

LFSPROShiny: an interactive R/Shiny app for prediction and visualization of cancer risks in families with deleterious germline *TP53* mutations

Nam H Nguyen^{1,2}, Elissa B Dodd-Eaton¹, Gang Peng³, Jessica L. Corredor⁴, Wenwei Jiao^{1,5}, Jacynnda Woodman-Ross⁴, Banu K. Arun^{4,6}, Wenyi Wang^{1,*}

¹Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX; ²Department of Statistics, Rice University, Houston, TX; ³Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN; ⁴Department of Clinical Cancer Genetics, The University of Texas MD Anderson Cancer Center, Houston, TX; ⁵Department of Statistics, North Carolina State University, Raleigh, NC; ⁶Department of Breast Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX.

*Corresponding author: Wenyi Wang, PhD; Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, 1400 Pressler Street, Houston, TX 77030; email: wwang7@mdanderson.org.

Our research is supported by the Cancer Prevention and Research Institute of Texas [RP200383] and the National Institutes of Health [R01CA239342, P30 CA016672].

Abstract

Purpose: LFSPRO is an R library that implements risk prediction models for Li-Fraumeni syndrome (LFS), a genetic disorder characterized by deleterious germline mutations in the *TP53* gene. To facilitate the use of these models in clinics, we developed LFSPROShiny, an interactive R/Shiny interface of LFSPRO that allows genetic counselors (GCs) to perform risk predictions without any programming components, and further visualize the risk profiles of their patients to aid the decision-making process.

Methods: LFSPROShiny implements two models that have been validated on multiple LFS patient cohorts: a competing-risk model that predicts cancer-specific risks for the first primary, and a recurrent-event model that predicts the risk of a second primary tumor. Starting with a visualization template, we keep regular contact with GCs, who ran LFSPROShiny in their counseling sessions, to collect feedback and discuss potential improvement. Upon receiving the family history as input, LFSPROShiny renders the family into a pedigree, and displays the risk estimates of the family members in a tabular format. The software offers interactive overlaid side-by-side bar charts for visualization of the patients' cancer risks relative to the general population.

Results: We walk through a detailed example to illustrate how GCs can run LFSPROShiny in clinics, from data preparation to downstream analyses and interpretation of results with an emphasis on the utilities that LFSPROShiny provides to aid decision making.

Conclusion: Since Dec 2021, we have applied LFSPROShiny to over 100 families from counseling sessions at MD Anderson Cancer Center. Our study suggests that software tools with easy-to-use interfaces are crucial for the dissemination of risk prediction models in clinical settings, hence serving as a guideline for future development of similar models.

Background

Germline mutations in cancer susceptibility genes can significantly increase the likelihood of developing certain cancer types during a person's lifetime¹. Risk prediction models built on family cancer history and Mendelian transmission to assess such increased cancer risks have been widely developed²⁻⁶. A prominent example is Li-Fraumeni syndrome (LFS)⁷, a rare autosomal-dominant hereditary disorder characterized by deleterious germline mutation in the *TP53* tumor suppressor gene⁷ that significantly increases the risks of a spectrum of cancer diagnoses, most notably breast cancer, soft-tissue sarcoma and osteosarcoma^{7,8}. Hence, when assessing cancer risks for LFS patients, there exist additional challenges to quantify the risk of each cancer type separately. To address this issue, Shin et al (2019)⁹ proposed a Bayesian semi-parametric model that estimates cancer-specific age-at-onset penetrances for the first primary cancer. Since multiple primary cancers are particularly prevalent among LFS patients, Shin et al (2020)¹⁰ developed a complementary model that estimates penetrances beyond the first primary cancer, but does not differentiate between the cancer types. We will refer to the two models as the cancer-specific (CS) model and the multiple primary cancer (MPC) model. The penetrance estimates from the two models can be utilized for personalized risk predictions on unseen patients given their family history. The risk predictions have been validated on independent patient cohorts^{11,12}, and have been shown to outperform the currently used clinical criteria when predicting the carrier status of *TP53* mutations¹³, including the Classic¹⁴ and the Chompret criteria^{15,16}. We implemented the risk prediction models in a publicly available R package, called LFSPRO.

Conversations with families whose family history is suggestive of LFS regarding clinical decisions such as genetic testing and cancer screening, have been challenging. This is partly due to the fact that clinicians and genetic counselors (GCs) could only provide general, as compared to personalized, lifetime risks associated with LFS¹⁷. The good predictive performance of LFSPRO supports its potential utility to assist medical professionals with risk communication and decision-making in real-world clinical settings, hence providing a plausible solution to this bottleneck. For this purpose, it is highly desirable to have an interactive computer program that can be readily executed by the end users without dealing with the

programming components. In this paper, we present LFSPROShiny, an R/Shiny implementation of LFSPRO that provides GCs and clinicians with an easy-to-use interface, comprehensive information on patients' risk landscapes, and effective visualizations.

Methods

Overview of LFSPRO

LFSPRO is an R package that implements risk predictions based on penetrance estimates from two statistical models^{9,10}. Both models were trained and validated on independent LFS family datasets^{11,12}. The CS model⁹ was motivated by the fact that LFS patients are at increased risks of different cancer types^{7,8}, which can be regarded as competing risks. Let K be the number of competing outcomes. The CS model computed the cancer-specific age-at-onset penetrance, denoted by $q_k^{cs}(t|\mathbf{X})$. This is defined as the probability of developing cancer of type k by age t given the patient's covariates \mathbf{X} . We considered four outcomes (i.e. $K = 4$): (i) breast cancer, (ii) sarcoma, (iii) all other cancer types combined, and (iv) competing mortality. The MPC model was motivated by the frequent cases of MPC among LFS patients. The model pooled all cancer types together, but extended beyond the first primary cancer. Let L be the number of cancer occurrences. The MPC model estimated the age-at-onset penetrance $q_l^{mpc}(t|\mathbf{X}(u))$, which is defined as the probability of developing the l th primary cancer by age t given the patient's covariates and previous cancer occurrence at time u . We restricted our attention to the first and second primary cancers (i.e., $L = 2$).

LFSPRO utilizes the CS model to predict cancer-specific risks for the first primary cancer. While useful for drawing generic epidemiological insights, the penetrance estimates cannot immediately be used for risk prediction in clinics because they do not take into account the patient's current age, denoted by a_p . Let T denote the age at diagnosis of the first primary. The probability of developing the k th cancer type within the next m years, given that the patient has not developed cancer up to age a_p , is given by

$$q_k^{cs}(a_p + m|\mathbf{X}, T > a_p) = \frac{q_k^{cs}(a_p + m|\mathbf{X}) - q_k^{cs}(a_p|\mathbf{X})}{\prod_{k=1}^K [1 - q_k^{cs}(a_p|\mathbf{X})]}$$

Due to the strong detrimental effect of *TP53* mutations, one of the most important covariates is the patient's genotype ($G = 1$ for mutation carrier, $G = 0$ for wildtype). Most patients, however, have not undergone genetic testing (i.e., G is unknown). LFSPRO estimates the posterior probability that a patient carries deleterious germline *TP53* mutation given his or her family history. Let G_0 be the genotype of the patient and $\mathbf{H} = \{H_1, \dots, H_n\}$ be the history of the n family members. Following our previous study¹³, we set the prevalence of *TP53* mutations in the general population, denoted by $P[G_0 = 1]$, to be 0.0006. LFSPRO computes the posterior probabilities $P[G_0 = g|\mathbf{H}]$, $g \in \{0,1\}$, by assuming Mendel transmission. We refer the readers to **Supplementary Materials** for the detailed computations. The risk of developing the k th cancer type within the next m years, $r_k^{cs}(a_p + m|\mathbf{X}_0^G, T > a_p)$, is then given by a weighted sum

$$r_k^{cs}(a_p + m|\mathbf{X}_0^G, T > a_p) = \sum_{g \in \{0,1\}} q_k^{cs}(t|\mathbf{X}_0^G \cup \{G_0 = g\})P[G_0 = g|\mathbf{H}]$$

where \mathbf{X}_0^G denotes the covariates without genotype.

LFSPRO utilizes the MPC model to predict the risk of a first primary cancer for patients who have not had cancer, and the risk of a second primary cancer for those who have had one primary cancer. Let T_l denote the age at diagnosis of the l th primary cancer. The probability of developing the l th primary cancer in the next m years given the patient's current age a_p and the previous cancer occurrence at t_{l-1} is given by

$$q_l^{mpc}(a_p + m|\mathbf{X}(t_{l-1}), T_l > a_p) = \frac{q_l^{mpc}(a_p + m|\mathbf{X}(t_{l-1})) - q_l^{mpc}(a_p|\mathbf{X}(t_{l-1}))}{1 - q_l^{mpc}(a_p|\mathbf{X}(t_{l-1}))}$$

LFSPRO then follows similar, albeit not identical, calculations to compute the risk estimates for patients with unknown genotypes. Currently, LFSPRO is configured to report the 5-year, 10-year and 15-year risks.

Development of LFSPROShiny

Prior programming knowledge in the R statistical software (v4.2.2; R Core Team 2022) is required to install and run LFSPRO. In clinics, it is highly desirable for GCs or clinicians to have a software tool that

can be run without programming components. Furthermore, LFSPRO offers numerical output without any visualization capability. Not only does effective visualization help GCs to quickly understand the risk landscapes of their patients and make informed decisions regarding cancer screening and monitoring, but it also aids communication of cancer risks with the patients to justify such vital decisions. These shortcomings motivated us to develop LFSPROshiny, an R/Shiny implementation of LFSPRO that offers the end users an easy-to-use and interactive interface.

The development of LFSPROshiny was an iterative process. The first version offered a basic graphical user interface (GUI) and functionality. The development team was in regular contact with the GCs at MD Anderson Cancer Center (MDACC) to continually improve the application. Although most of the discussions took place in weekly meetings, the GCs also reached out to us when they encountered errors or unexpected results when running LFSPROshiny. Some changes were made to fix technical errors that resulted in incorrect risk predictions, and to improve the application's robustness to deal with a variety of datasets that did not conform to the expected input data (e.g., family members with missing ages at cancer diagnosis or ages at last contact). Usually, such changes had to be implemented in the underlying LFSPRO package. Nevertheless, most of the changes were made to add useful functionalities to the application and to improve visualization. Overlaid bar charts were added for quick comparison between a patient's predicted risks and the general population. The plots were further made interactive for easy downloading and to better show the numerical values of the bars.

The implementation of LFSPROshiny consists of: (1) a ui.R file that determines the overall GUI of the application, including the widgets (e.g., file input, radio button) and the output components (e.g., plot, table), (2) a server.R file that processes the input and displays the output on the application interface, and (3) a functions.R script, which contains user-defined functions that are invoked by server.R, in addition to the dependencies, to aid processing of the input. The dependencies include shiny, shinyjs, shinyalert, kinship2, DT, data.table, reshape2, ggplot2, ggsci and LFSPRO.

User interface of LFSPROshiny

LFSPROShiny provides a GUI that requires no programming knowledge (**Figure 1**). Following the instructions in **Figure 1A**, GCs supply two input datasets that are formatted as .csv files (**Figure 1B**). The first is a family dataset that provides information such as age-at-last-contact, gender and *TP53* mutation status (if known) of the family members, each of whom is uniquely identified by an identification number, or ID. For each family member, the dataset contains the IDs of his or her parents, which can be used by LFSPROShiny to construct a pedigree plot based on the inferred familial relationship (**Figure 1C**). The second dataset describes the cancer occurrences, including the cancer types and ages at diagnosis. The cancer types must be coded according to the LFSPRO cancer spectrum. **Table 1** provides descriptions of the columns in the input datasets. Examples are provided in the **Results** section.

[Table 1]

The output is displayed in a tabular format (**Figure 1D**), in which each record corresponds to a patient in the family. The main results of LFSPRO are the carrier probabilities for patients that have not undergone genetic testing, as well as the cancer risk predictions (as detailed below). By default, LFSPROShiny incorporates the *TP53* mutation status of family members who have been tested, but also provides an option on the left panel (**Figure 1B**) to disable this information. The MPC model is preselected to compute the carrier probabilities, but LFSPROShiny has an option for the users to choose the CS model (**Figure 1B**). A patient is predicted to be a germline *TP53* mutation carrier if the output probability is greater than or equal to 0.2. This decision threshold was shown to achieve good trade-offs between sensitivity and specificity on multiple LFS datasets¹³. If necessary, GCs can set their preferred cutoff values (**Figure 1B**). Prediction results from the Classic¹⁴ and Chompret^{15,16} criteria, which are the current standards for detection of *TP53* mutation carriers, are also provided.

For patients without cancer, LFSPROShiny invokes the CS model to provide separate probabilities for sarcoma, breast cancer, and all other cancer types combined, in the next 5, 10, and 15 years. For those who have already had one primary cancer, LFSPROShiny invokes the MPC model to calculate the probability of a second primary malignancy. LFSPROShiny plots the patient's predicted risks, along with a hypothetical noncarrier of the same age and gender, in an overlaid interactive bar chart (**Figures 1E** and

1F). Since LFS is a rare disorder¹⁸, it is reasonable to assume that the risks of noncarriers are representative of the general population.

[Figure 1]

Deployment of LFSPROShiny

LFSPROShiny is supported across all major operating systems (Windows, macOS, and Linux). The source code is available on GitHub (<https://github.com/wwylab/LFSPRO-ShinyApp>). In the repository, the users can also find links to the underlying LFSPRO package. For public use, we host the live app on Shinyapps.io, which provides a convenient way to access LFSPROShiny on the web (<https://namhnguyen.shinyapps.io/lfspro-shinyapp-master/>). For clinical use at MDACC, LFSPROShiny is deployed on a virtual server, which is part of a VMware cluster. The virtual server has 16 processor cores and 64GB of memory, and operates on the Red Hat Enterprise Linux 7 operating system. This internal hosting structure ensures sufficient security controls for protected health information, and allows for immediate input of patient data without access issues. Since Dec 2021, we have applied LFSPROShiny to over 100 families from counseling sessions at MDACC, as GCs run LFSPROShiny on the family history data right after finishing the counseling session .

Results

We provide a walk-through of how GCs can run LFSPROShiny in clinics. Due to the confidentiality of patient data, we create hypothetical family and cancer history datasets. **Figure 2 (left)** shows an example family dataset, and **Figure 2 (right)** shows the cancer history of this family. GCs may include additional columns that contain their notes regarding the patients' ages at death. These columns are based on the handwritten notes that they take when additional information becomes available after the counseling sessions. LFSPROShiny automatically extracts information from the notes to complete the missing ages at last contact. For example, the person with ID 2 is estimated to have died at age 65.

[Figure 2]

These datasets, once formatted as csv files, can be uploaded to LFSPROShiny for downstream analyses. In **Supplementary Materials**, we provide addition examples to show how LFSPROShiny

robustly handles incomplete datasets. LFSPROShiny infers the relationships among the family members based on their IDs, and graphically displays the family as a pedigree tree (**Figure 3**).

[Figure 3]

One of the most important outputs is the probability of carrying a deleterious germline *TP53* mutation for an untested family member. GCs can use these numbers to identify at-risk individuals and potentially recommend genetic testing for them. **Figure 4** shows the mutation probabilities from the CS and MPC models. The tab for each family member can be expanded to show basic information and cancer history. Using the MPC model, the predicted mutation probability for ID 5 is 0.56, whereas the CS model gives a near-zero probability. This observation can be explained by the characteristics of the two models. The CS model focuses on the first primary cancer only: it was not until 50 years old that the person developed brain cancer, which, although part of the LFS spectrum, is much less prevalent than breast cancer and sarcoma among LFS patients. Hence, there is little evidence to predict that the person has *TP53* mutation from the perspective of the CS model. On the other hand, the MPC model, which does not differentiate between cancer types, considers only the fact that the person has had two primary cancers, leading to a much higher probability. The CS model might have overlooked the second primary tumor and, consequently, underestimated the carrier probability. Had the first primary cancer be early-onset osteosarcoma, the CS model might have been the better choice since it would have perceived this cancer occurrence as a strong signal for LFS. Currently, we provide both models in LFSPROShiny for GCs to use them on a case-by-case basis. We also discuss a potential solution in the **Conclusion** section. From **Figure 3**, the proband, who has been tested positive, is an offspring of IDs 5 and 6. Given that the proband has two siblings with osteosarcoma at very young ages (IDs 9 and 12), it is highly likely that her *TP53* mutation is not de novo and that one of the parents is also a mutation carrier. Hence, in the CS model, the near-zero probability for ID 5 is accompanied by a probability that is close to 1 for ID 6. The Chompret criteria does not take into account this information and predicts wildtype for both IDs 5 and 6.

[Figure 4]

LFSPROShiny provides interactive overlaid bar charts for GCs to visualize the cancer risks of their patients relative to the general population. For ID 9, who has already had osteosarcoma at age 11, LFSPROShiny invokes the MPC model to compute the risk of a second primary tumor (**Figure 5, top left**). This individual has a very high mutation carrier probability (**Figure 4**), thus the predicted cancer risks (blue) are well above the general population (red). ID 1 has not had cancer, so LFSPROShiny invokes the CS model to compute the risks of breast cancer, sarcoma, and all other cancer types combined, for the first primary cancer (**Figure 5, top right**). This person is predicted at near-zero probability to carry a mutation (**Figure 4**), hence the predicted cancer risks (blue) are only slightly higher than the general population (red). The overall risk of developing a first primary tumor can then be computed by summing up the cancer-specific risks. Risk prediction is not available for ID 4, who already died at age 87 (**Figure 5, bottom left**). Currently, LFSPROShiny does not offer the capability to predict cancer-specific risks beyond the first primary, hence risk prediction is not available for ID 5, who has already had brain and lung cancers (**Figure 5, bottom right**).

[Figure 5]

Conclusion

In this paper, we present LFSPROShiny, a GUI software application that is built upon the previously validated LFSPRO R package to provide a much-needed interface for medical professionals to calculate, visualize and assess a patient's personalized mutation carrier probability and cancer risks based on their family history. At-risk patients can be advised accordingly to pursue genetic testing and, if appropriate, undergo the current LFS cancer screening protocols for timely detection of cancer development. Providing patients with personalized risk assessment can help patients better understand their diagnosis and may motivate them to follow the current annual cancer screening regime, which remains an intensive and sometimes overwhelming process. The application requires no programming knowledge, is supported across multiple platforms, and produces the output in seconds. Thus, GCs can access the application on their own computers to assist real-time decision making in their counseling sessions. Our study suggests

that software tools with GUI and minimum requirements of programming knowledge can improve and expedite the clinical utility of personalized cancer risk predictions.

By continually increasing the user group of GCs, LFSPROShiny will be expanded to include additional functionalities. Currently, the software can only be used to predict cancer-specific risks of the first primary cancer for patients who have never had cancer, and the risks of a second primary malignancy for those who have had one primary cancer. Risk predictions for cancer survivors are becoming increasingly important. A recent study reported that there were over 18 million living cancer survivors in the United States on January 1, 2022¹⁹. This number is projected to exceed 20 million by 2026 and 26 million by 2040²⁰. Numerous studies have confirmed the dependence of the subsequent primary tumors on the type and timing of the previous occurrences^{21,22}. Nevertheless, the current clinical management of cancer survivors is largely indifferent from at-risk individuals, thus not accounting for these relationships. To address the needs in this growing population, we will expand LFSPROShiny to include a unified model that predicts cancer-specific risks beyond the first primary cancer, recently developed by Nguyen et al (2023)²³, after it is further validated on multiple independent patient cohorts. We will also update the existing CS and MPC models as needed. Given the generality of these frameworks, they can be retrained on more complex family datasets to include additional competing risks or extend the risk predictions to the third primary cancer. Additional features, such as the type of treatment for the first primary cancer, can be included to improve the models' performance²⁴⁻²⁶. Finally, the visualization components of LFSPROShiny will continue to be updated to further advance effective risk communication in the new generation of genetic and clinical counseling.

Acknowledgement

The authors thank Dr. Seung Jun Shin and Jingxiao Chen for their contributions to LFSPRO.

Support

Cancer Prevention and Research Institute of Texas [RP200383], National Institutes of Health [R01CA239342, P30 CA016672].

Data sharing statement

The source code of LFSPROShiny and example datasets are publicly available on GitHub (<https://github.com/wwylab/LFSPRO-ShinyApp>). The live app is hosted on Shinyapps.io (<https://namhnguyen.shinyapps.io/lfspro-shinyapp-master/>). The underlying LFSPRO package can be found on GitHub (<https://github.com/wwylab/LFSPRO>).

References

1. Knudson AG: Cancer genetics. *American Journal of Medical Genetics* 111:96–102, 2002.
2. Antoniou AC, Cunningham AP, Peto J, et al: The BOADICEA model of genetic susceptibility to breast and ovarian cancers: Updates and extensions. *British Journal of Cancer* 98:1457–1466, 2008.
3. Wang W, Niendorf KB, Patel D, et al: Estimating CDKN2A carrier probability and personalizing cancer risk assessments in hereditary melanoma using MelaPRO. *Cancer Research* 70:552–559, 2010.
4. Parmigiani G, Berry DA, Aguilar O: Determining carrier probabilities for breast cancer-susceptibility genes BRCA1 and BRCA2. *American Journal of Human Genetic* 62:145-158, 1998.
5. Chen S, Wang W, Lee S, et al: Prediction of germline mutations and cancer risk in the Lynch syndrome. *JAMA* 296:1479–1487, 2006.
6. Wang W, Chen S, Brune KA, et al: PancPRO: Risk assessment for individuals with a family history of pancreatic cancer. *Journal of Clinical Oncology* 25:1417–1422, 2007.

7. Li FP, Fraumeni JF: Soft-tissue sarcomas, breast cancer, and other neoplasms. *Annals of Internal Medicine* 71:747–752, 1969.
8. Malkin D, Li FP, Strong LC, et al: Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* 250:1233–1238, 1990.
9. Shin SJ, Yuan Y, Strong LC, et al: Bayesian semiparametric estimation of Cancer-specific age-at-onset penetrance with application to Li-Fraumeni syndrome. *Journal of the American Statistical Association* 114:541–552, 2019.
10. Shin SJ, Li J, Ning J, et al: Bayesian estimation of a semiparametric recurrent event model with applications to the penetrance estimation of multiple primary cancers in Li-Fraumeni syndrome. *Biostatistics* 21:467–482, 2020.
11. Shin, S. J. et al. Penetrance estimates over time to first and second primary cancer diagnosis in families with Li-Fraumeni syndrome: A single institution perspective. *Cancer Research* 80:347–353, 2020.
12. Shin SJ, Dodd-Eaton EB, Gao F, et al: Penetrance of different cancer types in families with Li-Fraumeni syndrome: A validation study using multicenter cohorts. *Cancer Research* 80:354–360, 2020.
13. Peng G, Bojadzieva J, Ballinger ML, et al: Estimating TP53 mutation carrier probability in families with li-fraumeni syndrome using LFSPRO. *Cancer Epidemiology Biomarkers and Prevention* 26:837–844, 2017.
14. Li FP, Fraumeni JF, Mulvihill JJ, et al: A cancer family syndrome in twenty-four kindreds. *Cancer Research* 48:5358-5362, 1988.
15. Bougeard G, Renaux-Petel M, Flaman J, et al: Revisiting Li-Fraumeni syndrome from TP53 mutation carriers. *Journal of Clinical Oncology* 33:2345–2352, 2015.
16. Tinat J, Bougeard G, Baert-Desurmont S, et al: 2009 version of the Chompret criteria for Li Fraumeni syndrome. *Journal of Clinical Oncology* 27:e108-109, 2009.

17. Ross J, Bojadzieva J, Peterson S, et al: The psychosocial effects of the Li-Fraumeni Education and Early Detection (LEAD) program on individuals with Li-Fraumeni syndrome. *Genetics in Medicine* 19:1064–1071, 2017.
18. Lalloo F, Varley J, Ellis D, et al: Prediction of pathogenic mutations in patients with early-onset breast cancer by family history. *The Lancet* 361:1101–1102, 2003.
19. Miller KD, Nogueira L, Devasia T, et al: Cancer treatment and survivorship statistics, 2022. *CA: A Cancer Journal for Clinicians* 72:409–436, 2022.
20. Bluethmann SM, Mariotto AB, Rowland JH: Anticipating the ‘silver tsunami’: Prevalence trajectories and comorbidity burden among older cancer survivors in the United States. *Cancer Epidemiology Biomarkers and Prevention* 25:1029–1036, 2016.
21. Deng W, Wang Y, Liu X, et al: Assessment of trends in second primary cancers in patients with metastatic melanoma from 2005 to 2016. *JAMA Network Open* 3:e2028627, 2020.
22. Demoor-Goldschmidt C, De Vathaire F: Review of risk factors of secondary cancers among cancer survivors. *The British Journal of Radiology* 92:20180390, 2019.
23. Nguyen NH, Shin SJ, Dodd-Eaton EB, et al: Personalized risk prediction for cancer survivors: A Bayesian semi-parametric recurrent event model with competing outcomes. *bioRxiv* doi:10.1101/2023.02.28.530537, 2023.
24. Dracham CB, Shankar A, Madan R: Radiation induced secondary malignancies: A review article. *Radiation Oncology Journal* 36:85–94, 2018.
25. Xiang M, Chang DT, Pollom EL: Second cancer risk after primary cancer treatment with three-dimensional conformal, intensity-modulated, or proton beam radiation therapy. *Cancer* 126:3560–3568, 2020.
26. Bagshaw HP, Arnow KD, Trickey AW, et al: Assessment of second primary cancer risk among men receiving primary radiotherapy vs surgery for the treatment of prostate cancer. *JAMA Network Open* 5:e2223025, 2022.

http://127.0.0.1:5707 Open in browser

Predicting mutation carrier status and future cancer risk for families with Li-Fraumeni syndrome

The online tool requires two files as input: Family Data File and Cancer Data File. Please check the two links below to download the example files:

A [Family Data File](#)
A [Cancer Data File](#)

The cancer type columns in the Cancer Data File are based on the LIS spectrum of cancers that are associated with germline TP53 mutations. The package also provides functions for identifying an LIS individual based on the Classic LIS and Chromopet criteria.

After running LSPRO, there will be a figure and a table shown on the right. The figure shows the pedigree structure of the input family. The cancer history will be shown by clicking [in the table](#).

Choose Family Data File
 Browse... [fam_data.csv](#) Upload complete

Choose Cancer Data File
 Browse... [cancer_data.csv](#) Upload complete

Sample ID:

Run LSPRO

Use known testing results?
 Yes No

Which model do you want to use to calculate carrier probability?
 MPC CS

Clairif for probability: 1

Showing 1 to 10 of 22 entries

ID	Gene	TP53/PCPO	Chromopet	Classic	Breast Cancer	Sarcoma	Other Cancers	Second Primary Cancer	Figure
	Testing	carrier	criteria	criteria	years	years	years	years	
0	1	1.00	Yes	Yes	Yes				Risk trend
1		0.14	No	No	0.00	0.00	0.00	0.00	0.18
2		0.07	No	No					Risk trend
3		0.01	Yes	No					Risk trend
4		0.01	No	No					Risk trend
5		0.06	Yes	No					Risk trend
6		0.08	Yes	No					Risk trend
9		0.05	Yes	Yes				0.05	0.12
10		0.00	No	No	0.00	0.00	0.00	0.01	0.02
11		0.00	Yes	No	0.07	0.15	0.25	0.03	0.04
							0.04	0.09	0.15

Previous 1 2 3 Next

Search:

Legend: ● Affected ○ Unaffected ○ Proband

C

Cancer Risk

Sample ID: 9
 Sex: Male

Probability (%)

15Yr: Personalized cancer risk: 20.233%
 Population cancer risk: 0.15%

5Yr: Personalized cancer risk: 1.415%
 Population cancer risk: 0.15%

E

Cancer Risk

Sample ID: 1
 Sex: Male

Probability (%)

15Yr: Personalized cancer risk: 3.66%
 Population cancer risk: 4.83%

5Yr: Personalized cancer risk: 0.483%
 Population cancer risk: 0.483%

F

Figure 1: (A) Instructions for using LSPROShiny; (B) Control panel for uploading the input datasets, running LSPROShiny, and customizing the output; (C) Pedigree plot describing the relationship among family members, as constructed from the input family dataset; (D) Output describing the risk profile of each family member, including the probabilities of being TP53 mutation carrier from the CS and MPC models; (E) Cancer-specific risk prediction from the CS model for a patient with no cancer history; (F) Risk of a second primary cancer for a patient with one primary cancer occurrence in the past.

Family dataset	
id	An identification number (ID) that uniquely identifies a family member. The numbering should start from zero.
fid	ID of the father. The value is set to NA if the father is unknown.
mid	ID of the mother. The value is set to NA if the mother is unknown.
test	<i>TP53</i> mutation status. The value is set to 0 for noncarriers and 1 for mutation carriers. If the person has not undergone genetic testing, the column should be left empty.
gender	0 for female and 1 for male.
age	Age of the person at the last follow-up. The column should be left empty if the age at last contact is unknown.
vital	“A” for alive and “D” for dead.
proband	“Y” if the person is the proband (i.e., the person who receives genetic counseling from the GC) and “N” if not.
PedigreeNotes (optional)	Handwritten notes from the GCs that may contain additional information regarding the age at last contact of the person. If provided, LFSPRO will use this column to fill in the missing age information.
Cancer history dataset	
id	An identification number that uniquely identifies a family member. This should be the same as id in the family dataset.
cancer.type	The type of cancer, which must be coded according to the LFSPRO cancer spectrum: ost: osteosarcoma, sts: soft-tissue sarcoma, breast, lung, brain, leukemia, acc: adrenocortical carcinoma, and non.lfs: all other cancer types combined.
diag.age	The age at cancer diagnosis.

Table 1: Column-by-column descriptions of the input family dataset and cancer history dataset.

id	fid	mid	gender	test	age	vital	proband	PedigreeNotes1	PedigreeNotes2	PedigreeNotes3
0	5	6	0	1	24	A	Y			
1	NA	NA	1			A	N	50-70s		
2	NA	NA	0			D	N		d.60s	
3	NA	NA	1		63	D	N			
4	NA	NA	0		87	D	N			
5	1	2	1		53	A	N			
6	3	4	0			D	N	brain any		
9	5	6	1		13	A	N			
10	NA	NA	1		34	A	N			
11	5	6	0		21	A	N			
12	5	6	1		5	A	N			
13	5	6	1			A	N	40s		
14	10	11	1		2	A	N			
15	16	36	0			A	N			
16	3	4	1	1	68	D	N			
17	3	4	1		63	D	N			
18	16	36	1			A	N		18 m +	
19	16	36	1		37	A	N			
20	17	37	1		39	A	N			
21	17	37	0			A	N	18 months		
36	NA	NA	0	0	24	A	N			
37	NA	NA	0		59	A	N			

id	cancer.type	diag.age
0	ost	15
0	breast	23
5	brain	50
5	lung	50
9	ost	11
12	ost	4
14	leukemia	2
36	non.lfs	24
37	non.lfs	54

Figure 2: Family dataset with the GCs' notes in the last three columns (left) and cancer history dataset that describes the cancer occurrences among family members (right).

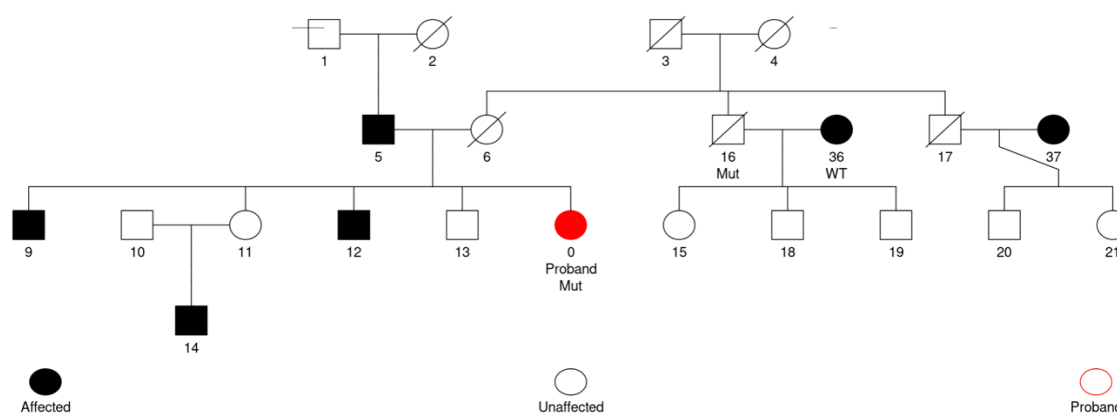


Figure 3: Pedigree tree that describes the relationship among family members. By convention, square denotes male and circle denotes female, and deceased family members are crossed with a diagonal. The proband is displayed in red. Individuals with at least one cancer occurrence (i.e., affected) is color-filled. Family members, who have undergone genetic testing, are annotated with their test results, with Mut denoting mutation carriers and WT denoting wildtypes.

Details	ID	Gene Testing	ProbLFSPRO	LFSPRO-carrier	Chompret criteria
	0	1	1.00	Yes	Yes
	1		0.14	No	No
	2		0.07	No	No
	3		0.61	Yes	No
	4		0.01	No	No
	5		0.56	Yes	No
53 years old male; brain at age 50; lung at age 50;					
	6		0.58	Yes	No
	9		0.95	Yes	Yes
	10		0.00	No	No
	11		0.90	Yes	No

Details	ID	Gene Testing	ProbLFSPRO	LFSPRO-carrier	Chompret criteria
	0	1	1.00	Yes	Yes
	1		0.00	No	No
	2		0.00	No	No
	3		0.51	Yes	No
	4		0.48	Yes	No
	5		0.01	No	No
53 years old male; brain at age 50; lung at age 50;					
	6		0.99	Yes	No
	9		0.90	Yes	Yes
	10		0.00	No	No
	11		0.72	Yes	No

Figure 4: TP53 mutation probabilities (ProbLFSPRO) produced by the MPC model (left) and the CS model (right). The binary decision (LFSPRO-carrier) is determined by the default cutoff probability of 0.2. The Shiny app allows users to modify the cutoff probability. Prediction from the Chompret criteria, which is part of the current NCCN guideline is also provided.

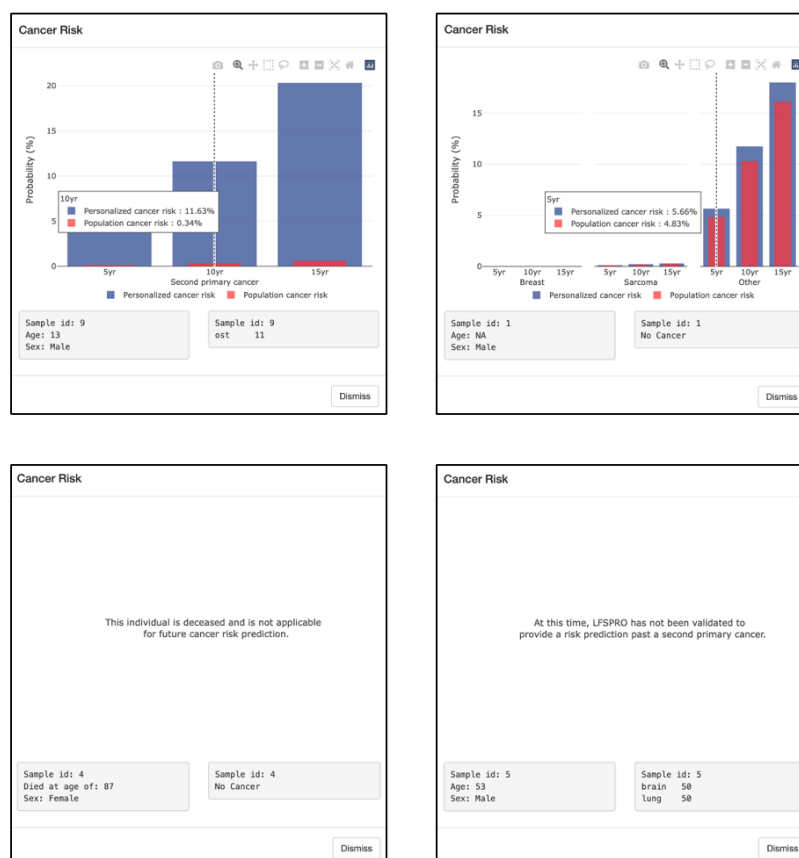


Figure 5: LFSPROShiny predicts the risks of a second primary tumor for individuals with one cancer (top left), and predicts cancer-specific risks (breast cancer, sarcoma, and all other cancer types combined) for those without cancer history (top right). The personalized risks are shown in blue, while the risks in the general population are shown in red. The predicted risks are provided for the next 5, 10 and 15 years. Risk predictions are not available for deceased individuals (bottom left), and for those who have had more than one cancer occurrence (bottom right).