RISK MODELING

Predicting cancer risk based on family history

A new software package provides more accurate cancer risk prediction profiles and has the ability to integrate more genes and cancer types in the future.

MICHELLE F JACOBS

Countless hours have been dedicated to researching cancer – how to prevent it, how to diagnose it early, and how to treat it. Yet, cancer remains a leading cause of death worldwide, accounting for almost 10 million fatalities in 2020.

Most cancers are caused by changes to genes that happen over a person's lifetime. In rarer cases (about 5–10%), they start due to inherited genetic mutations that produce a predisposition to cancer. In these instances, also known as familial or hereditary cancer syndromes, the mutation is passed down from generation to generation. In these families, more members tend to develop cancers than expected – often of the same or related type – which can also start at a particularly early age.

It is important to identify people with such genetic mutations so that they – and any family members at higher risk – can undergo enhanced cancer screening. Family history can be a useful predictor of hereditary cancer risk

(*Blackford and Parmigiani, 2010*). As such, risk prediction models that incorporate family history to estimate a person's chance of having a mutation in a cancer predisposition gene or of developing cancer have been employed for many years (*Chen et al., 2004*).

Historically, such models have been particularly valuable for deciding who to offer genetic testing to when only few and often costly genetic tests were available (*Fasching et al., 2007*). In some cases, insurance companies require the risk estimate related to carrying a cancer-related genetic mutation to exceed a certain threshold (typically 5 or 10%) to reimburse the cost of a genetic test (*Chen et al., 2006*). As research advances, the number of genes available for cancer-related genetic testing has now reached over 100 and is likely to continue increasing. Nevertheless, older risk modeling programs generally include only a small number of genes in their predictions. Now, in eLife, Danielle Braun and colleagues – including Gavin Lee and Jane Liang as joint first authors – report on a new software package that has the capacity to evolve alongside advances in cancer research (*Lee et al., 2021*).

The researchers, who are based at ETH Zürich, EPFL, Harvard, the Dana-Farber Cancer Institute, and the Broad Institute, developed PanelPRO, a tool that uses evidence gathered from extensive literature reviews to model the complex interplay between genes and cancer risk. PanelPRO's workflow consists of four main parts: input, preprocessing, algorithm, and output (*Figure 1*).
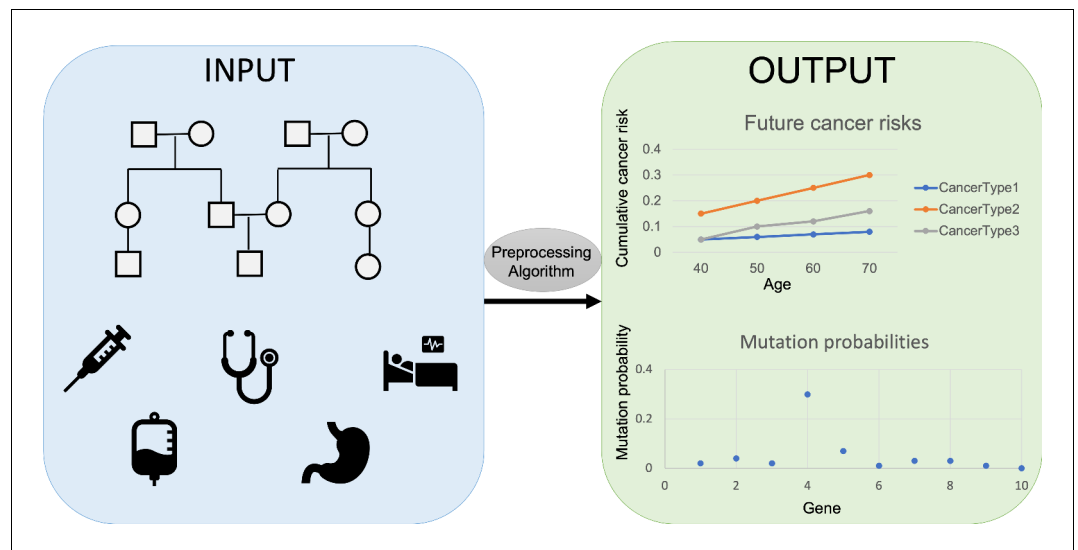
Figure 1. Workflow for PanelPRO. First, information on family history, including cancer diagnoses, age of relatives, and cancer risk factors is added into the risk modeling software PanelPRO (input, blue box on the left). Then, PanelPRO validates data formatting (preprocessing, grey oval), and analyses information about frequency and cancer risks for family cancer syndromes (algorithm, grey box) to estimate the likelihood of a person in a family having a mutation in a gene linked to an increased risk of cancer (output, green box on the right). Mutation probability and cumulative cancer risk are given as a probability between 0.0 (no risk) and 1.0 (100% risk).

The user first adds information about a history of cancers in a family – such as ages and cancer diagnoses – and other factors that might affect cancer risk. These include any risk-reducing surgeries in relatives, or tumors with biomarkers that might indicate a potential hereditary cause of their cancer. The software then adds information on the frequency of different hereditary cancer syndromes and assesses their associated cancer risks. PanelPRO can currently accommodate 18 types of cancer and generate predictions of probable mutations for 24 genes, but its code allows for the addition of new cancers or cancer-related genes that may be identified in the future.

During the preprocessing stage, the software verifies the input for any missing information and data, and also for any family relationships not supported by the software, such as 'double cousins', which occur when two siblings have children with two siblings from another family. Messages, warnings, or errors may be given to the user if any issues are detected.

After the information has been checked and modified as needed, the model proceeds to the algorithm stage. To calculate the output, the algorithm uses probabilities based on the family history, the frequency of hereditary cancer syndromes in the population, and the cancer history that would be expected if a cancer syndrome were present. The program then estimates the likelihood of a person in a family to have a mutation in a gene linked to an increased risk of cancer. These calculations can also be easily run for other family members using the existing information. It also shows a personalized estimate of future cancer risks. Users can choose which cancer types and genes to display.

However, some outstanding issues remain. Misreported family history information, such as an inaccurate cancer diagnosis or unknown age of diagnosis, can significantly affect estimates, highlighting that accuracy of patient-reported information is key to producing correct estimates (*Katki, 2006*). While patients have been shown to generally provide exact information on cancer history for first-degree relatives, the accuracy of these reports decreases for more distant relations (*Augustinsson et al., 2018*; *Murff et al., 2004*).

Moreover, analyses with a similar risk modeling software have revealed that a strict adherence to a 10% risk threshold to qualify for a test for a probable mutation in the BRCA gene (which is linked to an increased risk of developing breast, ovarian, and other cancers) would miss around 25% of individuals carrying a mutation when compared to genetic testing outcomes (*Varesco et al., 2013*). This is likely because cancer risks associated with hereditary

cancer syndromes are more variable than initially appreciated, and not all family histories may exhibit a predictable pattern of cancer, even when a mutation is present (*Okur and Chung, 2017*). This complicates risk assessments and argues against making decisions about genetic testing solely based on risk prediction models. Today, broader insurance coverage guidelines and lower costs for genetic tests have increased clinicians' ability to order these tests, even if certain risk thresholds are not met based on family history.

Nevertheless, the higher number of genes and cancer types supported by PanelPRO compared to other risk models are impressive and its ability to incorporate new genes and cancer types as testing advances are key in this fast-paced, constantly advancing field.

Michelle F Jacobs is in the Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, United States

mfjac@med.umich.edu

https://orcid.org/0000-0002-0458-1952

Competing interests: The author declares that no competing interests exist.

References

Augustinsson A, Ellberg C, Kristoffersson U, Borg Åke, Olsson H. 2018. Accuracy of self-reported family history of cancer, mutation status and tumor characteristics in patients with early onset breast cancer. *Acta Oncologica* **57**:595–603. DOI: https://doi.org/10.1080/0284186X.2017.1404635

Blackford A, Parmigiani G. 2010. Familial cancer risk assessment using BayesMendel. In: Michael F. O, John T. C (Eds). *Biomedical Informatics for Cancer Research*. Springer. p. 301–314. DOI: https://doi.org/10.1007/978-1-4419-5714-6

Chen S, Wang W, Broman KW, Katki HA, Parmigiani G. 2004. BayesMendel: an R environment for Mendelian risk prediction. *Statistical Applications in Genetics and Molecular Biology* **3**:1–19. DOI: https://doi.org/10.2202/1544-6115.1063

Chen S, Wang W, Lee S, Nafa K, Lee J, Romans K, Watson P, Gruber SB, Euhus D, Kinzler KW, Jass J, Gallinger S, Lindor NM, Casey G, Ellis N, Giardiello FM, Offit K, Parmigiani G, Colon Cancer Family Registry. 2006. Prediction of germline mutations and cancer risk in the Lynch syndrome. *JAMA* **296**:1479–1487. DOI: https://doi.org/10.1001/jama.296.12.1479, PMID: 17003396

Fasching PA, Bani MR, Nestle-Krämling C, Goecke TO, Niederacher D, Beckmann MW, Lux MP. 2007. Evaluation of mathematical models for breast cancer risk assessment in routine clinical use. *European Journal of Cancer Prevention* **16**:216–224. DOI: https://doi.org/10.1097/CEJ.0b013e32801023b3, PMID: 17415092

Katki HA. 2006. Effect of misreported family history on Mendelian mutation prediction models. *Biometrics* **62**:478–487. DOI: https://doi.org/10.1111/j.1541-0420.2005.00488.x, PMID: 16918912

Lee G, Liang JW, Zhang Q, Huang T, Choirat C, Parmigiani G, Braun D. 2021. Multi-syndrome, multi-gene risk modeling for individuals with a family history of cancer with the novel R package PanelPRO. *eLife* **10**:e68699. DOI: https://doi.org/10.7554/eLife.68699, PMID: 34406119

Murff HJ, Spigel DR, Syngal S. 2004. Does this patient have a family history of cancer? An evidence-based analysis of the accuracy of family cancer history. *JAMA* **292**:1480–1489. DOI: https://doi.org/10.1001/jama.292.12.1480, PMID: 15383520

Okur V, Chung WK. 2017. The impact of hereditary cancer gene panels on clinical care and lessons learned. *Molecular Case Studies* **3**:a002154. DOI: https://doi.org/10.1101/mcs.a002154, PMID: 29162654

Varesco L, Viassolo V, Viel A, Gismondi V, Radice P, Montagna M, Alducci E, Della Puppa L, Oliani C, Tommasi S, Caligo MA, Vivanet C, Zuradelli M, Mandich P, Tibiletti MG, Cavalli P, Lucci Cordisco E, Turchetti D, Boggiani D, Bracci R, et al. 2013. Performance of BOADICEA and BRCAPRO genetic models and of empirical criteria based on cancer family history for predicting BRCA mutation carrier probabilities: a retrospective study in a sample of Italian cancer genetics clinics. *The Breast* **22**:1130–1135. DOI: https://doi.org/10.1016/j.breast.2013.07.053, PMID: 24011770