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: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished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 have accelerated discovery by enabling standardized exploration of molecular and clinical features (e.g., cBioPortal (Cerami et al., 2012; Gao et al., 2013) and AACR Project GENIE (The AACR Project GENIE Consortium, 2017)). However, C-CAT secondary-use data are typically analyzed in controlled environments under data use agreements, limiting the utility of hosted public portals. Moreover, clinically relevant real-world evidence (RWE) questions in CGP practice often require analysis features that are uncommon in general-purpose genomics portals—for example, survival modeling that accounts for delayed testing and left truncation, which can meaningfully bias CGP-based outcome analyses (Ikegami, 2023; Tamura et al., 2023). 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.

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 and governance-aware design

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

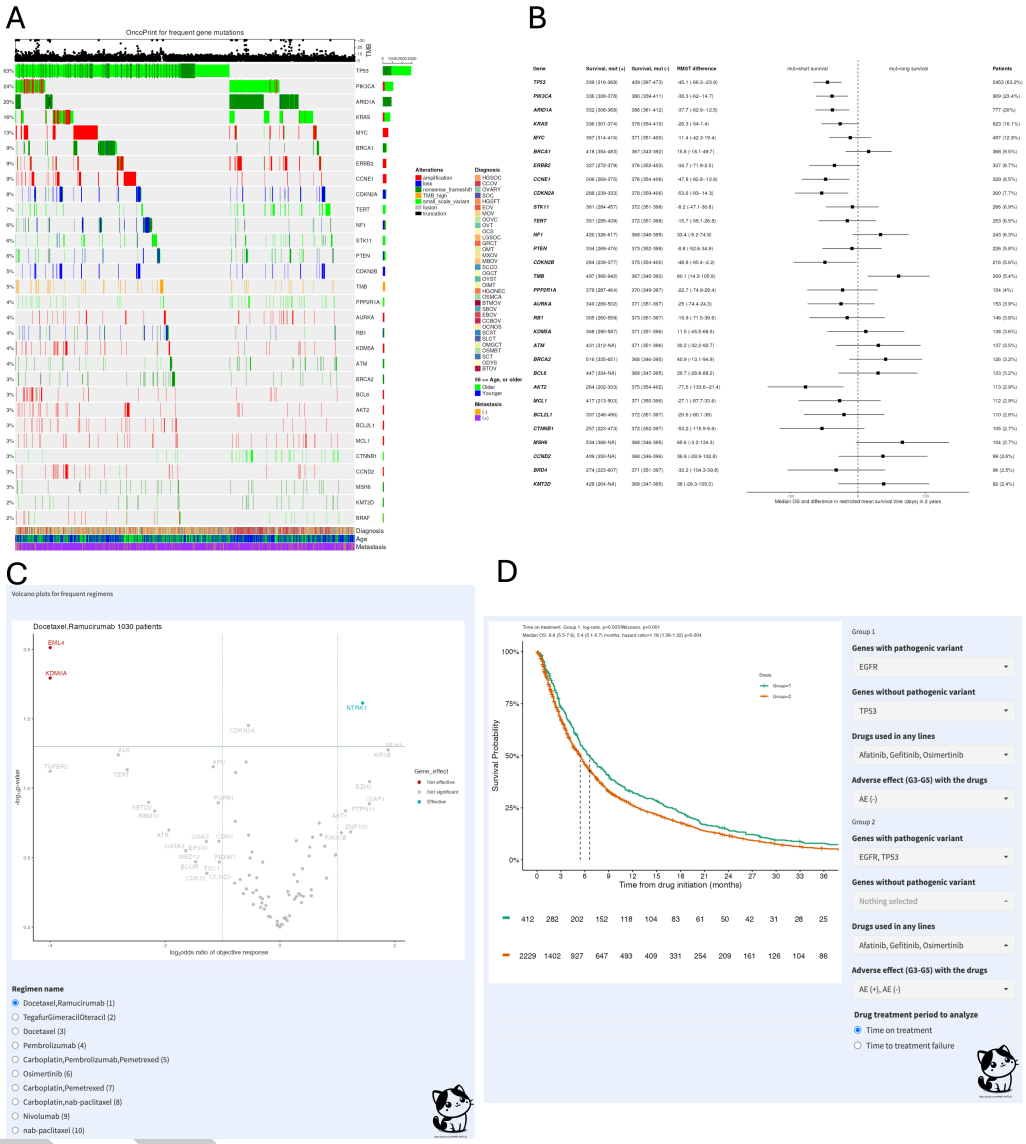

Figure 1: Representative analyses performed using FELIS. (A) OncoPrint summarizing frequently mutated genes across the selected cohort. (B) Forest plot showing the estimated effects of gene alterations on survival outcomes. (C) Volcano plot illustrating gene-level associations with drug response, highlighting 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).

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.

The Journal of Open Source Software

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