UOT 575.112

DEVELOPMENT OF THE QUICKGEO WEB TOOL FOR FUNCTIONAL GENOMICS DATA ANALYSIS

ULKAR HUSEYNOVA MUSTAFAYEVA[1*], SHAIG MEHDIYEV[2] LATIFA HASANLI[1], ORKHAN MUSTAFAYEV[2]

[1]Genetic Resources Institute, Ministry of Science and Education of the Republic of Azerbaijan ,155 Azadlig Ave., AZ 1106, Baku, Azerbaijan
[2]Baku State University, 23 Academician Zahid Khalilov St., AZ 1148, Baku,Azerbaijan
ulker.huseynova97@gmail.com

The Gene Expression Omnibus (NCBI GEO) is the largest public repository for functional genomics data, including gene expression, methylation, chromatin accessibility, and transcription factor binding profiles. Despite its rich content and scientific value, many researchers—particularly those without computational training—struggle to effectively retrieve and process data from GEO due to its complex structure, fragmented metadata, and non-standardized formats. To address these challenges, we developed QuickGEO, a user-friendly web-based application that simplifies access to GEO data and automates the essential stages of data extraction and preprocessing. QuickGEO allows users to search datasets by keyword or accession number, select relevant platforms and samples, and filter based on experimental conditions such as tissue type or disease state. The system then compiles a unified, analysis-ready table containing both expression data and sample annotations. Built using Python, MySQL, and modern web technologies, QuickGEO performs all operations server-side, eliminating the need for software installation or programming skills. Unlike most existing tools, which are limited to specific platforms or analytical functions, QuickGEO supports diverse data types and provides flexible options for dataset customization and export. It is particularly suitable for educators, students, and experimental biologists who wish to explore publicly available high-throughput data without requiring advanced bioinformatics training. By lowering technical barriers and streamlining the data preparation process, QuickGEO promotes the wider reuse of functional genomics data in research and education. The tool is freely accessible at https://quickgeo.bioset.org, offering a practical solution for integrating GEO data into biomedical and life sciences research pipelines.

Keywords: *functional genomics; Gene Expression Omnibus; web tool; data visualization; QuickGEO*

INTRODUCTION

Functional genomics is a branch of molecular biology focused on understanding gene functions and their interactions. One of the key resources in this field is the Gene Expression Omnibus (NCBI GEO), an open-access repository that provides large-scale gene expression and related high-throughput data. GEO is widely used by the research community as a source of primary experimental (Barrett et al. 2009, 2011). The utility of GEO continues to grow as researchers find increasingly diverse applications for its datasets — from comparative gene expression analysis to multi-omics integration and predictive modeling (Wang et al. 2020; Wang et al. 2021). To date, GEO data have been cited in more than 1,000 scientific publications, underscoring their widespread adoption and importance. Furthermore, these data are often repurposed to address research questions far beyond the scope of the original studies, highlighting their versatility and relevance across numerous areas of molecular biology and bioinformatics. Examples of such applications include testing new computational algorithms (Bhattacharya and De 2008), characterizing functional properties of genes (Pierre et al. 2010), conducting large-scale meta-analyses across thousands of independent datasets (Liu, 2010), and identifying diagnostic protein biomarkers for various diseases (Chen et al. 2010).

https://doi.org/10.61642/202511
Available online 10 December 2025

Despite its scale and significance, NCBI GEO presents several challenges that limit its usability, particularly for non-specialists. One of the key issues is that the meta-information describing experiments is often provided as arbitrary, unstructured free text (Wang et al., 2016), which significantly complicates automated data extraction and downstream analysis. As a result, users without a background in bioinformatics frequently encounter difficulties when attempting to interpret and utilize the data. Moreover, the current architecture of GEO does not support automated retrieval of a unified table that includes both platform metadata and expression values for all samples within a single series. To conduct a meaningful analysis, users are typically required to manually extract and merge data from multiple separate components of a GEO record. This process is time-consuming, error-prone, and creates a significant barrier for researchers lacking programming experience or technical support.

Currently, a number of tools have been developed to facilitate access to GEO data, including ShinyGEO, ExAtlas, and GEO2R (Dumas et al.,2016; Han et al. 2020a; Rey et al. 2020). However, most of these tools are either designed for highly specific analytical tasks or require users to have at least basic programming skills, which limits their accessibility for a broader community of researchers. In this context, the development of a universal, user-friendly web-based tool that enables efficient searching, extraction, preprocessing, and visualization of GEO data—without the need for coding or bioinformatics expertise—is particularly relevant (Dumas et al.,2016; Gundersen et al. 2015). Such a tool would lower the entry barrier for researchers from diverse backgrounds and support the wider reuse of publicly available genomic data in both research and education.

This study presents QuickGEO, a web-based tool developed to simplify access to data stored in the NCBI GEO repository. The main objective of the tool is to enhance the accessibility of functional genomics data by automating the key stages of data extraction, preprocessing, and presentation. Designed with usability in mind, the system targets a broad audience, including researchers, educators, and students who may lack specialized training in programming or bioinformatics.

MATERIALS AND METHODS

GEO Data Architecture

The data in the NCBI GEO repository are structured into three interconnected entities: platforms (GPL), samples (GSM), and series (GSE), each assigned a unique identifier. Metadata for these entities is typically provided in unstructured, free-text form, which significantly complicates automated processing and integration. Additionally, data are available in several formats, including SOFT, MINiML, Series Matrix, and others, further increasing the complexity of data unification and standardization. The architecture of the proposed tool is implemented as a web-based application, ensuring cross-platform compatibility. This design enables users to work with GEO data regardless of the operating system or hardware specifications of their device.

Software Implementation

The system is developed as a web application using the Python programming language and the Apache 2.4 web server. Table visualization and interactive result display are facilitated by the jQuery DataTables library, while the user interface is designed using the Bootstrap 4 framework, ensuring responsiveness and ease of use. Metadata storage and keyword-based search functionalities are powered by the MySQL relational database management system. For automated data extraction and processing, custom parsers were implemented to support the interpretation of XML, JSON, and various text archive formats retrieved from GEO servers.

RESULTS AND DISCUSSION

The developed web application QuickGEO provides a complete workflow for accessing and processing functional genomics data from the NCBI GEO repository — from keyword- or ID-based search to the generation of a final, exportable data table. One of the key advantages of the system lies in its web-based architecture, which eliminates the need for local installation and does not rely on the user's computing resources. All resource-intensive operations, including data retrieval, unpacking, and preliminary processing, are executed on the server side. This design ensures high availability and performance, enabling access to GEO data from any Internet-connected device, regardless of its operating system or hardware configuration.

The platform supports the entire workflow from data discovery to export through a user-friendly and intuitive interface. Users can search for GEO series using keywords, thanks to a built-in form that interacts with a locally synchronized relational database, allowing the identification of relevant datasets even without knowledge of specific GSE identifiers. When series contain samples from multiple platforms (e.g., microarray and sequencing), QuickGEO prompts users to select the relevant platform and dynamically displays the associated fields, such as sample annotations and label characteristics.

Customization is a core feature of QuickGEO. Users can select fields such as label ID, chromosome, genomic coordinates, tissue type, and other biological attributes to tailor the data extraction process to their specific research needs. All available samples are displayed in a structured table with detailed metadata, and a built-in filtering system allows for efficient selection even in large datasets—based on variables such as donor age, tissue type, or disease status.

Once selections are made, QuickGEO automatically retrieves and processes the relevant files from the GEO repository, generating a unified output table that combines both metadata and expression values. The output is readily exportable in common formats like .csv and .json, supporting immediate use in tools such as Excel, R, or Python.

Compared to existing web-based tools such as shinyGEO (Duma et al.,2016), GEO2R (Han et al., 2020b), and ImaGEO (Petralia et al., 2020; Rahman et al., 2021), which are often limited to differential expression analysis and microarray data, QuickGEO offers a broader, more flexible alternative. Many of these tools either do not support sequencing data or require programming skills (e.g., GEOquery, which runs through the Bioconductor R framework; Gentleman et al., 2004), limiting accessibility for non-specialists.

Moreover, challenges such as fragmented metadata, non-standard formats (SOFT, MINiML), and manual processing complexity have been widely acknowledged in the literature as significant barriers to effective data reuse (Han et al., 2020a; Rahman et al., 2021). These limitations disproportionately affect researchers without computational training, impeding broad adoption of GEO-based reanalysis in biomedical studies.

QuickGEO directly addresses these issues by providing:

a. Platform-independent access to various GEO data types;
b. Automated backend processing, removing the need for installation or code;
c. User-friendly customization and filtering, which enables even non-specialists to prepare analysis-ready datasets.

These features make QuickGEO especially useful in educational settings and research laboratories, where users may lack computational expertise. By lowering the technical entry barrier, the tool opens new possibilities for re-analysis, reinterpretation, and integration of public genomic data into biomedical research and teaching.

CONCLUSION

Our web-based tool QuickGEO provides a simple, intuitive, and efficient solution for accessing functional genomics datasets stored in the NCBI GEO repository. The system supports a complete data processing workflow — from searching for series by keywords and extracting

data via unique identifiers to generating final, analysis-ready tables. All computationally intensive operations are performed on the server side, fully relieving users of the need for programming skills or experience with bioinformatics tools. The tool addresses one of the major challenges in working with GEO data — the difficulty of retrieving and structuring information due to its fragmented and non-standardized representation. In contrast to most existing solutions, QuickGEO is not restricted to specific data types or platforms, which makes it suitable for a wide range of biological applications and research scenarios. Thanks to its flexible architecture and scalability, the system can be easily adapted to support new data formats and extended with additional modules for visualization and statistical analysis. Thus, QuickGEO serves as a universal and accessible tool for researchers, educators, and students working in the fields of molecular biology, genomics, and bioinformatics. The system is freely available at: https://quickgeo.bioset.org

REFERENCES

Barret T., Troup D.B., Wilhite S.E., Ledoux P., Evangelista C., Kim I.F., Tomashevsky M., et al. NCBI GEO: Archive for Functional Genomics Data Sets-10 Years On. *Nucleic Acids Research* 39(SUPPL. 1): 1005–10.2011. doi:10.1093/nar/gkq1184.

Barret T., Troup D.B., Wilhite S.E., Ledoux P., Rudnev D., Evangelista C., Kim I.F. et al. NCBI GEO: Archive for High-Throughput Functional Genomic Data. *Nucleic Acids Research* 37(SUPPL. 1): 885–90. 2009. doi:10.1093/nar/gkn764.

Bhattacharya A., and Rajat K. De. Divisive Correlation Clustering Algorithm (DCCA) for Grouping of Genes: Detecting Varying Patterns in Expression Profiles. *Bioinformatics* 2008;24(11): 1359–66. doi:10.1093/bioinformatics/btn133.

Chen R., Sigdel T.K., Li L., Kambham N., Dudley J.T., Hsieh S.Ch., R. Klassen B. et al. Differentially Expressed RNA from Public Microarray Data Identifies Serum Protein Biomarkers for Cross-Organ Transplant Rejection and Other Conditions. *PLoS Computational Biology* 2010;6(9). doi:10.1371/journal.pcbi.1000940.

Dumas J., Gargano M.A., and Dancik G.M. ShinyGEO: A Web-Based Application for Analyzing Gene Expression Omnibus Datasets. *Bioinformatics* 2016;32(23): 3679–81. doi:10.1093/bioinformatics/btw519.

Gentleman R.C., Carey V.J., Bates D.M., Bolstad B., Dettling M., Dudoit S., Ellis B. et al. "Bioconductor: Open Software Development for Computational Biology and Bioinformatics." *Genome biology* 2004;5(10). doi:10.1186/gb-2004-5-10-r80.

Gundersen G.W., Jones M.R., Rouillard A.D., Kou Y., Monteiro C.D., Feldmann A.S., Hu, K.S. and Ma'Ayan A. GEO2Enrichr: Browser Extension and Server App to Extract Gene Sets from GEO and Analyze Them for Biological Functions. *Bioinformatics* 2015;31(18): 3060–62. doi:10.1093/bioinformatics/btv297.

Han B., Yang X., Zhang P., Zhang Y., Tu Y., He Z., Li Y. et al. DNA Methylation Biomarkers for Nasopharyngeal Carcinoma. *PloS one* 2020a;15(4): e0230524. doi:10.1371/journal.pone.0230524.

Han B., Yang X., Zhang P., Zhang Y., Tu Y., He Z., Li Y. et al DNA Methylation Biomarkers for Nasopharyngeal Carcinoma. *PLoS ONE* 2020, b;15(4): 1–16. doi:10.1371/journal.pone.0230524.

Liu S. Increasing Alternative Promoter Repertories Is Positively Associated with Differential Expression and Disease Susceptibility. *PLoS ONE* 2010, 5(3): 1–7. doi:10.1371/journal.pone.0009482.

Petralia M.C., Ciurleo R., Saraceno A., Pennisi M., Basile M.S., Fagone P., Bramanti P., Nicoletti F., and Cavalli E. Meta-Analysis of Transcriptomic Data of Dorsolateral Prefrontal Cortex and of Peripheral Blood Mononuclear Cells Identifies Altered Pathways in Schizophrenia. *Genes* 2020, 11(4). doi:10.3390/genes11040390.

Pierre M., DeHertogh B., Gaigneaux A., DeMeulder B., Berger F., Bareke E., Michiels, and Depiereux C.E. Meta-Analysis of Archived DNA Microarrays Identifies Genes Regulated by Hypoxia and Involved in a Metastatic Phenotype in Cancer Cells. *BMC Cancer* 2010:10. doi:10.1186/1471-2407-10-176.

Rahman Md R., Islam T., Nicoletti F., Petralia M.C., Ciurleo R., Fisicaro F., Pennisi M., et al. Identification of Common Pathogenetic Processes between Schizophrenia and Diabetes Mellitus by Systems Biology Analysis. *Genes* 2021;12(2): 1–14. doi:10.3390/genes12020237.

Rey R., Suaud-Chagny M.F., Dorey J.M., Teyssier J.R. and d'Amato T. Widespread Transcriptional Disruption of the MicroRNA Biogenesis Machinery in Brain and Peripheral Tissues of Individuals with Schizophrenia. *Translational Psychiatry* 2020;10(1). doi:10.1038/s41398-020-01052-5.

Wang J., Wang Y., Kong F., Han R., Song W., Chen D., Bu L. et al. Identification of a Six-Gene Prognostic Signature for Oral Squamous Cell Carcinoma. *Journal of Cellular Physiology* 2020;235(3): 3056–68. doi:10.1002/jcp.29210.

Wang J., Dai P., Zou T., Lv Y., Zhao W., Zhang X., and Zhang Y. Transcriptome Analysis of the Transdifferentiation of Canine BMSCs into Insulin Producing Cells. *BMC Genomics* 2021;22(1): 1–15. doi:10.1186/s12864-021-07426-3.

Wang Z., Monteiro C.D., Jagodnik K.M., Fernandez N.F., G.W.Gundersen, Rouillard A.D., Jenkins Sh.L. et al. Extraction and Analysis of Signatures from the Gene Expression Omnibus by the Crowd. *Nature Communications* 2016, 7 doi:10.1038/ncomms12846.

FUNKSİONAL GENOMİKA MƏLUMATLARININ TƏHLİLİ ÜÇÜN QUICKGEO VEB ALƏTİNİN İNKİŞAFI

Ülkər Hüseynova Mustafayeva[1*], Şaiq Mehdiyev[2], Lətifə Həsənli[1], Orxan Mustafayev[2]
[1]Azərbaycan Respublikası Elm və Təhsil Nazirliyi Genetik Ehtiyatlar İnstitutu
[2]Bakı Dövlət Universiteti

Gene Expression Omnibus (NCBI GEO) funksional genomika məlumatları üçün ən böyük açıq ictimai repozitoriyadır və burada gen ekspressiyası, metilasiya, xromatinə əlçatanlıq və transkripsiya faktorlarının birləşmə profilləri kimi məlumatlar toplanır. Məzmununun zənginliyinə və elmi dəyərinə baxmayaraq, xüsusilə hesablama sahəsində təlim keçməmiş tədqiqatçılar GEO məlumatlarını səmərəli şəkildə əldə etməkdə və işləməkdə çətinlik çəkirlər. Bu çətinliklərə səbəb GEO-nun mürəkkəb quruluşu, parçalanmış metadata və standartlaşdırılmamış formatlardır. Bu problemləri aradan qaldırmaq üçün biz QuickGEO adlı istifadəsi asan, veb-əsaslı bir tətbiq hazırladıq. Bu tətbiq GEO məlumatlarına çıxışı sadələşdirir və məlumatların çıxarılması və ilkin emal mərhələlərini avtomatlaşdırır. QuickGEO istifadəçilərə açar söz və ya GEO accession nömrəsinə əsasən datasetləri axtarmağa, müvafiq platforma və nümunələri seçməyə, həmçinin toxuma növü və ya xəstəlik vəziyyəti kimi eksperimental şərtlərə əsasən filtrləməyə imkan verir. Sistem nəticədə həm ekspressiya məlumatlarını, həm də nümunə anotasiya məlumatlarını birləşdirən, analizə hazır yekun cədvəl təqdim edir. Python, MySQL və müasir veb texnologiyaları ilə hazırlanmış QuickGEO bütün əməliyyatları server tərəfində həyata keçirir, bu da əlavə proqram təminatı qurulmasına və ya proqramlaşdırma bacarığına ehtiyacı aradan qaldırır. Əksər mövcud alətlərdən fərqli olaraq, hansı ki, ya müəyyən platformalarla məhdudlaşır, ya da yalnız spesifik analitik funksiyalar təklif edir, QuickGEO müxtəlif məlumat növlərini dəstəkləyir və datasetlərin fərdiləşdirilməsi və ixracı üçün çevik seçimlər təqdim edir. QuickGEO xüsusilə müəllimlər, tələbələr və bioinformatika üzrə dərin biliklərə malik olmayan eksperimental bioloqlar üçün əlverişlidir. Texniki baryerləri azaltmaqla və məlumatların hazırlanması prosesini sadələşdirməklə bu alət funksional genomika məlumatlarının tədqiqat və təhsil məqsədilə daha geniş şəkildə istifadəsini təşviq edir. Veb-səhifə https://geo.bioset.org ünvanında sərbəst şəkildə əlçatandır və GEO məlumatlarını biotibb və həyat elmləri üzrə tədqiqat proseslərinə inteqrasiya etmək üçün praktik həll təqdim edir.

***Açar sözlər**: funksional genomika; Gen İfadə Omnibusu (GEO); veb alət; məlumatların vizuallaşdırılması; QuickGEO*

РАЗРАБОТКА ВЕБ-ИНСТРУМЕНТА QUICKGEO ДЛЯ АНАЛИЗА ДАННЫХ ФУНКЦИОНАЛЬНОЙ ГЕНОМИКИ

Улькар Гусейнова Мустафаева[1]*, Шаиг Мехтиев[2], Латифа Гасанлы[1], Орхан Мустафаев[2]
[1]Институт генетических ресурсов Министерства науки и образования Азербайджанской Республики
[2]Бакинский государственный университет

Gene Expression Omnibus (NCBI GEO) – крупнейший общедоступный репозиторий данных функциональной геномики, включая профили экспрессии генов, метилирования, доступности хроматина и связывания факторов транскрипции. Несмотря на богатый контент и научную ценность GEO, многим исследователям, особенно не имеющим компьютерной подготовки, сложно эффективно извлекать и обрабатывать данные из GEO из-за его сложной структуры, фрагментированных метаданных и нестандартизированных форматов. Для решения этих проблем мы разработали QuickGEO – удобное веб-приложение, которое упрощает доступ к данным GEO и автоматизирует основные этапы извлечения и предварительной обработки данных. QuickGEO позволяет пользователям искать наборы данных по ключевому слову или идентификационному номеру номеру, выбирать релевантные платформы и образцы, а также фильтровать данные по экспериментальным условиям, таким как тип ткани или состояние заболевания. Затем система составляет единую, готовую к анализу таблицу, содержащую как данные об экспрессии, так и аннотации образцов. QuickGEO, созданный с использованием Python, MySQL и современных веб-технологий, выполняет все операции на стороне сервера, устраняя необходимость в установке программного обеспечения или навыках программирования. В отличие от большинства существующих инструментов, ограниченных определёнными платформами или аналитическими функциями, QuickGEO поддерживает различные типы данных и предоставляет гибкие возможности настройки и экспорта наборов данных. Он особенно подходит преподавателям, студентам и биологам-экспериментаторам, желающим изучать общедоступные высокопроизводительные данные без необходимости углубленного изучения биоинформатики. Снижая технические барьеры и оптимизируя процесс подготовки данных, QuickGEO способствует более широкому использованию данных функциональной геномики в исследованиях и образовании. Инструмент доступен бесплатно по адресу https://geo.bioset.org и предлагает практическое решение для интеграции данных GEO в исследовательские процессы в области биомедицины и наук о жизни.

Ключевые слова: *функциональная геномика; Gene Expression Omnibus (GEO); веб-инструмент; визуализация данных; QuickGEO*

Çapa təqdim etmişdir: İlham Şahmuradov AMEA-nın müxbir üzvü, b.e.d.
Redaksiyaya daxil olma tarixi: 25.07.2025
Təkrar işlənməyə göndərilmə tarixi: 20.08.2025
Çapa qəbul edilmə tarixi: : 24.09.2025