


SOFTWARE

Open Access



# ROGUE: an R Shiny app for RNA sequencing analysis and biomarker discovery

Alvin Farrel<sup>1,2,3,4\*</sup> , Peng Li<sup>1,2</sup>, Sharon Veenbergen<sup>1,5,7</sup>, Khushbu Patel<sup>3,4</sup>, John M. Maris<sup>3,6</sup> and Warren J. Leonard<sup>1,2\*</sup>

\*Correspondence:

farrela@chop.edu; wjl@helix.nih.gov

<sup>1</sup> Laboratory of Molecular Immunology, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA

<sup>2</sup> Immunology Center, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA

<sup>3</sup> Division of Oncology and Center for Childhood Cancer Research, Children's Hospital of Philadelphia, Philadelphia, PA, USA

<sup>4</sup> Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA, USA

<sup>5</sup> Laboratory of Pediatric Gastroenterology, Erasmus University Medical Center, Rotterdam, The Netherlands

<sup>6</sup> Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

<sup>7</sup> Present Address: Laboratory of Medical Immunology, Department of Immunology, Erasmus University Medical Center, Rotterdam, The Netherlands

## Abstract

**Background:** The growing power and ever decreasing cost of RNA sequencing (RNA-Seq) technologies have resulted in an explosion of RNA-Seq data production. Comparing gene expression values within RNA-Seq datasets is relatively easy for many interdisciplinary biomedical researchers; however, user-friendly software applications increase the ability of biologists to efficiently explore available datasets.

**Results:** Here, we describe ROGUE (RNA-Seq Ontology Graphic User Environment, <https://marisshiny.research.chop.edu/ROGUE/>), a user-friendly R Shiny application that allows a biologist to perform differentially expressed gene analysis, gene ontology and pathway enrichment analysis, potential biomarker identification, and advanced statistical analyses. We use ROGUE to identify potential biomarkers and show unique enriched pathways between various immune cells.

**Conclusions:** User-friendly tools for the analysis of next generation sequencing data, such as ROGUE, will allow biologists to efficiently explore their datasets, discover expression patterns, and advance their research by allowing them to develop and test hypotheses.

**Keywords:** R Shiny, RNA-Seq, Differential expression, Biomarkers, GSEA, Gene ontology

## Background

RNA sequencing (RNA-Seq) has become an extremely powerful tool for understanding biological pathways and molecular mechanisms. Technological advancements, both wet-lab and computational, have transformed RNA-Seq into a more accessible tool, giving biomedical researchers access to a less biased view of RNA biology and transcriptomics [1–3]. The growing power and ever decreasing cost of RNA-Seq technologies have resulted in a marked increase in RNA-Seq dataset production.

The explosion of computational algorithms and pipelines in the last decade has given researchers the ability to perform rigorous analyses and explore RNA-Seq data [4–9]. Differential expression analysis (DEA) [10–13], which is the most common analysis performed on RNA-Seq, is used to estimate steady-state mRNA levels. There are multiple bioinformatics pipelines and packages used to perform DEA [13], including edgeR [10],



This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

DESeq[11], and limma-voom [12]. Different combinations of the various algorithms to analyze sequence reads and perform DEA can affect the biological conclusions drawn from the data [7, 14–16]. Researchers must carefully select the optimal combination of tools based on their specific biological questions and the available computational resources to perform deep dives and thorough exploration of their RNA-Seq data [7].

DEA is often combined with gene ontology (GO) analysis, pathway analysis, and clustering algorithms to characterize data and elucidate the processes and dynamics involved in transcription [17]. These studies give new insights into gene regulatory networks and expression. Gene enrichment analysis is a standard GO approach to evaluate upregulated pathways and processes [17–20]. Dimensionality reduction methods, such as multidimensional scaling (MDS) [10, 21], principal component analysis (PCA) [22, 23], and t-distributed stochastic neighbor embedding (t-SNE) [24], are used to identify RNA-Seq libraries with similar gene expression profiles. Moreover, while many other sophisticated RNA-Seq technologies exist, such as isoform analyses, single-cell RNA-Seq, and spatially resolved RNA-Seq methods, bulk RNA-Seq remains a powerful tool that continues to shape our understanding of biology.

The availability of RNA sequencing datasets is becoming more common due to increased support of open data by academicians and requirements by scientific journals and funding agencies to make publication-affiliated datasets publicly available. This has gifted the scientific community with an extensive repository of datasets [25–27] derived from cell lines, animal models, and patient-derived samples of a wide variety of tissues and diseases. Researchers can explore these datasets of interest to generate or test hypotheses. However, even standard DEA and GO analyses often requires a bioinformatician or a computationally savvy biologist.

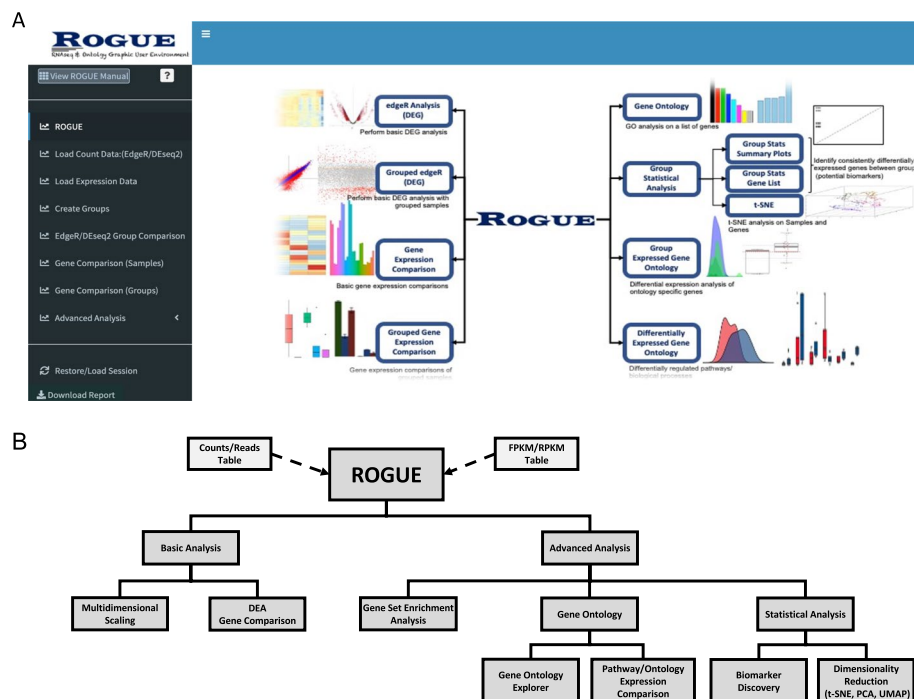
User-friendly tools for RNA-Seq analyses will allow biomedical scientists with limited programming experience to explore these datasets. Here we present RNA-Seq Ontology Graphic User Environment (ROGUE), an R Shiny application that allows biologists to perform differentially expressed gene analysis, gene ontology and pathway enrichment analysis, potential biomarker identification, and advanced statistical analyses. We demonstrate the capability of ROGUE by exploring the basic differences between CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells, and natural killer (NK) cells. Furthermore, we show how ROGUE can be used to identify biomarkers and differentially enriched pathways present in similar immune cells in different diseases.

We propose that ROGUE will allow scientists to explore their datasets and also compare their findings with publicly available datasets, increasing the potential of data-driven biomedical discovery.

## Methods

### Workflow

ROGUE is an R Shiny web app with a graphic user interface (GUI) (Fig. 1A) that takes expression data as input such as raw read counts, length-normalized counts, expression units including fragments per kilobase of transcript per million mapped reads (FPKM), reads per kilobase of transcript per million mapped reads (RPKM), and transcripts per million (TPM). Users can generate their own RNA-Seq matrix or download publicly available RNA-Seq expression data from databases such as gene expression omnibus



**Fig. 1** **A** ROGUE R Shiny app graphic user interface. **B** ROGUE workflow. ROGUE takes raw read counts, normalized counts, or quantified expression values (RPKM, FPKM, TPM) as input. The user can quickly look at the clustering of all samples based on the expression values of all genes, perform differential expression analysis, and compare genes between samples or groups. ROGUE also includes statistical tools for gene set enrichment analysis (GSEA), gene ontology (GO) analysis, biomarker discovery, and dimensionality reduction by t-SNE, PCA, or UMAP

(GEO) [25], ArrayExpress [26], The genotype tissue expression (GTEx) Project [27], and the cancer genome atlas (TCGA) [28]. An online manual is available at <https://maris.shiny.research.chop.edu/ROGUE/Instructions.pdf>. When the input is raw read counts or length-normalized counts quantified by packages such as HT-seq [29] or RSEM [8], ROGUE generates RPKM tables and can perform DEA using edgeR [10] or DESeq2 [11] which are two of the state-of-the-art R packages for DEA analysis [13] and has been shown to outperform other methods in various applications [30, 31]. ROGUE also allows users to perform more advanced analyses such as biomarker discovery based on gene expression, dimensionality reduction, gene set enrichment analysis, and gene ontology analysis (Fig. 1B).

Gene expression comparison between samples and groups can be visualized with heatmaps, bar plots, and boxplots. Users can also use ROGUE to predict possible biomarkers by ranking genes with maximized fold change and minimized coefficients of variation in gene expression between groups of samples. The Welch's t-test and the Wilcoxon Rank Sum Test can also be used to rank genes by their difference in expression distribution between the groups using the Biomarker Discovery Tool.

Gene set enrichment analysis (GSEA) is a computational method that determines whether a pre-ranked (i.e., log fold change) gene list shows statistically significant, concordant differences between two biological states (e.g., CD4<sup>+</sup> vs. CD8<sup>+</sup> T cells). GSEA between individual samples or groups can be performed using the Fast Gene Set

Enrichment Analysis (fgsea) R package [32] with data imported from the Molecular Signatures Database (MSigDB) [18, 33]. Alternatively, gene ontology analysis on a list of differentially expressed genes can be performed using the Gene Ontology Resource [17, 34], which is imported into ROGUE. Furthermore, ROGUE can determine differentially expressed gene sets using the Gene Ontology Resource. This resource uses the Wilcoxon rank sum test to determine if the expression of all genes within a biological process or molecular function are statistically different between samples or groups.

Dimensionality reduction methods can be applied to the datasets and visualized using 2-dimensional and 3-dimensional plots. ROGUE performs PCA using the 'prcomp' R function, t-SNE using the 'Rtsne' R package [35], and Uniform Manifold Approximation and Projection (UMAP) method for dimensionality reduction using the 'uwot' R package [36–38].

The source code for ROGUE is available at <https://github.com/afarrel/ROGUE>. All packages and implementation of the tools are described at this repository.

## Datasets

We performed basic analyses on datasets GSE60424 [39], GSE102317 [40], and GSE40350 [41] and GSE101470 [42] from the GEO Database to illustrate the basic features of ROGUE. Human CD4<sup>+</sup> and CD8<sup>+</sup> T cells, NK cells, neutrophils, and monocytes from healthy subjects and subjects diagnosed with type 1 diabetes, amyotrophic lateral sclerosis, sepsis, and multiple sclerosis were retrieved from GSE60424. RNA-Seq data from mouse CD4<sup>+</sup> and CD8<sup>+</sup> T cells and NK cells were retrieved from GSE102317, GSE40350, and GSE101470, respectively, for additional analyses. Dataset GSE102317 contains RNA-Seq data from CD4<sup>+</sup> T cells treated with IL-2 and IL-21 for 0 (control), 2, 4, and 24 h. Dataset GSE40350 contains CD8<sup>+</sup> T cells treated with IL-2 and IL-15 for 0 (control), 4, and 24 h. Dataset GSE101470 includes RNA-Seq from mature CD11b<sup>-</sup>/CD27<sup>-</sup>, CD11b<sup>-</sup>/CD27<sup>+</sup>, CD11b<sup>+</sup>/CD27<sup>+</sup>, and CD11b<sup>+</sup>/CD27<sup>low</sup> NK cells as well as *Stat5* double knock-in mice with N-terminal mutations in STAT5A and STAT5B that prevent STAT5 tetramerization but not dimerization.

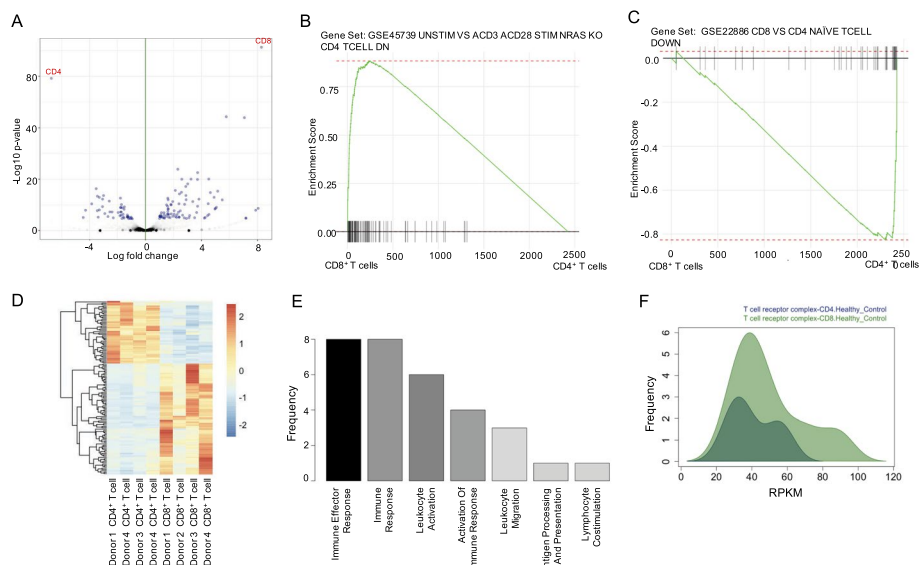
## Results and discussion

### Workflow

We demonstrate the capability of ROGUE by exploring some basic differences between CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells, and natural killer (NK) cells in datasets downloaded from the GEO Database. First, we performed DEA using edgeR [10] and compared the expression of genes of interest between cell types. We then performed GSEA, GO analysis, and biomarker discovery based on gene expression to understand functional differences between the cells and discover possible biomarkers. We used ROGUE to perform dimensionality reduction by t-SNE to evaluate if the transcriptome of these cells were distinct enough to cluster each sample by cell type. Finally, we searched for differentially expressed gene sets from the GO Resource to evaluate changes in pathways pre and post-interferon beta (IFN $\beta$ ) treatment in immune cells from patients with multiple sclerosis (MS).

### Basic DEA and GO analysis

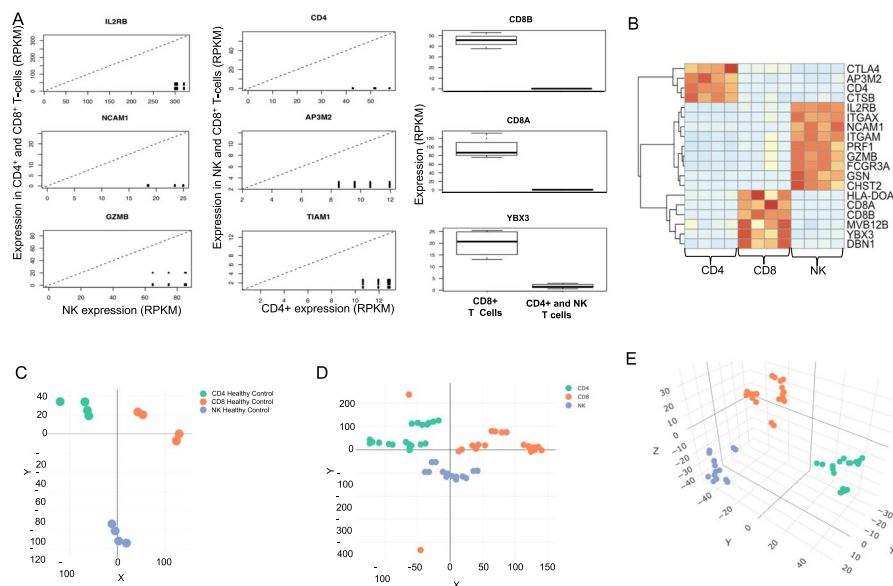
To illustrate the basic features of ROGUE, we first performed DEA on CD4<sup>+</sup> T cells versus CD8<sup>+</sup> T cells from healthy humans in dataset GSE60424 using edgeR [10] and generated a volcano plot showing the differentially expressed genes (Fig. 2A). We next performed GSEA using the ‘fgsea’ R package to identify enriched gene signatures from the differentially expressed genes between CD4<sup>+</sup> T cells and CD8<sup>+</sup> T cells from healthy humans (Fig. 2B, C, Additional file 1). For this illustration, we expect to see gene sets with enhanced expression in experiments with stimulated CD8<sup>+</sup> T cells or with lower expression in CD4<sup>+</sup> T cells to be enriched in our CD8<sup>+</sup> T cells RNA-Seq libraries and lower in our CD4<sup>+</sup> T cell libraries. Interestingly, the most enriched gene set for CD8<sup>+</sup> T cells when compared to CD4<sup>+</sup> T cells was a set (GSE45739) of genes downregulated in CD4<sup>+</sup> T cells with *Nras* knockout (KO) mice (Fig. 2B). While CD4<sup>+</sup> thymocyte differentiation is not affected in *Nras* KO mice, CD8<sup>+</sup> thymocyte differentiation has been shown to be significantly reduced [43]. Not surprisingly, the most enriched gene set for CD4<sup>+</sup> human T cells, was a set (GSE22886) of genes downregulated in naïve CD8<sup>+</sup> T cells when compared to CD4<sup>+</sup> T cells (Fig. 2C). A heatmap was used to display the distinct expression patterns of the differentially expressed genes between CD4<sup>+</sup> and CD8<sup>+</sup> T cells from the four healthy donors in the dataset (Fig. 2D). Basic GO analysis of genes upregulated in CD8<sup>+</sup> T cells showed enrichment in genes related to immune effector process, immune response, and leukocyte activation (Fig. 2E). We next used the gene ontology comparison tool to evaluate which type of T cell expresses more genes related to the T cell receptor (TCR) complex. This analysis interestingly revealed that the TCR complex was more enriched in the CD8<sup>+</sup> T cells as they expressed more genes at greater RPKM than the CD4<sup>+</sup> T cells (Fig. 2F).



**Fig. 2** Basic analysis of CD4<sup>+</sup> T cells versus CD8<sup>+</sup> T cells in healthy individuals. **A** Volcano plot showing differentially expressed genes. **B** GSEA showing the most enriched gene set when CD8<sup>+</sup> T cells were compared to CD4<sup>+</sup> T cells. **C** GSEA shows that the gene set downregulated in naïve CD8<sup>+</sup> T cells when compared to naïve CD4<sup>+</sup> T cells followed the same pattern in the current dataset. **D** Heatmap showing top differentially expressed genes between CD8<sup>+</sup> and CD4<sup>+</sup> T cells. **E** Gene ontologies of genes upregulated in CD8<sup>+</sup> T cells. **F** Distribution of expressed genes related to the T cell receptor complex

### Biomarker discovery

Biomarker discovery is essential in biomedical and pharmaceutical research [44–46]. Although mRNA is not always translated into protein, one can infer potential biomarkers from RNA-Seq data. ROGUE uses an optional combination of the coefficient of variation (CV), Wilcoxon-ranked sum test, or t-test for biomarker discovery between RNA-Seq library groups. ROGUE was used to identify potential biomarkers between CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells, and NK cells using the Biomarker Discovery tool (Fig. 3A), and a subset of these potential biomarkers was compared across the various cell types using a heatmap (Fig. 3B). The expression values of the potential biomarkers were used to perform t-SNE on all the RNA-Seq libraries. A 2-dimensional plot of the t-SNE results shows that RNA-Seq libraries from CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells, and NK cells from healthy controls clustered reasonably well based on the potential biomarkers discovered (Fig. 3C). Clusters were not as distinct when t-SNE was performed on T and NK cell libraries from both healthy controls and patients in 2 dimensions (Fig. 3D), but the clusters in a 3-dimensional plot generated by t-SNE were more homogeneous (Fig. 3E). We evaluated the occurrence of these biomarkers in mouse immune cells and observed that only a few of the biomarkers can be used across all datasets in both species (Additional file 2: A–B). As expected, CD4 and CTLA4 were identified as potential biomarkers for differentiating CD4<sup>+</sup> T cells from CD8<sup>+</sup> T cells and NK cells across both datasets while CD8A and CD8B were identified as potential biomarkers for CD8<sup>+</sup> T cells. Gene expression of the potential human NK cell biomarkers were enriched in mouse NK cells that expressed CD27 (Additional file 2: C–D). t-SNE was performed on the mouse datasets

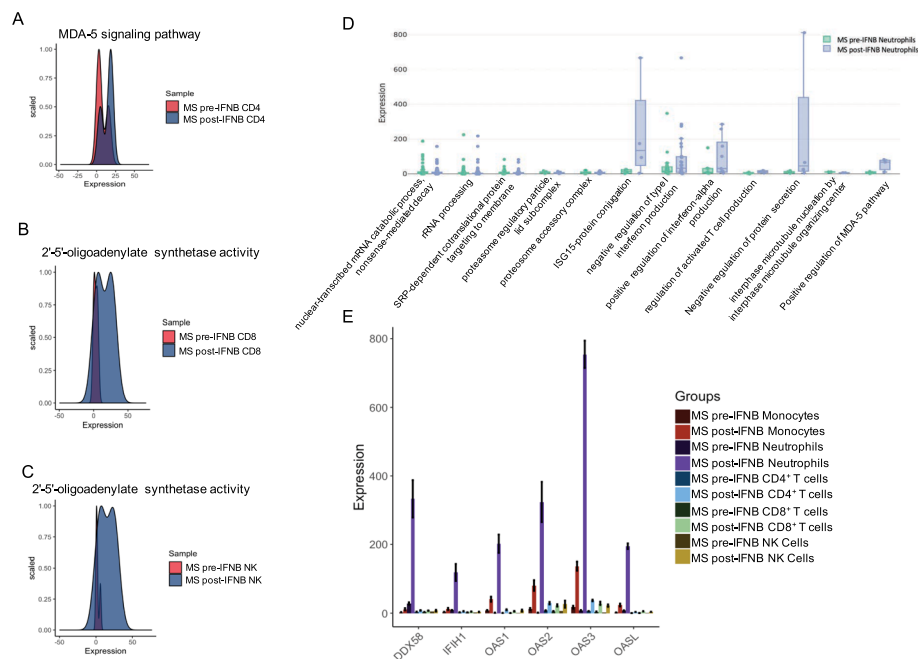


**Fig. 3** Biomarker Discovery among CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells, and NK cells. **A** The Biomarker tool shows genes with high expression in one cell type and very low expression in the other cell types, suggesting that they may be potential biomarkers. **B** Heatmap showing relative expression values of the potential biomarkers' gene expression for CD4<sup>+</sup> T cell, CD8<sup>+</sup> T cell, and NK cell groups. **C** 2-dimensional t-SNE plot of CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells, and NK cells from healthy controls using the identified potential biomarkers. **D** 2-dimensional t-SNE plot of CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells, and NK cells from both healthy and diseased groups using the identified potential biomarkers. **E** 3-dimensional t-SNE plot using the identified potential biomarkers emphasizes separation between clusters

using the gene expressions of the potential biomarkers. The enrichment of the potential human NK cell biomarkers in mouse CD27<sup>+</sup> NK cells was reflected in the t-SNE plot as they formed a distinct cluster from the other NK cells (Additional file 2: E). It is worth noting that even though the mouse immune cells cluster well using the biomarkers ascertained from the human immune cells, it is possible that the immune cells cluster well due to a batch effect instead of gene expression signature (Additional file 2: F).

### Comparison of biological pathways after treating multiple sclerosis patients with IFN $\beta$

Dataset GSE60424 contains RNA-Seq data from CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells, NK cells, neutrophils, and monocytes of MS patients before and after IFN $\beta$  treatment. MS is an inflammatory demyelinating disease of the central nervous system [47]. IFN $\beta$  treatment is a safe and reasonably effective treatment for MS patients [48–51] due to its anti-inflammatory and immunomodulatory effects [52, 53]. While this is a widely-used treatment, the precise mechanism is unknown. To identify potential hypotheses of the mechanism downstream of IFN $\beta$  treatment, we used ROGUE to identify differentially expressed biological processes in CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells, and NK cells isolated from patients pre- or post-treatment with IFN $\beta$ . CD4<sup>+</sup> T cells showed upregulation of the MDA-5 signaling pathway, among other biological processes (Fig. 4A and Additional file 3). CD8<sup>+</sup> T cells and NK cells showed upregulation of 2′–5′-oligoadenylate synthetase activity (Fig. 4B, C). Given that the MDA-5 signaling pathway and 2′–5′-oligoadenylate synthetase activity are both involved in interferon signaling in innate immunity [54–57], and both pathways were upregulated in CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells, and NK cells (Additional file 4: A), we examined the differentially expressed biological processes in neutrophils and monocytes. As expected, we observed an increase in pathways related to interferon production, protein secretion, as well as positive regulation of MDA-5 pathway (Fig. 4D and Additional file 4: B). This led us to examine the expression of genes related to MDA-5 and 2′–5′-oligoadenylate synthetase in all five cell types pre- and post-treatment, as this might give insights into the underlying mechanism. Furthermore, there is at least one report that polymorphisms in MDA-5 (IFIH1) are associated with MS [58], although another report states that this association does not exist in a specific French population [59]. Nevertheless, the MDA-5 signaling pathway and 2′–5′-oligoadenylate synthetase activity were upregulated in all five cell types (Fig. 4E). We then confirmed that both MDA-5 and RIG-I (DDX58) are upregulated in MS patients' immune cells following treatment with IFN $\beta$ , as they are involved in the induction of IRF7 expression and constitutively-expressed IRF3 [57, 60] (Additional file 4: C). A well-defined mechanism of interferon-stimulated gene (ISG) expression is that IRF3 and IRF7 regulate the expression of type 1 interferons, which then induce ISGs through JAK-STAT signaling, including OAS1A and OAS1B [56]. However, IFN $\alpha$  and IFN $\beta$  mRNAs are not expressed, which suggests that administered IFN $\beta$  rather than endogenously produced IFN $\beta$  induces ISGs through the JAK-STAT pathway. This model is consistent with our data, as ISGs were upregulated in all five cell types after IFN $\beta$ -treatment with significantly greater expression of MDA-5, RIG-1 and ISGs observed in neutrophils (Fig. 4E and Additional file 4: C). Given that 2′–5′-oligoadenylate synthetase can induce apoptosis in tumors [61], perhaps this alternative role of 2′–5′-oligoadenylate



**Fig. 4** Using differentially expressed pathways to generate and/or explore hypotheses. **A** Distribution of MDA-5 signaling pathway in untreated (red) and IFN $\beta$ -treated (blue) CD4<sup>+</sup>T cells showing an upregulation of genes related to MDA-5 signaling. **B, C** Distribution of 2'-5'-oligoadenylate synthetase activity in untreated (red) and IFN $\beta$ -treated (blue) CD8<sup>+</sup>T cells (**B**) and NK cells (**C**) showing an upregulation of genes 2'-5'-oligoadenylate synthetase activity. **D** Boxplots showing pathways that may be differentially regulated in neutrophils with multiple genes consistently upregulated or downregulated post-IFN $\beta$  treatment. **E** Bar plot showing upregulated MDA-5 (IFIH1), RIG-I (DDX58), and genes related to 2'-5'-oligoadenylate synthetase in IFN $\beta$ -treated monocytes, neutrophils, CD4<sup>+</sup>T cells, CD8<sup>+</sup>T cells, and NK cells

synthetase also occurs in immune cells, giving it a pro-inflammatory role as well as an anti-inflammatory role by promoting apoptosis and regulating cell growth and proliferation [57]. Furthermore, the disproportionate upregulation of genes with pro-apoptotic and antiproliferative roles in neutrophils supports Hasselbalch and Søndergaard's report of a higher neutrophil-to-lymphocyte ratio, which is a marker of systemic inflammation, before treatment with IFN $\beta$  when compared to controls by [62]. Moreover, Pierson et al. demonstrated that depleting neutrophils in the MS animal model reduces the progression of the disease and Naegel et al. showed evidence that the increase in neutrophils in relapsing–remitting MS is likely due to decreased apoptosis [63, 64]. If this potential pro-apoptotic anti-inflammatory role of 2'-5'-oligoadenylate synthetase exists, it could be the mechanism by which IFN $\beta$  treatment positively impacts MS patients. In addition to 2'-5'-oligoadenylate synthetase activity, IFN $\beta$  may be involved in another pathway that explains the MDA-5/RIG-1 upregulation. Shimoni et al. suggested that IFN $\beta$  can bind cell surface receptors and promote the induction of RIG-1 as part of a positive feedback loop [65]. Wang et al. further showed that RIG-1 and MDA5 signaling induces tumor necrosis factor (TNF) in macrophages [66], and TNF has been shown to have anti-inflammatory effects in MS [67]. The anti-inflammatory effects of TNF coupled with the pro-apoptotic role of



2′–5′-oligoadenylate synthetase may be part of the downstream mechanism contributing to the positive response induced by IFN $\beta$  in MS patients.

## Conclusion

ROGUE is designed to be a user-friendly R Shiny application that allows users to perform basic tasks with available RNA-Seq data such as differentially expressed gene analysis and gene ontology analysis. While other freely available web tools and portals have been developed to allow researchers to address discrete questions based on molecular and genomic datasets without the need for strong computational skills [68, 69], ROGUE allows deeper dataset exploration, allowing users to compare gene expression and gene set enrichments between samples and groups. For example, users can explore similarities of expression profiles using the dimensionality reduction methods such as t-SNE, PCA, UMAP, and MDS and search for potential biomarkers between groups of RNA-Seq libraries, to our knowledge making it the only currently available tool to allow this range of dataset analysis (Additional file 5). Furthermore, users have the option to download their session and continue their analysis at a later time. Users can also restore a session if the web application gets disconnected from the server. In addition to the case study presented here, we successfully tested ROGUE on ten diverse human and mouse case studies downloaded from Expression Atlas to illustrate the various applications and robustness (Additional file 6). It is worth noting that ROGUE is an R Shiny application thus allowing the inclusion of many statistical and graphical functions by the R community as well as the ability to be implemented on both local and web servers; however, like all R Shiny applications there are limitations. One of these limitations is that R Shiny applications that are implemented on web servers may perform slowly and sometimes disconnect from the server resulting in a subsequent crash when processing large datasets or performing computationally intensive functions. For this reason, we recommend downloading the local version of ROGUE from <https://github.com/afarrel/ROGUE> when processing large datasets. Here, we show that a user can explore RNA-Seq data obtained from public databases and use ROGUE to analyze that data to generate or support new or existing hypotheses. ROGUE provides non-R programmers access to many statistical and graphical R packages for RNA-Seq analyses through a GUI so they can analyze their data and create figures. Ideally, tools like ROGUE will allow more biomedical researchers to take advantage of genomic data available and help expedite needed bioinformatics analyses. ROGUE is available at <https://marisshiny.research.chop.edu/ROGUE/>.

## Availability and requirements

Project Name: ROGUE.

Project Home Page: <https://marisshiny.research.chop.edu/ROGUE/>.

Github: <https://github.com/afarrel/ROGUE>.

Operating System: Platform independent.

Programming language: R.

Other requirements: R environment and included packages. Tested on R version 3.6.

Any restrictions to use by non-academics: none.

**Abbreviations**

CV	Coefficient of variation
DEA	Differential expression analysis
FGSEA	Fast gene set enrichment analysis
FPKM	Fragments per kilobase of transcript per million mapped reads
GEO	Gene expression omnibus
GO	Gene ontology
GSEA	Gene set enrichment analysis
GTE <sub>x</sub>	The genotype tissue expression
GUI	Graphic user interface
IFN <sub>α</sub>	Interferon alpha
IFN <sub>β</sub>	Interferon beta
ISG	Interferon-stimulated gene
MDS	Multidimensional scaling
MSigDB	Molecular signatures database
MS	Multiple sclerosis
NK	Natural killer
PCA	Principal component analysis
RNA-Seq	RNA sequencing
ROGUE	RNA-Seq ontology graphic user environment
RPKM	Reads per kilobase of transcript per million mapped reads
TCGA	The cancer genome atlas
TCR	T cell receptor
TNF	Tumor necrosis factor
TPM	Transcripts per million
t-SNE	T-distributed stochastic neighbor embedding
UMAP	Uniform manifold approximation and projection

**Supplementary Information**

The online version contains supplementary material available at <https://doi.org/10.1186/s12859-023-05420-y>.

**Additional file 1:** GSEA analysis of healthy human CD8<sup>+</sup> T cells vs CD4<sup>+</sup> T cells.

**Additional file 2:** Evaluating biomarkers found in human CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells, and NK cells in mouse immune cells from different datasets.

**Additional file 3:** Distribution of gene expression profiles in the differentially expressed pathways.

**Additional file 4:** Evaluation of MD5A-signaling, RIG-1 signaling, and 2'-5'-oligoadenylate synthetase pre- and post-IFN<sub>β</sub> treatment.

**Additional file 5:** Available Rshiny RNAseq analysis tools.

**Additional file 6:** List of case studies.

**Acknowledgements**

We thank Jian-Xin Lin for discussions and suggestions in reviewing the application and the manuscript. We are grateful to Grace Coggins and Laura Scolaro for using this application in their research and suggesting changes to increase usability.

**Author contributions**

Conceptualization, AF; Methodology, AF, PL, and SV; Software, AF, PL and KP; Formal Analysis, AF; Funding Acquisition, WJL and JMM; Writing, AF, PL, SV, JMM and WJL; Supervision, AF and WJL. All authors read and approved the final manuscript.

**Funding**

This work was supported by the Division of Intramural Research, National Heart, Lung, and Blood Institute, NIH, and NIH grants R35 CA220500, P01 CA217959 and U54 CA232568 (JMM).

**Availability of data and materials**

RNA sequencing expression data from human immune cells: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE60424>. RNA sequencing expression data from mouse immune cells: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE102317>. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE40350>. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE101470>

**Declarations****Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

### Competing interests

The authors declare that they have no competing interest.

Received: 13 November 2020 Accepted: 18 July 2023

Published online: 29 July 2023

### References

1. Garalde DR, Snell EA, Jachimowicz D, Sipos B, Lloyd JH, Bruce M, et al. Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods*. 2018;15:201–6.
2. Smith AM, Jain M, Mulroney L, Garalde DR, Akeson M. Reading canonical and modified nucleobases in 16S ribosomal RNA using nanopore native RNA sequencing. *PLoS ONE*. 2019;14: e0216709.
3. Byrne A, Beaudin AE, Olsen HE, Jain M, Cole C, Palmer T, et al. Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. *Nat Commun*. 2017;8:16027.
4. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28:511–5.
5. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc*. 2016;11:1650–67.
6. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 2016;34:525–7.
7. Sahraeian SME, Mohiyuddin M, Sebra R, Tilgner H, Afshar PT, Au KF, et al. Gaining comprehensive biological insight into the transcriptome by performing a broad-spectrum RNA-seq analysis. *Nat Commun*. 2017;8:59.
8. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform*. 2011;12:323.
9. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 2012;7:562–78.
10. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinform Oxf Engl*. 2010;26:139–40.
11. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
12. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*. 2014;15:R29.
13. Costa-Silva J, Domingues D, Lopes FM. RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS ONE*. 2017;12: e0190152.
14. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. *Genome Biol*. 2016;17:13.
15. Seyednasrollah F, Laiho A, Elo LL. Comparison of software packages for detecting differential expression in RNA-seq studies. *Brief Bioinform*. 2015;16:59–70.
16. Williams CR, Baccarella A, Parrish JZ, Kim CC. Empirical assessment of analysis workflows for differential expression analysis of human samples using RNA-Seq. *BMC Bioinform*. 2017;18:38.
17. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25:25–9.
18. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;102:15545–50.
19. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res*. 2019;47:D419–26.
20. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S. AmiGO: online access to ontology and annotation data. *Bioinform Oxf Engl*. 2009;25:288–9.
21. Loraine AE, Blakley IC, Jagadeesan S, Harper J, Miller G, Firon N. Analysis and visualization of RNA-Seq expression data using RStudio, bioconductor, and integrated genome browser. *Methods Mol Biol Clifton NJ*. 2015;1284:481–501.
22. Jolliffe I. Principal component analysis. In: Lovric M, editor. *International encyclopedia of statistical science*. Berlin: Springer; 2011. p. 1094–6. [https://doi.org/10.1007/978-3-642-04898-2\\_455](https://doi.org/10.1007/978-3-642-04898-2_455).
23. Jolliffe I. Principal component analysis. In: *Wiley StatsRef: Statistics Reference Online*. American Cancer Society; 2014. <https://doi.org/10.1002/9781118445112.stat06472>.
24. van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res*. 2008;9:2579–605.
25. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCB gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002;30:207–10.
26. Athar A, Füllgrabe A, George N, Iqbal H, Huerta L, Ali A, et al. ArrayExpress update—from bulk to single-cell expression data. *Nucleic Acids Res*. 2019;47:D711–5.
27. Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A, Buia SA, et al. A novel approach to high-quality postmortem tissue procurement: the GTEx project. *Biopreserv Biobank*. 2015;13:311–9.
28. The Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*. 2013;45:1113–20.
29. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31:166–9.

30. Schurch NJ, Schofield P, Gierliński M, Cole C, Sherstnev A, Singh V, et al. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA NY N*. 2016;22:839–51.
31. Assefa AT, De Paepe K, Everaert C, Mestdagh P, Thas O, Vandesompele J. Differential gene expression analysis tools exhibit substandard performance for long non-coding RNA-sequencing data. *Genome Biol*. 2018;19:96.
32. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis. *bioRxiv*. 2021; 060012.
33. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinform Oxf Engl*. 2011;27:1739–40.
34. Gene Ontology Consortium. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res*. 2021;49:D325–34.
35. Krijthe JH. Rtsne: T-Distributed stochastic neighbor embedding using Barnes–Hut implementation. 2015. <https://github.com/krijthe/Rtsne>.
36. Melville J. uwot: the uniform manifold approximation and projection (UMAP) method for dimensionality reduction. 2020. <https://CRAN.R-project.org/package=uwot>.
37. McInnes L, Healy J, Melville J. UMAP: uniform manifold approximation and projection for dimension reduction. 2020.
38. Becht E, McInnes L, Healy J, Dutertre C-A, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*. 2018;37:38–44.
39. Linsley PS, Speake C, Whalen E, Chaussabel D. Copy number loss of the interferon gene cluster in melanomas is linked to reduced T cell infiltrate and poor patient prognosis. *PLoS ONE*. 2014;9:e109760.
40. Li P, Mitra S, Spolski R, Oh J, Liao W, Tang Z, et al. STAT5-mediated chromatin interactions in superenhancers activate IL-2 highly inducible genes: Functional dissection of the Il2ra gene locus. *Proc Natl Acad Sci USA*. 2017;114:12111–9.
41. Lin J-X, Du N, Li P, Kazemian M, Gebregiorgis T, Spolski R, et al. Critical functions for STAT5 tetramers in the maturation and survival of natural killer cells. *Nat Commun*. 2017;8:1320.
42. Ring AM, Lin J-X, Feng D, Mitra S, Rickert M, Bowman GR, et al. Mechanistic and structural insight into the functional dichotomy between IL-2 and IL-15. *Nat Immunol*. 2012;13:1187–95.
43. Pérez de Castro I, Diaz R, Malumbres M, Hernández M-I, Jagirdar J, Jiménez M, et al. Mice deficient for N-ras: impaired antiviral immune response and T-cell function. *Cancer Res*. 2003;63:1615–22.
44. Liotta LA, Ferrari M, Petricoin E. Clinical proteomics: written in blood. *Nature*. 2003;425:905.
45. Goossens N, Nakagawa S, Sun X, Hoshida Y. Cancer biomarker discovery and validation. *Transl Cancer Res*. 2015;4:256–69.
46. Norouzinia M, Chaleshi V, Alizadeh AHM, Zali MR. Biomarkers in inflammatory bowel diseases: insight into diagnosis, prognosis and treatment. *Gastroenterol Hepatol Bed Bench*. 2017;10:155–67.
47. Wootla B, Eriguchi M, Rodriguez M. Is multiple sclerosis an autoimmune disease? *Autoimmune Dis*. 2012;2012:969657.
48. Jacobs L, Brownscheidle CM. Appropriate use of interferon beta-1a in multiple sclerosis. *BioDrugs*. 1999;11:155–63.
49. Clerico M, Contessa G, Durelli L. Interferon- $\beta$ 1a for the treatment of multiple sclerosis. *Expert Opin Biol Ther*. 2007;7:535–42.
50. Bertolotto A, Gilli F. Interferon-beta responders and non-responders. A biological approach. *Neurol Sci Off J Ital Neurol Soc Ital Soc Clin Neurophysiol*. 2008;29(Suppl 2):S216-217.
51. Freedman MS. Long-term follow-up of clinical trials of multiple sclerosis therapies. *Neurology*. 2011;76(Supplement 1):S26.
52. Kasper LH, Reder AT. Immunomodulatory activity of interferon-beta. *Ann Clin Transl Neurol*. 2014;1:622–31.
53. Dhib-Jalbut S, Marks S. Interferon- $\beta$  mechanisms of action in multiple sclerosis. *Neurology*. 2010;74(Supplement 1):S17.
54. Loo Y-M, Gale MJ. Immune signaling by RIG-I-like receptors. *Immunity*. 2011;34:680–92.
55. Qiu L, Wang T, Tang Q, Li G, Wu P, Chen K. Long non-coding RNAs: regulators of viral infection and the interferon antiviral response. *Front Microbiol*. 2018;9:1621.
56. Pulit-Penalzo JA, Scherbik SV, Brinton MA. Type 1 IFN-independent activation of a subset of interferon stimulated genes in West Nile virus Eg101-infected mouse cells. *Virology*. 2012;425:82–94.
57. Choi UY, Kang J-S, Hwang YS, Kim Y-J. Oligoadenylate synthase-like (OASL) proteins: dual functions and associations with diseases. *Exp Mol Med*. 2015;47:e144–e144.
58. Martínez A, Santiago JL, Cénit MC, de Las HV, de la Calle H, Fernández-Arquero M, et al. IFIH1-GCA-KCNH7 locus: influence on multiple sclerosis risk. *Eur J Hum Genet EJHG*. 2008;16:861–4.
59. Couturier N, Gourraud P-A, Cournu-Rebeix I, Gout C, Bucciarelli F, Edan G, et al. IFIH1-GCA-KCNH7 locus is not associated with genetic susceptibility to multiple sclerosis in French patients. *Eur J Hum Genet EJHG*. 2009;17:844–7.
60. Malathi K, Dong B, Gale M Jr, Silverman RH. Small self-RNA generated by RNase L amplifies antiviral innate immunity. *Nature*. 2007;448:816–9.
61. Mullan PB, Hosey AM, Buckley NE, Quinn JE, Kennedy RD, Johnston PG, et al. The 2,5 oligoadenylate synthetase/RNaseL pathway is a novel effector of BRCA1- and interferon- $\gamma$ -mediated apoptosis. *Oncogene*. 2005;24:5492–501.
62. Hasselbalch IC, Søndergaard HB, Koch-Henriksen N, Olsson A, Ullum H, Sellebjerg F, et al. The neutrophil-to-lymphocyte ratio is associated with multiple sclerosis. *Mult Scler J Exp Transl Clin*. 2018;4:2055217318813183–2055217318813183.
63. Pierson ER, Wagner CA, Goverman JM. The contribution of neutrophils to CNS autoimmunity. *Clin Immunol Orlando Fla*. 2018;189:23–8.
64. Naegele M, Tillack K, Reinhardt S, Schippling S, Martin R, Sospedra M. Neutrophils in multiple sclerosis are characterized by a primed phenotype. *J Neuroimmunol*. 2012;242:60–71.
65. Shimoni Y, Nudelman G, Hayot F, Sealfon SC. Multi-scale stochastic simulation of diffusion-coupled agents and its application to cell culture simulation. *PLoS ONE*. 2011;6:e29298–e29298.

66. Wang F, Gao X, Barrett JW, Shao Q, Bartee E, Mohamed MR, et al. RIG-I mediates the co-induction of tumor necrosis factor and type I interferon elicited by myxoma virus in primary human macrophages. *PLoS Pathog.* 2008;4:e1000099.
67. Liu J, Marino MW, Wong G, Grail D, Dunn A, Bettadapura J, et al. TNF is a potent anti-inflammatory cytokine in autoimmune-mediated demyelination. *Nat Med.* 1998;4:78–83.
68. Sundararajan Z, Knoll R, Hombach P, Becker M, Schultze JL, Ulas T. Shiny-Seq: advanced guided transcriptome analysis. *BMC Res Notes.* 2019;12:432.
69. Su W, Sun J, Shimizu K, Kadota K. TCC-GUI: a Shiny-based application for differential expression analysis of RNA-Seq count data. *BMC Res Notes.* 2019;12:133.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

