The Journal of Open Source Software

FELIS: web application for integrated analysis of Japan's national clinicogenomic database

Masachika Ikegami [1,2]

1 Tokyo Metropolitan Cancer and Infectious Diseases Center Komagome Hospital, Tokyo, Japan **2** Division of Cellular Signaling, National Cancer Center Research Institute, Tokyo, Japan

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor:

Submitted: 20 December 2025
Published: unpublished

Summary

Japan's national cancer genomic medicine program aggregates comprehensive genomic profiling (CGP) results and linked clinical information at the Center for Cancer Genomics and Advanced Therapeutics (C-CAT) (Kohno et al., 2022). These data create unique opportunities for large-scale, real-world clinicogenomic studies; however, their effective use depends on clinically meaningful questions formulated by domain experts, while the need to design and implement bespoke analytical pipelines remains a substantial barrier to analysis.

FELIS (Flexible Exploration for LIquid and Solid tumor clinical sequencing data) is an open-source, locally deployable web application (R/Shiny) designed for no-code interactive analysis of secondary-use C-CAT datasets. It provides a point-and-click interface for cohort construction, clinicogenomic summarization, visualization, and bias-aware outcome analyses. By lowering technical barriers while remaining compatible with secured/offline environments, FELIS helps clinicians and translational researchers iterate rapidly from a clinical question to a reproducible analysis output.

Several analytical components implemented in FELIS are based on methods previously described in peer-reviewed publications, including variant-based clustering analysis and survival modeling accounting for delayed entry and left truncation. The scientific contribution of FELIS lies in the robust software implementation, integration, and practical accessibility of these methods rather than the introduction of new statistical methodology.

Statement of need

Large clinicogenomic resources create unique opportunities for real-world evidence (RWE) generation. However, utilizing Japan's national C-CAT secondary-use data presents specific challenges. These datasets are accessed in strictly controlled, offline environments under data use agreements, which prevents the use of cloud-based analysis tools. Furthermore, clinically relevant questions in comprehensive genomic profiling (CGP) practice often require addressing specific statistical biases, such as delayed testing and left truncation, which can meaningfully distort outcome analyses if ignored (Ikegami, 2023; Tamura et al., 2023). There is a lack of accessible tools that allow domain experts to perform these complex, bias-aware analyses within the required security constraints without extensive programming skills. FELIS addresses these gaps by offering:

- **No-code cohort building** from de-identified, preprocessed tables derived from secondary-use C-CAT datasets.
- **Bias-aware survival analysis** suitable for CGP settings with delayed entry/left truncation.
- **Clinically oriented outputs** (tables and figures) designed for downstream reporting and manuscript preparation.

- **Privacy-preserving deployment** on a local workstation or institutional server (including offline/containerized setups), so sensitive data remain within the user's controlled environment.

State of the Field

The domain of cancer genomics visualization is currently supported by robust, publicly hosted platforms such as cBioPortal (Cerami et al., 2012; Gao et al., 2013) and AACR Project GENIE (The AACR Project GENIE Consortium, 2017). These tools have established the standard for exploring large-scale, open-access genomic datasets. However, they are often ill-suited for secondary-use clinical datasets governed by strict governance and privacy controls, like those from C-CAT, which typically prohibit data upload to external hosted services.

While general-purpose R/Bioconductor packages offer the statistical flexibility required for such analyses, they present a steep learning curve for clinicians and translational researchers lacking programming expertise. Furthermore, standard genomic analysis pipelines often overlook specific biases inherent to real-world evidence (RWE), such as left truncation and delayed entry, which require specialized statistical handling not typically found in off-the-shelf genomic visualization tools.

FELIS addresses this specific niche by bridging the gap between inflexible hosted portals and code-heavy statistical packages. A "build" approach was chosen over contributing to existing platforms to satisfy two critical constraints: (1) the need for a lightweight, locally deployable architecture compatible with offline secure environments, and (2) the integration of bias-aware survival analysis methods into a no-code interface. By operationalizing these specific methodological and governance requirements, FELIS provides a unique scholarly contribution that enables reproducible RWE generation in restricted clinical environments.

Software Design

FELIS was designed with three primary constraints in mind: governance-aware deployment, accessibility for non-programming users, and analytical rigor for real-world clinicogenomic research.

The software is implemented as a Shiny application. This architecture allows FELIS to leverage the extensive statistical ecosystem of R while providing a browser-based interface suitable for clinicians and translational researchers. A modular design was chosen to separate cohort definition, genomic summarization, and outcome analysis, enabling incremental extension without entangling analytical logic.

A key design trade-off was prioritizing local deployment over hosted scalability. While this limits immediate multi-user web access, it ensures compatibility with secured and offline environments required for secondary-use clinical data. Containerized deployment options further support reproducibility across heterogeneous computing systems.

Analytically, FELIS integrates established methods—such as variant-based clustering and survival models with delayed entry—into a unified workflow. The design emphasizes transparency and reproducibility over black-box automation, allowing users to understand and validate each analytical step.

Research Impact Statement

FELIS has been developed and applied in the context of Japan's national cancer genomic medicine program and has supported multiple clinicogenomic research projects using secondary-use C-CAT data. Analytical components implemented in FELIS have contributed to peer-

reviewed publications, including studies on variant-based clustering in cancer genomics and outcome analyses accounting for delayed entry and left truncation.

The software is actively used by clinicians and researchers within secured institutional environments to perform exploratory analyses, generate publication-ready figures, and support hypothesis generation for translational studies. By lowering technical barriers while maintaining analytical rigor, FELIS enables broader participation of domain experts in clinicogenomic research.

The open-source release of FELIS, together with containerized deployment options and example workflows, provides a foundation for future extensions and adoption by other groups working with governance-restricted clinicogenomic datasets.

Software description

Architecture and deployment

FELIS is distributed as an R package that launches an interactive Shiny application. It supports (i) direct installation from source and (ii) container-based deployment to promote reproducibility across heterogeneous computing environments. The application is designed to be usable in restricted networks commonly required for secondary-use clinical data.

Data inputs

FELIS operates on research-use datasets prepared from C-CAT secondary-use programs. Users load standardized, de-identified tables (e.g., patient-level clinical variables, tumor metadata, treatments, and variant-level calls) generated by local preprocessing within their authorized environment. This design keeps FELIS open source while accommodating access control and governance constraints of national clinicogenomic data.

Core functionality

FELIS provides interactive modules that cover common clinicogenomic workflows:

- **Cohort definition and stratification**: filter and intersect clinical variables (e.g., age, sex, tumor type, stage, lines of therapy) and genomic alterations (genes, variant classes, panels).
- **Genomic summaries and visualization**: alteration frequency summaries, oncoprint-style views, co-alteration exploration, and subgroup comparisons.
- **Outcome analysis**: Kaplan–Meier and regression-based survival analyses with options to handle delayed entry/left truncation in CGP settings.
- **Treatment pattern summarization**: descriptive analyses of real-world treatment sequences and therapy exposure among genomically defined subgroups.
- **Export for reporting**: download-ready plots and tables to facilitate communication with multidisciplinary teams and manuscript preparation.

The Journal of Open Source Software

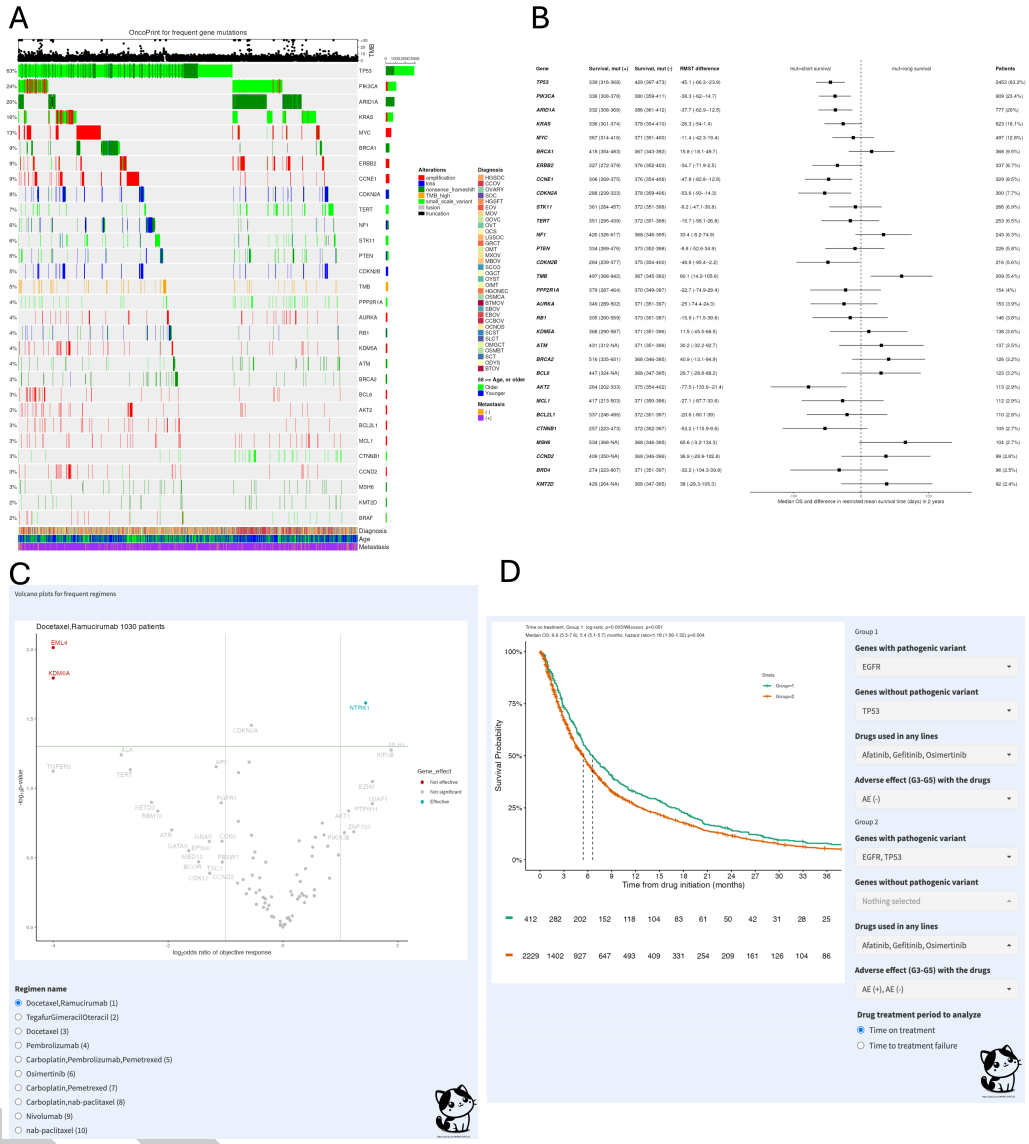

Figure 1: Representative analyses performed using FELIS. (A) OncoPrint summarizing frequently mutated genes across the selected cohort. (B) Forest plot showing the estimated effects of gene alterations on survival outcomes. (C) Volcano plot illustrating gene-level associations with drug response, highlighting effect sizes and statistical significance. (D) Kaplan–Meier survival curves comparing two groups stratified by a user-defined factor within the FELIS interface.

Some of the analytical methods implemented in FELIS have been previously reported in the literature. In particular, the variant-based clustering analysis follows the approach described by Mochizuki(Mochizuki et al., 2024), and the bias-aware survival analysis with correction for delayed entry and left truncation is based on Tamura(Tamura et al., 2023). FELIS provides a unified and reproducible software implementation of these methods tailored to the data structure and governance constraints of Japan's national clinicogenomic database (C-CAT).

AI usage disclosure

No generative AI tools were used in the development of the FELIS software or in the analysis performed by the software. Generative AI was not used to generate scientific results or figures.

The Journal of Open Source Software

Acknowledgements

We thank patients, participating hospitals, and C-CAT data governance teams for enabling secondary use of national clinicogenomic resources. We also thank collaborators and early users who provided feedback on clinical workflows and software usability.

Funding

This work was supported by grants by Boehringer Ingelheim, Daiwa Securities Foundation; Japan Cancer Association and Kobayashi Foundation for Cancer Research Young Investigator Research Grant; Japan Agency for Medical Research and Development (#25kk0305033h0001); the Japan Society for the Promotion of Science KAKENHI (22K15571 and 24K18565).

References

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A. P., Sander, C., & Schultz, N. (2012). The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discovery*, *2*(5), 401–404. https://doi.org/10.1158/2159-8290.CD-12-0095

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., & Schultz, N. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science Signaling*, *6*(269), pl1. https://doi.org/10.1126/scisignal.2004088

Ikegami, M. (2023). Letter to the editor: Left-truncation bias should be considered in prognostic analysis using national genomic profiling database. *Japanese Journal of Clinical Oncology*, *53*(11), 1091–1091. https://doi.org/10.1093/jjco/hyad098

Kohno, T., Kato, M., Kohsaka, S., Sudo, T., Tamai, I., Shiraishi, Y., Okuma, Y., Ogasawara, D., Suzuki, T., Yoshida, T., & Mano, H. (2022). C-CAT: The national datacenter for cancer genomic medicine in japan. *Cancer Discovery*, *12*(11), 2509–2515. https://doi.org/10.1158/2159-8290.CD-22-0417

Mochizuki, T., Ikegami, M., & Akiyama, T. (2024). Factors predictive of second-line chemotherapy in soft tissue sarcoma: An analysis of the national genomic profiling database. *Cancer Science*, *115*(2), 575–588. https://doi.org/10.1111/cas.16050

Tamura, T., Ikegami, M., Kanemasa, Y., Yomota, M., Furusawa, A., Otani, R., Saita, C., Yonese, I., Onishi, T., Kobayashi, H., Akiyama, T., Shimoyama, T., Aruga, T., & Yamaguchi, T. (2023). Selection bias due to delayed comprehensive genomic profiling in japan. *Cancer Science*, *114*(3), 1015–1025. https://doi.org/10.1111/cas.15651

The AACR Project GENIE Consortium. (2017). AACR project GENIE: Powering precision medicine through an international consortium. *Cancer Discovery*, *7*(8), 818–831. https://doi.org/10.1158/2159-8290.CD-17-0151