

## Article

# ARGEOS: A New Bioinformatic Tool for Detailed Systematics Search in GEO and ArrayExpress

Gleb E. Gavrish <sup>1</sup>, Dmitry V. Chistyakov <sup>2,\*</sup> and Marina G. Sergeeva <sup>2</sup>

<sup>1</sup> Faculty of Bioengineering and Bioinformatics, Moscow Lomonosov State University, 119234 Moscow, Russia; ggavrish@fbb.msu.ru

<sup>2</sup> Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, 119992 Moscow, Russia; sergeeva@belozersky.msu.ru

\* Correspondence: chistyakov@belozersky.msu.ru; Tel.: +7-495-939-4332

**Simple Summary:** A systematic search for datasets of transcriptome data is a hefty task. Therefore, we developed the ARGEOS web tool, which simplifies the search and selection of datasets from various public databases. In addition, the service carries out an advanced analysis of a dataset, including collecting detailed protocols, information on the number of datasets, and providing additional reference information. An example of a cell polarization study exemplifies the effectiveness of the tool.

**Abstract:** Conduct a reanalysis of transcriptome data for studying intracellular signaling or solving other experimental problems is becoming increasingly popular. Gene expression data are archived as microarray or RNA-seq datasets mainly in two public databases: Gene Expression Omnibus (GEO) and ArrayExpress (AE). These databases were not initially intended to systematically search datasets, making it challenging to conduct a secondary study. Therefore, we have created the ARGEOS service, which has the following advantages that facilitate the search: (1) Users can simultaneously send several requests that are supposed to be used for systematic searches, and it is possible to correct the requests; (2) advanced analysis of information about the dataset is available. The service collects detailed protocols, information on the number of datasets, analyzes the availability of raw data, and provides other reference information. All this contributes to both rapid data analysis with the search for the most relevant datasets and to the systematic search with detailed analysis of the information of the datasets. The efficiency of the service is shown in the example of analyzing transcriptome data of activated (polarized) cells. We have performed a systematic search of studies of cell polarization (when cells are exposed to different immune stimuli). The web interface for ARGEOS is user-friendly and straightforward. It can be used by a person who is not familiar with database searching.

**Keywords:** GEO; ArrayExpress; polarization; systematics search; gene expression; bioinformatic tool; transcriptome



**Citation:** Gavrish, G.E.; Chistyakov, D.V.; Sergeeva, M.G. ARGEOS: A New Bioinformatic Tool for Detailed Systematics Search in GEO and ArrayExpress. *Biology* **2021**, *10*, 1026. <https://doi.org/10.3390/biology10101026>

Academic Editors:  
Ioannis Michalopoulos and  
Apostolos Maltras

Received: 20 August 2021  
Accepted: 4 October 2021  
Published: 11 October 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Conducting secondary studies using transcriptome data is increasingly used as a primary method for studies of intracellular signaling [1,2], the study of molecular signatures of various diseases and conditions [3,4], the search for biomarkers of diseases [5], drug development [6], or other experimental problems. This approach makes searching for data to solve the problems of systematic analysis necessary.

Gene expression data are archived as microarray or RNA-seq datasets mainly in two public databases: Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>, 4 October 2021) and ArrayExpress (AE; <https://www.ebi.ac.uk/arrayexpress/>, 4 October 2021). Newly appearing local public databases have already been integrated or are planned to be integrated with one of these large databases. For example, the Japanese GEA data will be indexed by AE [7]. To date, these databases contain more than 156,000 records

(datasets) and more than 4.5 million samples. Each dataset can contain one to three to several thousand samples. At the same time, it may be noted that these databases were not initially intended for searching datasets for systematic analysis, which makes it challenging to conduct secondary research.

Many services make it easier to work with the data itself (see review by Wang et al. 2019 [8]). The main directions of such programs are advanced processing of user-selected datasets with a specific focus. For example, at the moment, many tools have been developed for analyzing Array data from GEO: GEO2Enrich [9], ImaGEO [10], and GEOacle [11]. These programs allow users to analyze differentially expressed genes (DEGs) of the selected datasets by considering metadata. However, they are limited to microarray datasets. A package for R has also been developed that allows interaction with GEO data, called GEO2R [12]. Some services such as GREIN—an open-source resource for re-using GEO RNA-seq data [13]—or TACITuS—a web-based system that supports rapid query access to microarray and NGS repositories [14]—were offered to facilitate work with sequencing data.

At the same time, there is still no convenient tool for the primary selection of datasets on the topic selected by the user and, more often, the researcher has to perform this work manually, analyzing both GEO and AE. Indeed, most of the existing software solutions do not provide the ability to search and sort by research metadata, e.g., the concentration of a substance for cell experiments or the exposure time for experiments with animals, i.e., about the information stored in the research protocols. Consequently, there is a need to create a convenient tool for systematically searching datasets on a selected topic, integrating the two largest public gene expression databases, and extracting additional data from experimental protocols that characterize datasets in more detail.

Primary dataset search services include the work of Ivliev [15], which makes it easier to find individual datasets, but this complicates systematic analysis since it does not work with experiment protocols [15]. The solution to the problem of combining duplicate datasets when searching in GEO, Array, and the Genomic Expression Archive (GEA; <https://www.ddbj.nig.ac.jp/gea/>, 4 October 2021) is proposed in the OAE (All Of The Gene Expression) service, which was designed to index public gene expression data [16]. OAE allows users to observe the number of entries by keywords quickly and in the form of a histogram and provides information on the number of entries, dividing them by year, type, or organism. At the same time, the proposed tool has the same drawbacks as the work of Ivliev [15]. Indeed, in order to systematically search for suitable datasets, analyzing a large pool of information is necessary, including the number of samples in the dataset, logs and the year when the dataset was published, and, very importantly, experimental protocols that include data on the concentration of stimuli, the time of their use, and other details.

We faced the primary search for datasets for analysis when searching for transcriptome data on polarization experiments. Polarization is a complex of cellular changes that change the cellular phenotype [17,18]. Cell pro-inflammatory and anti-inflammatory phenotypes are determined, which are elicited by various stimuli, for example, lipopolysaccharide (LPS) and interleukin-4 (IL-4), respectively [17,18]. Functional phenotypes, which can be acquired by cells depending on the microenvironment or stimulation, are currently the focus of investigations into new anti-inflammatory therapeutic approaches [19,20]. Not surprisingly, there are many different transcriptome datasets in this area of research, which contain from two to thousands of samples in each and obtained from other cells, with different stimuli and different technical protocols.

Currently, there are 144,576 records in the GEO database. When searching for LPS, the GEO database produces 2178 records and 1011 records for IL-4. However, in the built-in search in GEO, filtering is possible only by organism, experiment type, or publication date. The user needs to filter the rest of the information about datasets manually. For a complex query that includes several keywords, it is necessary to formulate a complex

query, combining the keywords with the operators 'AND' and 'OR'. Thus, it is difficult to carry out a systematic search using standard tools.

ARGEOS aims to combine gene expression data and make them systematically searchable in a user-friendly form. We developed the ARGEOS algorithm by indexing the publicly available gene expression databases GEO and AE. The effectiveness of our ARGEOS service is shown by the example of searching for data on studies of cell polarization when the innate immune system is exposed to LPS, IL-4, and other stimuli.

## 2. Materials and Methods

### 2.1. Implementation

ARGEOS is written in Python 3.8 (Python Software Foundation, Wilmington, DE, USA). The search in GEO is carried out remotely by sending a request to National Center for Biotechnology Information (NCBI) E-Utilities [21]. Interaction with NCBI E-Utilities services occurs through the Entrezpy package [22]. The package provides a communication delay and fulfills other NCBI recommendations. The search in AE is performed by parsing the XML output, and the query is formed so that only the unique AE datasets are returned.

### 2.2. Obtaining GEO Data

Data analysis of GEO datasets is performed by analyzing an XML file located in the Miniml section on the NCBI GEO FTP server.

### 2.3. Obtaining AE Data

Parsing the records in the database is performed by querying the EBI AE service ([www.ebi.ac.uk/arrayexpress/](http://www.ebi.ac.uk/arrayexpress/), 4 October 2021). For each record found, an XML file is formed, which the program reads, distributes the information into variables, and prints in the final output. Protocol data are obtained separately and upon request by ID.

### 2.4. Data Analysis

GEO: The following information is collected from the block with information about the dataset: organism (organisms), number of samples (samples), experiment type (types), platform ID, dataset name, upload date, summary, and general design of the experiment. The ID (from the PubMed database) of the reference article is also collected. The following information is collected from the block of samples: cell type, treatment protocol, growth protocol, type of analyzed molecules, isolation protocol, and the characteristics field, which includes all other parameters of the sample. Sample information is filtered to remove duplicates. The user then receives one record entry. If the sample information is different, i.e., unique for each sample, then the user receives multiple protocols. For example, if the treatment protocol is the same for all samples, the user will receive it only once. However, if the authors provided a unique treatment protocol for each sample, the user will receive them all. AE: For each record found, an XML file is formed, which the program reads, distributes the information into variables, and prints in the final output. Thus, there is a union of two databases and the end-user receives the same information in the same columns, which conveys the appearance of data homogeneity. This is intended to facilitate the operations of filtering data and reading data for manual sorting or classification.

### 2.5. Analyzing PubMed Data

For each dataset from GEO and AE, if a reference article is specified, the program receives a PubMed ID or DOI, through which it searches PubMed by sending a request to NCBI E-Utilities [15]. The final table contains information about the title of the article, DOI, the name of the journal, and the impact factor calculated based on the name of the journal.

## 2.6. Data Export

After writing information about datasets into variables and bringing the information into the same format, the information is written to “EXPORT TABLE 2” in the following order.

Accession—record ID (GEO or AE); Organism—an organism (or organisms) specified by the authors; Samples—number of samples; Type—the type of experiment; Platform—the platform (or platforms) on which the raw data is received. Always indicated GEO, less commonly found in AE; Title—the name of the dataset; Year—the date of the first publication of the dataset; Summary—a summary dataset and a summary of the study provided by the authors; Link—link to the dataset (in GEO or AE); next is the block of reference articles. All articles indicated by the authors as related to this dataset are listed in the “All references” block. Among them, the article published in the journal with the highest impact factor stands out. Information about this article is listed in separate columns: Paper\_title—the title of the reference article; Journal—the journal in which the reference article was published; Impact factor—the impact factor of the magazine; DOI or PubMed—link to the article; All references—information on all articles (titles, magazines, and links); Type of molecule—the type of molecule analyzed in the study (for example, “total RNA” or “Genomic DNA”); BioProjectlink (NCBI)—Link to the BioProject page on the NCBI website; BioProjectlink (EBI)—Link to the BioProject page on the EBI website (EMBL); SRA—link to SRA (relevant for RNA-seq experiments); All protocols—the column contains all the textual information provided by the authors of the dataset. These can be protocols, sample characteristics, etc. If the number of characters exceeds the limit allowed for most software, then the information is split into several cells (in adjacent columns) in order to correctly read the file.

## 2.7. Source Code Availability

A stand-alone version of the program is available for download on Github (<https://github.com/gleb-gavrish/ARGEOS>, 4 October 2021). The program is available in the .py extension, which allows anyone to access the source code.

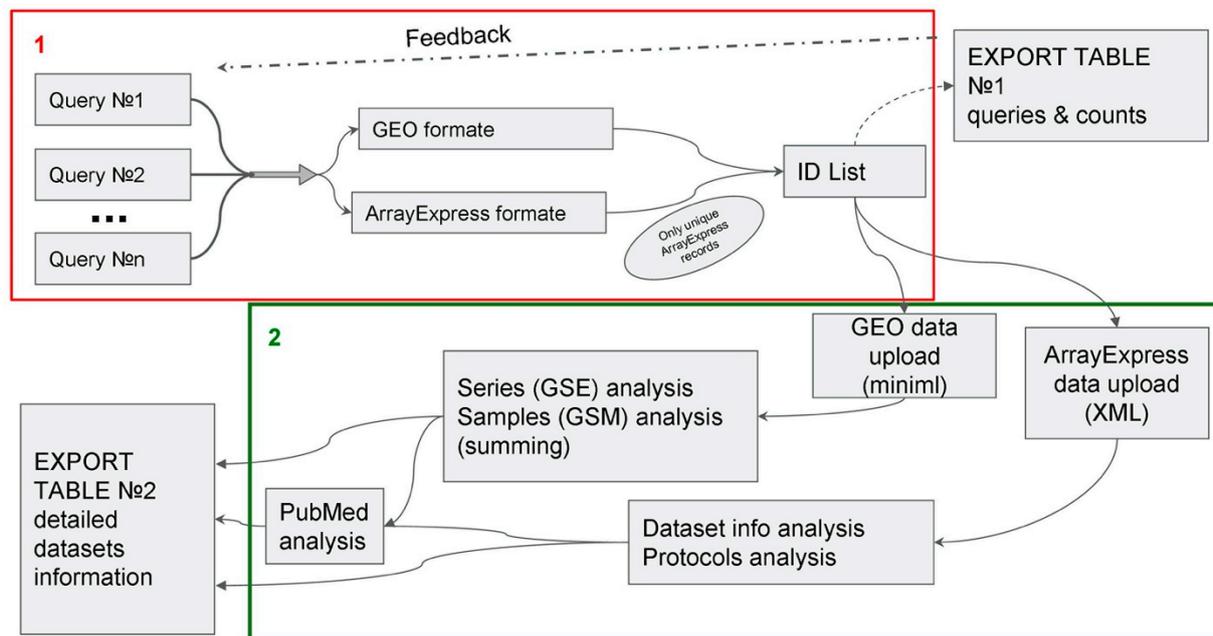
## 3. Results

### 3.1. ARGEOS Work Algorithm

For the task of systematic analysis of transcriptome data, we proposed the following scheme of the program (Figure 1). The circuit consists of two main blocks.

Block #1 Search: In this block, the user makes requests and enters the keywords. Next, the corresponding datasets are selected and filtered from the GEO and AE databases (they are compared and duplicates removed), and a list is created with the ID of all found datasets. It is important to note that the ID sheet contains data on both microarrays and RNA-seqs, methylation, and all other experiments presented in these databases. At the same time, a table (“EXPORT TABLE #1”) is saved with information about the request and the number of datasets corresponding to this request. This allows the user to adjust the search at this stage. Table S1 shows an example of the algorithm output for the queries for “polarization”, which includes 23 unique queries.

Block #2 Gathering information: Since GEO and AE have different recording formats for their databases, our program combines the data from these databases into a single form. In GEO, we receive detailed information from a file in the Miniml format (Figure 1). In parallel, we analyze information related to series (with an ID in the format GSEnnn) and samples (with an ID in the format GSMnnn). In the series section, basic information is recorded, such as the type of experiment, the number of samples, and summary. The sample section stores information about research protocols and various experiment parameters, such as cell type or incubation time. AE provides information in XML format (Figure 1). A file of this format contains all the basic information except for the research protocols. Only IDs for the corresponding protocols are indicated. By using these IDs, the texts of the protocols are separately collected, which are located in separate XML files.



**Figure 1.** Diagram of the ARGEOS web tool. There are two blocks in the ARGEOS scheme: (1) search; (2) collection of information. When performing a search, ARGEOS accepts search queries as inputs and searches the GEO and AE databases. Based on the results of the first block, a table is formed with the number of finds, as well as a list of IDs of all finds (without repetitions). The second block collects all information on each find. After bringing the information into a unified form, a final table is formed, with all the findings on request enriched with detailed information. For an explanation of the individual operations, see the main text.

The last stage of the analysis is collecting information about the PubMed article (Figure 1). In collecting basic information in the previous steps, article IDs are collected in the PubMed database. By using a call to the PubMed database, information is then collected about title of publications, year of publication of articles, impact factors of the journals, and DOIs. This ensures the same article information format for entries from GEO and AE. It is important to note that the program breaks down datasets with several organisms or experiment types into several rows (e.g., samples obtained on several organisms or in the course of experiments of different types). This makes it convenient to filter records even if there are multiple values in the original field. This feature is enabled by default but can also be disabled with the appropriate flag.

After block 2 is completed, “EXPORT TABLE #2” is created. The table includes the following data: Accession—ID of a dataset from GEO or AE; Organism—an organism or organisms (if several); Samples—number of samples; Type—the type or types of the experiment (chips, pH-sec, etc.); Platform—ID of the platform on which the experiment was carried out; Title—the name of the dataset; Year—the year the dataset was published; Summary—abstract of the dataset specified by the authors; Link—link to GEO or AE; Paper\_title—the title of the article; Journal—the journal in which the article is published; Impact factor 2018—impact factor; DOI or PubMed ID—link to the article; All references—information on all articles (if there are several); Type of molecule—a type of detected molecules; BioProject link (NCBI)—link to BioProject (page in the NCBI database); BioProject link (EBI)—link to BioProject (page in EBI database); SRA—link to record in SRA; All protocols—all protocols specified by the authors. An example of a complete table with the output of the algorithm for queries “polarization” is given in Table S2.

### 3.2. Graphical Web Interface

In order to simplify the user’s work with tools, we have developed a web interface for the program, which is available at [www.ar-geos.org](http://www.ar-geos.org), 4 October 2021.



For secondary analysis of transcriptomes, we chose RNA-seq experiments. Since the peculiarities of the GEO database are that individual samples are recorded as a new dataset when combined with a new ID (the so-called “SuperSeries”), we also filtered them. This resulted in a table with 472 datasets (Supplementary Table S3).

Next, we supplemented the table with a markup made by using manual binary classification (0 if there is no feature; 1 if there is a feature). This allowed us to mark the following features of interest: cell types (macrophages, microglia, astrocytes, and others), stimuli (LPS, IL-4, INF $\gamma$ , TNF $\alpha$ , etc.), and type of experiment (in vitro, in vivo, and single-cell). Table S3 includes all RNA-seq datasets devoted to polarization and contains information about the experimental parameters, which allows for targeted comparison of specific transcriptomes to obtain adequate valid data. Thus, using the ARGEOS program and the additional processing carried out, it was possible to create a local database to study polarization processes. Data on the ratio of dataset parameters are shown in Figure 3.

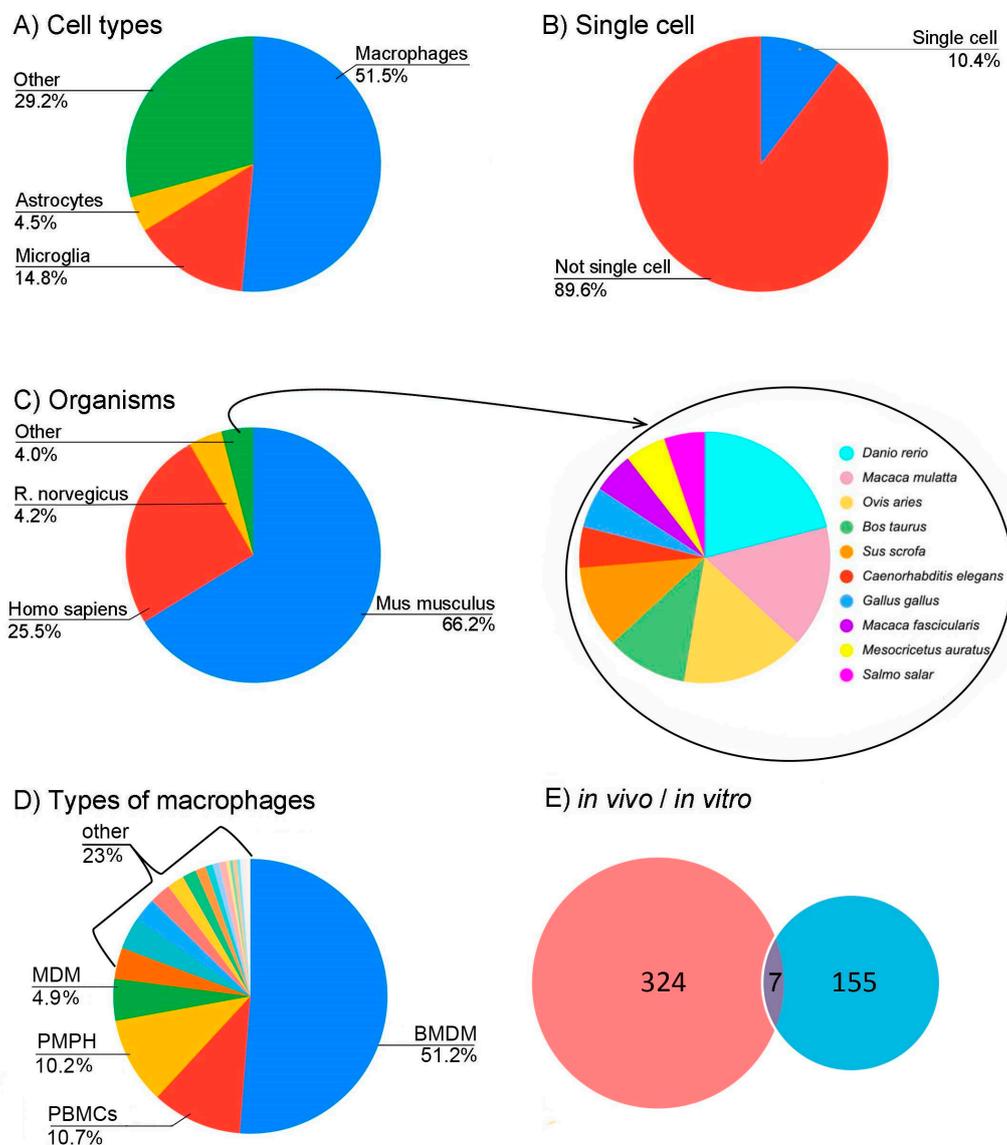
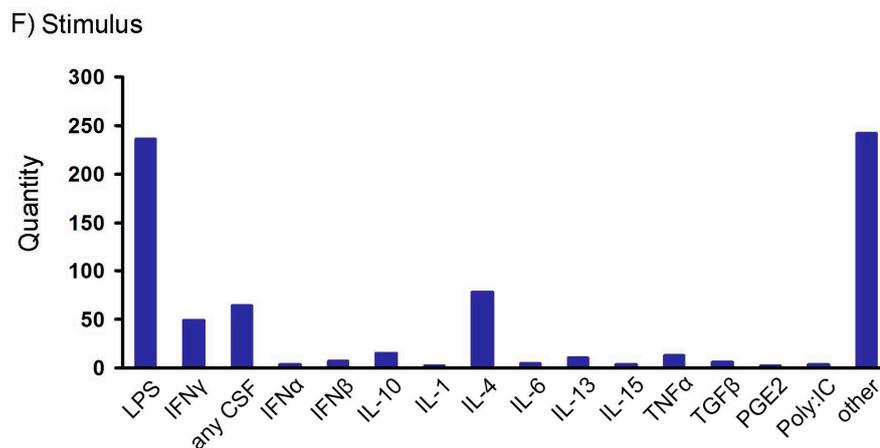


Figure 3. Cont.



**Figure 3.** Analysis of the obtained local database for queries related to polarization. Distribution of datasets (A) by the type of cells on which the studies were carried out; (B) by data on individual cells or populations; (C) by the organism from which the cells were obtained; (D) by methods for isolating macrophages (their subtypes); (E) by whether stimulation took place in vivo or in vitro; and (F) by stimuli used to change the cellular phenotype. Abbreviations: BMDM—Bone-marrow-derived macrophages; PBMCs—peripheral blood monocytes cells; PMPH—peritoneal macrophages; MDM—monocytes derived macrophages; LPS—lipopolysaccharide; IFN—interferon; CSF—colony-stimulating factor; IL—interleukin; TNF $\alpha$ —tumor necrosis factor alpha; TGF $\beta$ —transforming growth factor beta; PGE2—prostaglandin E2; Poly:IC—polyinosinic-polycytidylic acid.

Although polarization has been described for different cell types [32], the attention of researchers is focused on macrophages (51.5%) (Figure 3A). Some studies (10.4%) were carried out on individual cells (Figure 3B). In total, studies were carried out on 13 organisms, most often *Mus musculus* (66.2%) and then *Homo sapiens* (25.5%) (Figure 3C). When working with cell cultures for planning studies and analyzing results, it is informative to distinguish between obtaining cells. In total, we identified 21 different types of macrophages. Most studies were carried out on bone marrow monocyte-derived macrophages (BMDM; 51.2%) (Figure 3D). Datasets numbering 317 refer to in vitro experiments, 148 to in vivo, and both types of data are available for seven datasets (Figure 3E). We also analyzed the stimuli used to obtain different polarization studies' phenotypes (Figure 3F). The local database contains the distribution for all possible incentives, i.e., substances with which they acted on cells or organisms by injection. Other influences (such as hypoxia) and various miRNAs and siRNAs belong to the other field (Figure 3F).

Thus, the analysis of the local database data built with ARGEOS allows users to receive a primary idea of the state of research in this area of interest.

#### 4. Discussion

Transcriptome analysis technologies are important systems biology methods for investigations. The two most commonly used techniques for transcriptome analysis (microarray technology and next-generation sequencing) generate a large amount of data. Most existing tools dealt with the storage and analysis of initial data. Recently, the analysis of transcriptomic data has started to be used as a preliminary search before using usual molecular biology methods such as real-time PCR or overall planning of the experimental design of research. Such new tasks raise the problem of searching transcriptomic data.

There are currently two primary public databases, GEO (<http://www.ncbi.nlm.nih.gov/geo/>, 4 October 2021) and AE (<https://www.ebi.ac.uk/arrayexpress/>, 4 October 2021). The AE database is currently being migrated to the BioStudies database [33]. The complete transition is expected in 2021, after which AE will be completely phased out. However, old and new data will be available through the BioStudies database, and the presence of two large databases will remain. To date, GEO and AE contain more than 156,000 records, which contain information on more than 4.5 million samples. This number will increase exponentially. Therefore, the ARGEOS tool is a convenient search tool, as it

allows users to consider different formats for placing source data in databases and provides much more additional information for systematic analysis.

The ARGEOS service has the following advantages:

1. Users can send several queries simultaneously, which are supposed to be used for systematic searches.
2. It is possible to analyze the number of finds for each query (this allows the user to adjust queries in advanced searches).
3. Advanced analysis of information about the dataset is available.

The service collects detailed logs, information on the number of datasets, analyzes the availability of raw data, and provides other reference information including link and journal impact factors. All this contributes to both rapid data analysis with the search for the most relevant datasets and detailed search with detailed analysis of the information of the datasets. This is the first tool with these capabilities, which surpasses the previously proposed tools of Ivliev [15] or Bono [16].

We considered that researchers of different specialties could search transcriptome data. Therefore, we made a user-friendly and straightforward web interface for ARGEOS which can be used by biologists who are not familiar with database searching.

We have previously used transcriptomic data analysis to study gliomas [34], cell cilia [35], or systematic mining of gene co-expression networks in cancer [36]. In previous years, there was no problem with data searching, and special tools were not required. There were few data, or the searches did not need to retrieve protocol details. The problem arose when investigating the molecular mechanisms of inflammatory processes on the cellular level. The inflammatory response is part of the innate immune response and accompanies many diseases such as infection, chronic inflammation, metabolic diseases, neurodegeneration, mental diseases, and cancer [37]. Polarization is a small part of modern science, but large amounts of data have accumulated in transcriptome storage databases. As it was impossible to process these data manually, we created a special ARGEOS tool for solving the problems of systematic analysis of transcriptomic data.

We also demonstrated the possibilities of using the ARGEOS tool in the example of searching and conducting primary analysis of transcriptome data in the field of studying changes in the phenotype of cells under the action of stimuli associated with activation of the innate immune system. Using the ARGEOS tool allows users to create a local database on the selected topic. However, if the local database is large enough (for example, more than 100 records), its analysis requires different approaches. Since our local database contained more than 400 records, we described the steps we took for the analysis, including the manual data processing algorithm we used. This is performed to facilitate compiling and processing local databases for any study of this type. Creating local databases appears to be a practical solution to the problem of systematic analysis of transcriptome data. Such local databases, being in the public domain, can be an additional effective source of information. For example, the local database on the polarization problem will allow researchers to re-analyze data in future, conveniently identify standing markers of polarization coordinates, and identify the critical nodes of polarization metabolism. A new time has come and, for the first time, we offer a ready-made tool and pipeline for the systematic search of datasets.

The example with polarization provided in the work allowed us to perform the following: (1) isolate macrophages as cells for studying the polarization process; (2) postpone research on single cells (due to their weak representation); (3) choose a mouse as the most studied organism; (4) choose a source of macrophages isolation (BMDM); (5) choose a couple of stimuli LPS-IL-4; (6) choose incubation time with the stimulus; and (7) select datasets for analysis. It would take a long time to carry out this analysis without using ARGEOS. Thus, our tool not only simplifies large-scale data collection but also facilitates their selection for analysis and design of further experimental work.

## 5. Conclusions

The accumulation of data in open transcriptome databases and the intensive use of transcriptome analysis for secondary research or the design of experiments have set the task of developing a special service. ARGEOS is a bioinformatic tool for a deep scan of transcriptomic datasets that contributes to systematic search with detailed analysis on the information of the datasets.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/biology10101026/s1>, Table S1: Requests, Table S2: Program output, Table S3: Polarization database.

**Author Contributions:** Conceptualization, G.E.G. and D.V.C.; software, G.E.G.; investigation, G.E.G.; resources, M.G.S.; writing—original draft preparation, G.E.G., D.V.C. and M.G.S.; project administration, M.G.S. and D.V.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** The reported study was funded by RFBR according to the research project № 19-29-01243.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are contained within the article or supplementary material.

**Acknowledgments:** The authors acknowledge the Makarich HPC cluster (A.N. Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia) for the computational resources.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Cacheaux, L.P.; Ivens, S.; David, Y.; Lakhter, A.J.; Bar-Klein, G.; Shapira, M.; Heinemann, U.; Friedman, A.; Kaufer, D. Transcriptome profiling reveals TGF- $\beta$  signaling involvement in epileptogenesis. *J. Neurosci.* **2009**, *29*, 8927–8935. [[CrossRef](#)]
2. Stockhammer, O.W.; Zakrzewska, A.; Hegedűs, Z.; Spaink, H.P.; Meijer, A.H. Transcriptome Profiling and functional analyses of the zebrafish embryonic innate immune response to Salmonella infection. *J. Immunol.* **2009**, *182*, 5641–5653. [[CrossRef](#)] [[PubMed](#)]
3. Park, J.; Shim, J.K.; Yoon, S.J.; Kim, S.H.; Chang, J.H.; Kang, S.G. Transcriptome profiling-based identification of prognostic subtypes and multi-omics signatures of glioblastoma. *Sci. Rep.* **2019**, *9*, 1–11. [[CrossRef](#)] [[PubMed](#)]
4. Strunnikova, N.V.; Maminishkis, A.; Barb, J.J.; Wang, F.; Zhi, C.; Sergeev, Y.; Chen, W.; Edwards, A.O.; Stambolian, D.; Abecasis, G.; et al. Transcriptome analysis and molecular signature of human retinal pigment epithelium. *Hum. Mol. Genet.* **2010**, *19*, 2468–2486. [[CrossRef](#)]
5. Casamassimi, A.; Federico, A.; Rienzo, M.; Esposito, S.; Ciccodicola, A. Transcriptome profiling in human diseases: New advances and perspectives. *Int. J. Mol. Sci.* **2017**, *18*, 1652. [[CrossRef](#)] [[PubMed](#)]
6. Yang, X.; Kui, L.; Tang, M.; Li, D.; Wei, K.; Chen, W.; Miao, J.; Dong, Y. High-throughput transcriptome profiling in drug and biomarker discovery. *Front. Genet.* **2020**, *11*, 19. [[CrossRef](#)] [[PubMed](#)]
7. Kodama, Y.; Mashima, J.; Kosuge, T.; Ogasawara, O. DDBJ update: The Genomic Expression Archive (GEA) for functional genomics data. *Nucleic Acids Res.* **2019**, *47*, D69–D73. [[CrossRef](#)]
8. Wang, Z.; Lachmann, A.; Ma'ayan, A. Mining data and metadata from the gene expression omnibus. *Biophys. Rev.* **2019**, *11*, 103–110. [[CrossRef](#)]
9. Gundersen, G.W.; Jones, M.R.; Rouillard, A.D.; Kou, Y.; Monteiro, C.D.; Feldmann, A.S.; Hu, K.S.; Ma'ayan, A. GEO2Enrichr: Browser extension and server app to extract gene sets from GEO and analyze them for biological functions. *Bioinformatics* **2015**, *31*, 3060–3062. [[CrossRef](#)]
10. Toro-Domínguez, D.; Martorell-Marugán, J.; López-Domínguez, R.; García-Moreno, A.; González-Rumayor, V.; Alarcón-Riquelme, M.E.; Carmona-Sáez, P. ImaGEO: Integrative gene expression meta-analysis from GEO database. *Bioinformatics* **2019**, *35*, 880–882. [[CrossRef](#)]
11. Djordjevic, D.; Tang, J.Y.S.; Chen, Y.X.; Kwan, S.L.S.; Ling, R.W.K.; Qian, G.; Woo, C.Y.Y.; Ellis, S.J.; Ho, J.W.K. Discovery of perturbation gene targets via free text metadata mining in Gene Expression Omnibus. *Comput. Biol. Chem.* **2019**, *80*, 152–158. [[CrossRef](#)] [[PubMed](#)]
12. Barrett, T.; Wilhite, S.E.; Ledoux, P.; Evangelista, C.; Kim, I.F.; Tomashevsky, M.; Marshall, K.A.; Phillippy, K.H.; Sherman, P.M.; Holko, M.; et al. NCBI GEO: Archive for functional genomics data sets—Update. *Nucleic Acids Res.* **2013**, *41*, D991–D995. [[CrossRef](#)] [[PubMed](#)]
13. Al Mahi, N.; Najafabadi, M.F.; Pilarczyk, M.; Kouril, M.; Medvedovic, M. GREIN: An interactive web platform for re-analyzing GEO RNA-seq data. *Sci. Rep.* **2019**, *9*, 1–9. [[CrossRef](#)]

14. Alaimo, S.; Di Maria, A.; Shasha, D.; Ferro, A.; Pulvirenti, A. TACITuS: Transcriptomic data collector, integrator, and selector on big data platform. *BMC Bioinform.* **2019**, *20*, 1–11. [[CrossRef](#)] [[PubMed](#)]
15. Ivliev, A.E.; 't Hoen, P.A.C.; Villerius, M.P.; den Dunnen, J.T.; Brandt, B.W. Microarray retriever: A web-based tool for searching and large scale retrieval of public microarray data. *Nucleic Acids Res.* **2008**, *36*, W327–W331. [[CrossRef](#)] [[PubMed](#)]
16. Bono, H. All of gene expression (AOE): An integrated index for public gene expression databases. *PLoS ONE* **2020**, *15*, e0227076. [[CrossRef](#)] [[PubMed](#)]
17. Mosser, D.M.; Edwards, J.P. Exploring the full spectrum of macrophage activation. *Nat. Rev. Immunol.* **2008**, *8*, 958–969. [[CrossRef](#)]
18. Lawrence, T.; Natoli, G. Transcriptional regulation of macrophage polarization: Enabling diversity with identity. *Nat. Rev. Immunol.* **2011**, *11*, 750–761. [[CrossRef](#)]
19. Atri, C.; Guerfali, F.Z.; Laouini, D. Role of human macrophage polarization in inflammation during infectious diseases. *Int. J. Mol. Sci.* **2018**, *19*, 1801. [[CrossRef](#)]
20. Chistyakov, D.V.; Gavrish, G.E.; Goriainov, S.V.; Chistyakov, V.V.; Astakhova, A.A.; Azbukina, N.V.; Sergeeva, M.G. Oxylipin profiles as functional characteristics of acute inflammatory responses in astrocytes pre-treated with IL-4, IL-10, or LPS. *Int. J. Mol. Sci.* **2020**, *21*, 1780. [[CrossRef](#)]
21. Wheeler, D.L.; Barrett, T.; Benson, D.A.; Bryant, S.H.; Canese, K.; Chetvernin, V.; Church, D.M.; DiCuccio, M.; Edgar, R.; Federhen, S.; et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2006**, *34*, D173–D180. [[CrossRef](#)] [[PubMed](#)]
22. Buchmann, J.P.; Holmes, E.C. Entrezpy: A Python library to dynamically interact with the NCBI Entrez databases. *Bioinformatics* **2019**, *35*, 4511–4514. [[CrossRef](#)] [[PubMed](#)]
23. Mackaness, G.B. Cellular resistance to infection. *J. Exp. Med.* **1962**, *116*, 381–406. [[CrossRef](#)]
24. Martinez, F.O.; Gordon, S. The M1 and M2 paradigm of macrophage activation: Time for reassessment. *F1000Prime Rep.* **2014**, *6*, 13. [[CrossRef](#)] [[PubMed](#)]
25. Jang, E.; Lee, S.; Kim, J.H.; Kim, J.H.; Seo, J.W.; Lee, W.H.; Mori, K.; Nakao, K.; Suk, K. Secreted protein lipocalin-2 promotes microglial M1 polarization. *FASEB J.* **2013**, *27*, 1176–1190. [[CrossRef](#)] [[PubMed](#)]
26. Jang, E.; Kim, J.-H.; Lee, S.; Kim, J.-H.; Seo, J.-W.; Jin, M.; Lee, M.-G.; Jang, I.-S.; Lee, W.-H.; Suk, K. Phenotypic polarization of activated astrocytes: The critical role of lipocalin-2 in the classical inflammatory activation of astrocytes. *J. Immunol.* **2013**, *191*, 5204–5219. [[CrossRef](#)] [[PubMed](#)]
27. Tarassishin, L.; Suh, H.S.; Lee, S.C. LPS and IL-1 differentially activate mouse and human astrocytes: Role of CD14. *Glia* **2014**, *62*, 999–1013. [[CrossRef](#)] [[PubMed](#)]
28. Zamanian, J.L.; Xu, L.; Foo, L.C.; Nouri, N.; Zhou, L.; Giffard, R.G.; Barres, B.A. Genomic analysis of reactive astrogliosis. *J. Neurosci.* **2012**, *32*, 6391–6410. [[CrossRef](#)]
29. Cassetta, L.; Cassol, E.; Poli, G. Macrophage polarization in health and disease. *Sci. World J.* **2011**, *11*, 2391–2402. [[CrossRef](#)]
30. Siddaway, A.P.; Wood, A.M.; Hedges, L.V. How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annu. Rev. Psychol.* **2019**, *70*, 747–770. [[CrossRef](#)]
31. Linares-Espinós, E.; Hernández, V.; Domínguez-Escrig, J.L.; Fernández-Pello, S.; Hevia, V.; Mayor, J.; Padilla-Fernández, B.; Ribal, M.J. Methodology of a systematic review. *Actas Urol. Esp.* **2018**, *42*, 499–506. [[CrossRef](#)]
32. Liu, L.R.; Liu, J.C.; Bao, J.S.; Bai, Q.Q.; Wang, G.Q. Interaction of microglia and astrocytes in the neurovascular unit. *Front. Immunol.* **2020**, *11*, 1024. [[CrossRef](#)]
33. Sarkans, U.; Füllgrabe, A.; Ali, A.; Athar, A.; Behrang, E.; Diaz, N.; Fexova, S.; George, N.; Iqbal, H.; Kurri, S.; et al. From ArrayExpress to BioStudies. *Nucleic Acids Res.* **2021**, *49*, D1502–D1506. [[CrossRef](#)]
34. Ivliev, A.E.; 't Hoen, P.A.C.; Sergeeva, M.G. Coexpression network analysis identifies transcriptional modules related to proastrocytic differentiation and sprouty signaling in glioma. *Cancer Res.* **2010**, *70*, 10060–10070. [[CrossRef](#)] [[PubMed](#)]
35. Ivliev, A.E.; 't Hoen, P.A.C.; van Roon-Mom, W.M.C.; Peters, D.J.M.; Sergeeva, M.G. Exploring the transcriptome of ciliated cells using in silico dissection of human tissues. *PLoS ONE* **2012**, *7*, e35618. [[CrossRef](#)] [[PubMed](#)]
36. Ivliev, A.E.; 't Hoen, P.A.C.; Borisevich, D.; Nikolsky, Y.; Sergeeva, M.G. Drug repositioning through systematic mining of gene coexpression networks in cancer. *PLoS ONE* **2016**, *11*, e0165059. [[CrossRef](#)]
37. Liu, J.; Cao, X. Cellular and molecular regulation of innate inflammatory responses. *Cell. Mol. Immunol.* **2016**, *13*, 711–721. [[CrossRef](#)]