RESEARCH ARTICLE

Genetic Epidemiology

OFFICIAL JOURNAL
INTERNATIONAL GENETIC
EPIDEMIOLOGY SOCIETY
www.geneticepi.org

WILEY

Breast and bowel cancers diagnosed in people 'too young to have cancer': A blueprint for research using family and twin studies

John L. Hopper[1] | Shuai Li[1,2,3,4] ® | Robert J. MacInnis[1,5] | James G. Dowty[1] |
Tuong L. Nguyen[1] ® | Minh Bui[1] | Gillian S. Dite[1,6] | Vivienne F. C. Esser[1] |
Zhoufeng Ye[1] ® | Enes Makalic[1] | Daniel F. Schmidt[7] |
Benjamin Goudey[8,9] ® | Karen Alpen[1] | Miroslaw Kapuscinski[1] |
Aung Ko Win[1,10,11] | Pierre-Antoine Dugué[1,4,5] | Roger L. Milne[1,4,5] |
Harindra Jayasekara[1,5] | Jennifer D. Brooks[12] ® | Sue Malta[1] |
Lucas Calais-Ferreira[1] ® | Alexander C. Campbell[1,3] | Jesse T. Young[1,12,13,14,15,16] |
Tu Nguyen-Dumont[4] | Joohon Sung[17,18,19] | Graham G. Giles[1,4,5] |
Daniel Buchanan[20] | Ingrid Winship[21] | Mary Beth Terry[22] |
Melissa C. Southey[4,5,20] | Mark A. Jenkins[1,10]

Correspondence
John L. Hopper, Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, 207 Bouverie St, Carlton, VIC 3053, Australia.
Email: j.hopper@unimelb.edu.au

Abstract
Young breast and bowel cancers (e.g., those diagnosed before age 40 or 50 years) have far greater morbidity and mortality in terms of years of life lost, and are increasing in incidence, but have been less studied. For breast and bowel cancers, the familial relative risks, and therefore the familial variances in age-specific log(incidence), are much greater at younger ages, but little of these familial variances has been explained. Studies of families and twins can address questions not easily answered by studies of unrelated individuals alone. We describe existing and emerging family and twin data that can provide special opportunities for discovery. We present designs and statistical analyses, including novel ideas such as the VALID (Variance in Age-specific Log Incidence Decomposition) model for causes of variation in risk, the DEPTH (DEPendency of association on the number of Top Hits) and other approaches to analyse genome-wide association study data, and the within-pair, ICE FALCON (Inference about Causation from Examining FAmiliaL CONfounding) and ICE CRISTAL (Inference about Causation from

For affiliations refer to page 442.

Examining Changes in Regression coefficients and Innovative STatistical AnaLysis) approaches to causation and familial confounding. Example applications to breast and colorectal cancer are presented. Motivated by the availability of the resources of the Breast and Colon Cancer Family Registries, we also present some ideas for future studies that could be applied to, and compared with, cancers diagnosed at older ages and address the challenges posed by young breast and bowel cancers.

1 | INTRODUCTION

1.1 | The problem: Breast and bowel cancers diagnosed in persons 'too young to have cancer'

Breast and bowel (colorectal) cancers are two of the most common cancers worldwide, with average ages at diagnosis in the 60s. An important minority of these cancers are diagnosed at young ages (e.g., breast cancers diagnosed before age 40 or bowel cancers before age 50) before population screening typically commences, after clinical presentation and typically at a later stage. These young cancers have far greater physical, socioeconomic and psychological impact, have higher mortality and concerningly are increasing in incidence (Acheampong et al., 2020; Araghi et al., 2019; Jayasekara et al., 2019; Kehm et al., 2019; Young et al., 2015). There is a critical need to identify those who are at greatest risk and find out what is causing these malignancies in persons 'too young to have cancer' who are, at the moment, unprotected by population screening programmes.

1.2 | Knowledge gaps

Because they are less common, little is known about the risk factors, causes and outcomes of young cancers. More information on who is at greatest risk is needed to devise cost-effective programmes for their prevention and early detection. Clearly, prevention and screening must start early in life, and young age is a critical window for cancer prevention. However, epidemiological research on breast and bowel cancer has typically studied people at the ages when most cancers are diagnosed. For example, in case–control studies usually the age of the controls is by design similar to the age at diagnosis of the cases. Cohort studies of breast and bowel cancers rarely recruit young adults; the Breast and Colon Cancer Family Registries (Breast CFR and Colon CFR) are a major exception (see Section 2).

Familial relative risks, and therefore familial variances in risk (see Section 3.1), are much larger at younger ages (Graff et al., 2017; Moller et al., 2016) and therefore provide greater opportunity for discovery of their causes. Yet we still know little. Despite the many advances in genetic research over the last three decades, it is now clear from gene panel testing that high-risk, pathogenic variants (mutations) in the 'major' genes discovered in the mid-1990s (*BRCA1*, *BRCA2* and the DNA mismatch repair genes *MLH1*, *MSH2*, *MSH6* and *PMS2*) are responsible for only a small proportion of breast and bowel cancers, even of those diagnosed at young ages (see Section 3.2). For carriers of high-risk mutations in these major genes, there are substantial differences in their risks depending on their family history (Dowty et al., 2013; Kuchenbaecker et al., 2017; Win et al., 2021), the causes of which are not understood. Although known lifestyle factors are associated with modest risk gradients over all ages combined, it is not definitively known which, or even if, they apply to risk of young cancers.

1.3 | Genetic epidemiology, studies of families and twins, and heritability

Genetic Epidemiology is about how genes and the environment combine to determine disease and health. Studies of families, and especially of twins, have historically been fundamental because they can address questions not easily answered using studies of unrelated individuals alone. This was important before the Human Genome project, when the role of genes on health in general was difficult to establish empirically.

Modern epidemiological analyses of twin and family studies go well beyond the 'nature versus nurture' debate

captured by the concept of 'heritability'. Whereas heritability of a continuous trait defined as the proportion of variance caused by genetic factors laid the foundation for variance component analysis, Fisher pointed out that absolute genetic variance was the important parameter, not as a proportion, and referred to its 'hotch-potch of a denominator' (Fisher, 1951). When it comes to risk using binary outcomes under the essentially deterministic liability model, there is not even a finite denominator; total variance in risk is unbounded because perfect risk prediction is not possible.

The concept of 'heritability' in terms of 'percentage of risk explained' is misleading for guiding genetic research. Under the essentially deterministic liability model, the relationship between the tetrachoric correlation and risk discrimination depends on the disease prevalence (Wray et al., 2010). The more common a disease is, or the older the age at which its prevalence is defined, the higher is its supposed 'heritability', despite the fact that familial relative risks decline with age.

Empirical evidence shows that a more realistic view of the way genes and environment combine to determine risk is a multiplicative model of disease incidence, as clearly demonstrated by polygenic risk scores (Mavaddat et al., 2019; Thomas et al., 2020). The maximum variation in genetic risks, and therefore the maximum risk discrimination that can be achieved, is determined by the familial risk ratio for monozygotic (MZ) twin pairs (Hopper et al., 2023).

Moreover, using a measure of a person's underlying familial or genetic risk based on their family data, we found that fundamentally the more a person is at genetic risk, the more important are their nongenetic risk factors—and vice versa (Hopper et al., 2018). That is, genetic risks are not 'immutable'—much the opposite; there is enormous prospects for reducing (modifying) risk even for people at high genetic risk.

1.4 | A pathway to new knowledge through the use of cancer family and twin data

In this paper, we describe some existing and emerging family and twin data that can provide special opportunities for discovery. For example, family studies have been the cornerstone for determining cancer risks associated with variants that are rare but likely highly pathogenic, such as protein-truncating mutations in genes like *BRCA1*, starting with linkage studies of highly selected multiple-case families (Easton et al., 1995) through to population-based families sampled irrespective of their family history (Hopper, Southey, et al., 1999). Case-

family designs in which young cases are compared with their cancer-free siblings are less prone to participation, recall and other biases (Milne et al., 2011). Cancer family studies can have extensive gene panel and family history data to stratify cases and controls (Nguyen-Dumont et al., 2020). Family designs allow for making inference about causation (Li et al., 2020); see Sections 3.4. and 3.5. Family designs also allow new approaches to be taken, such as consideration of recessive genes (Li, MacInnis, et al., 2022), family-based genome-wide association studies (GWAS), and studies of nongenomic familial risk factors such as those for breast cancer being discovered from mammograms (Hopper et al., 2020).

We have identified barriers to discovery, such as a lack of understanding of the existence, utility and power of different family study designs and a lack of awareness of the availability of resources, especially family and twin studies. We also believe from experience that both analyses and translation can benefit from there being close connections between the analytics team and the designers and collectors of the data, and with experts in the content areas. Therefore, we offer thoughts and ideas for how 21st century genetic epidemiology could create a precision population health approach to young breast and bowel cancers, with hopefully much wider implications.

2 | RESOURCES

We are motivated by existing data and resources, and what might exist in the near future, starting with extraordinary long-term family-based cohorts arising from the Breast and Colon CFRs. Both Cancer Family Registries have been funded for a further 5 years and the breast cancer initiative has a focus on recruiting the next-generation and more young cancer families.

2.1 | Breast CFR

The Breast CFR is an international cohort study of multigenerational population- and clinic-based families with and without breast cancer funded continuously by the National Institutes of Health (USA) since 1995, conducted in the United States, Canada and Australia (John et al., 2004; Neuhausen et al., 2009; Terry et al., 2016). It comprises 33,000 women and 7000 men from 15,000 families and the 20-year follow-up has just been completed. At baseline, we collected blood, biospecimens, medical records, family pedigree, cancer and epidemiology data. There are 1508 *BRCA1* and *BRCA2* mutation carriers, 12,265 breast cancers (8755

diagnosed before age 45) at enrolment with age-matched population- and family-based controls, 2250 incident breast cancers including 1135 in those unaffected at baseline. There have been comprehensive genetic and environmental risk characterisations.

2.2 | Colon CFR

Similar to the Breast CFR, the Colon CFR has been funded continuously by the National Institutes of Health (USA) since 1997. Conducted in the United States, Canada and Australasia, it comprises 23,000 women and 19,000 men from 15,000 families and the 20-year follow-up has just been completed (Jenkins et al., 2018). At baseline, we collected blood, biospecimens, medical records, family pedigree, cancer and epidemiology data. There are 2300 DNA mismatch repair gene carriers, 12,000 bowel cancers (2000 diagnosed before age 50) with age-matched population- and family-based controls, 1058 incident bowel cancers including 445 in those unaffected at baseline. This is the world's most comprehensive data and biospecimens for Lynch syndrome and young bowel cancer research.

2.3 | Harmonisation and future expansions of the Breast CFR and the Colon CFR

The questionnaires, family history, genomic and other data of the Breast CFR and Colon CFR have been harmonised; the baseline questionnaires are designed to overlap. The combined Breast CFR and Colon CFR cohorts cover the full spectrum of risks for both breast and bowel cancers; that is, the Colon CFR includes families without breast cancer and vice versa.

2.4 | Some population-complete family and twin cancer data

National genealogies cross-referenced with health care registers are being increasingly established, starting with the Swedish cancer family database (Yu & Hemminki, 2020) and more recently in Denmark (Athanasiadis et al., 2022). Similarly, algorithms have been developed to identify familial relationships from large emergency health records databases in the United States (Polubriaginof et al., 2018).

There has been a long history of population-based cancer twin studies, recently combined as the Nordic Twin Study (Harris et al., 2019). The recent FinnGen study in Finland includes genomic data, such as polygenic risk scores (Mars et al., 2022). We have described a number of new analyses, for example, using Inference about Causation from Examining FAmiliaL CONfounding (ICE FALCON) and Inference about Causation from Examining Changes in Regression coefficients and Innovative STatistical Analysis (ICE CRISTAL), that could be performed using this resource (Li & Hopper, 2023; Li et al., 2020, 2024).

The UK Biobank offers the opportunity for family studies given participants who are genetically related can be identified using extensive GWAS data (Brumpton et al., 2020).

3 | DESIGNS AND NOVEL ANALYSES

There is a long history of statistical methods to analyse family and twin data on disease incidence from a genetic epidemiological viewpoint. Complex segregation analyses based on likelihood theory underpinned linkage studies to identify risk-associated major loci and estimate penetrance without, and then with, genomic data specific to the loci under consideration. These models have been extended to include polygenic factors (Lange, 2002) and incorporated into risk prediction models such as Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) (Antoniou et al., 2008; Lee et al., 2019, 2022). A recent example combined multiple measured major genes, in effect discovered families with a very high dominantly inherited risk, which turned out to be due to a known gene, while simultaneously estimating the role of residual familial factors as represented by a factor behaving like a polygenic risk (Li, MacInnis, et al., 2022). Below we present some recently developed approaches to analyse family and genomic data.

3.1 | Variance in Age-specific Log Incidence Decomposition (VALID)

Quantifying the familial, genetic, genomic and nonfamilial (environmental) variances in risk is critical for understanding causes and making personalised risk predictions. Under a multiplicative normal risk model for a measured or unmeasured predictor, the variance in risk—with risk defined as the age-specific log(incidence)—is related solely to the risk gradient on the log incidence scale—which is equal to Δ = the difference between cases and controls in the mean of the predictor—and not to the disease prevalence (Hopper & Carlin, 1992;

Hopper et al., 2023). This model refers to risk scores—risk factors transformed, adjusted for age and other confounders and standardised—and is based on the OPERA (Odds PER Adjusted standard deviation) concept that allows risk factors (predictors) to be put into perspective with one another (Hopper, 2015).

As explained in Hopper et al. (2023), there is a simple Unifying Equation that determines the familial risk ratio (FRR) due to a risk score that has a correlation r in relatives, namely,

$$FRR = \exp(r\Delta^2). \tag{1}$$

Therefore, there is a maximum to genetic variance given by the FRR for MZ twin pairs, but contrary to what is implied using the liability model and tetrachoric correlation (Wray et al., 2010), there is no natural limit to nongenetic variance. The maximum genetic risk discrimination is dictated by the FRR for MZ twin pairs (FRR$_{MZ}$), alone. Specifically, when Φ is the cumulative normal distribution and AUC is the area under the receiver operating curve

$$\text{Maximum AUC} = \Phi\{[\log(FRR_{MZ})/2]^{1/2}\}. \tag{2}$$

3.2 | Application of VALID to breast and bowel cancer

We applied VALID to breast cancer using the Nordic Twin Study of Breast Cancer summary statistics (Möller et al., 2016), a complex segregation analysis that included our Australian multigenerational family data (Li, MacInnis, et al., 2022), and other publications on breast cancer including our mammogram risk scores (Hopper et al., 2020; Nguyen et al., 2017, 2018, 2020) to decompose the variance in risk as shown in Figure 1. The gaps in knowledge are the shaded regions.

Familial variance is highly age-dependent, and high-risk mutations in *BRCA1*, *BRCA2* and other 'major genes' explain a proportion at young—but not old—age. There is still a vast amount of variance at a young age, the causes for which are 'unknown'. In addition, a recent genome-wide exome sequencing meta-analysis found that coding variants in other genes explain little variance (Wilcox et al., 2023). The polygenic risk score explains variance at older ages but not at young ages. The mammogram risk scores are correlated by about 0.5 in MZ twin pairs, so contribute to both familial and nonfamilial variance (Nguyen et al., 2022), and are virtually independent to the polygenic risk score (Li, Nguyen, et al., 2022). The epidemiological risk factors explain some variance, but more so at older ages.

A similar story applies to bowel cancer (see Figure 2), except that: (i) the only major genes of significance are the DNA mismatch repair genes, and they only explain 'nonpolyposis' (Lynch syndrome) bowel cancer. The genes implicated in the current polygenic risk scores for bowel cancer in general do not appear to be relevant to Lynch syndrome bowel cancer (Jenkins et al., 2021;
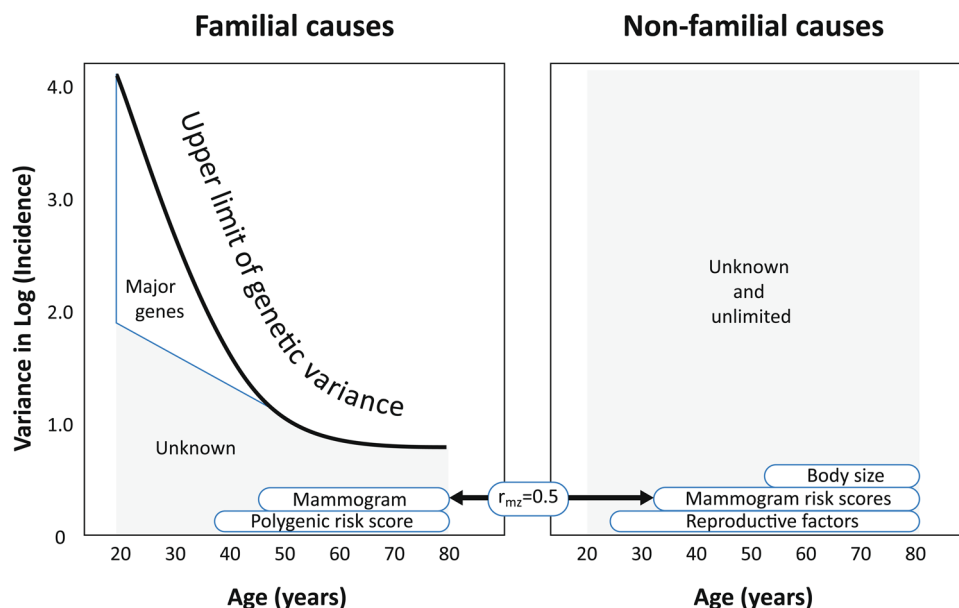

FIGURE 1 Variance in log(incidence) of breast cancer by age according to familial effects, including rare variants in major genes such as *BRCA1* and *BRCA2*, polygenic risk scores, mammogram risk scores that have a substantial familial component and some other epidemiological risk factors that are mostly nonfamilial, based on the literature (from Hopper et al., 2023).
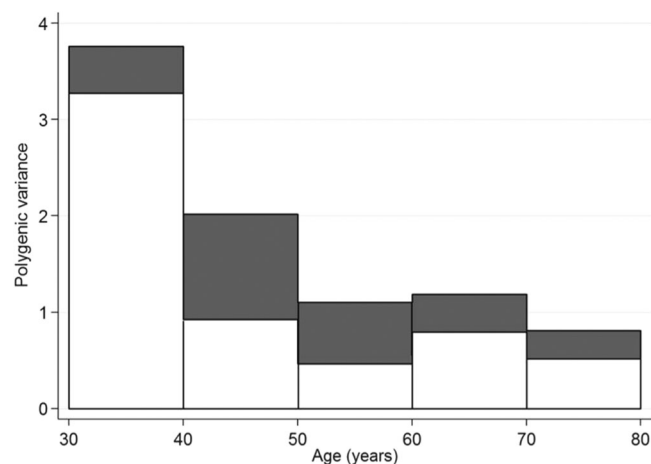
FIGURE 2 Familial variance explained by hypothetical unidentified major genes (dark grey) and the polygenic component (white) for each 10-year age group, after excluding families with pathogenic variants in known DNA mismatch repair genes (from Win et al., 2017).

Win et al., 2013). There is substantial evidence of other as yet undiscovered major genes for general bowel cancer, as we found from a large segregation analysis (Win et al., 2017). As for breast cancer, the epidemiological risk factors do not explain a substantial amount of variance in risk (Zheng et al., 2020).

3.3 | DEPendency of association on the number of Top Hits (DEPTH)

We developed the DEPTH method (Makalic et al., 2013) for analysing GWAS data that utilises the power of massively parallel computing to find genomic regions, rather than individual variants, implicated in risk. DEPTH is a discovery tool that can be applied to family data. It uses a sliding window of multiple variants to create a SONAR (Signal-Over-Noise Associated with Risk) for each overlapping region that can be interpreted as the log odds in favour of the fitted model. When plotted using a genome browser, the user can easily identify regions of interest. We have applied DEPTH to prostate cancer and identified 112 novel putative risk regions (MacInnis, Schmidt, et al., 2016).

We recently applied DEPTH to bowel cancer data from the Colon CFR (5735 cases and 3688 controls), and 8865 cases and 10,285 controls from other nonfamily studies and compared with 132 published risk regions (Lai et al., 2023). We identified 82 likely risk regions that would not be detected using conventional approaches nor had been identified in previous studies. We found four reproducible risk regions with the strongest signal from

the human leucocyte antigens (HLA) region on chromosome 6p21, implicating the immune system in causing bowel cancer.

3.4 | Within-sibship comparisons, familial confounding and causation

Within-sibship comparison designs naturally control for familial (i.e., genetic and nongenetic) factors shared by the siblings, and the extent of these familial confounders will vary across the types of siblings and their ages. However, finding a within-pair association does not imply the association is causal as it may still be prone to individual-level confounding, whereas not finding a within-pair association does not imply the association is not causal. Making inference about causation in the presence of familial confounding can be addressed by the ICE FALCON and ICE CRISTAL methods below (Sections 3.5 and 3.6).

For pairs selected due to being disease discordant, case–control analyses predict risk for controls who, by design, have a family history. Studies of disease-concordant siblings using the within-pair difference between ages at diagnosis predict risk for individuals with a strong family history, in that both have a diagnosis. Note that, using a within-sibship design, inference is being made about exposures of the magnitude of the sample within-pair differences in exposure, which will likely be less than the differences in exposure across the population for which unmatched studies are making inference. Selection of pairs most discordant for exposures (see e.g., Hopper & Seeman, 1994) will be more powerful per subject and enable more comparable comparisons of within- and between-pairs estimates.

3.5 | ICE FALCON

Causation is a pivotal issue through its aetiological relevance. It also has implications for prevention and risk reduction given that associations alone are not enough to 'confer' or 'convey' causation even if the associations are over time, within families or involve a genetic exposure. Randomised controlled trials are expensive, often cannot be conducted for ethical or practical reasons and typically take considerable time leaving the findings vulnerable to being out of date. Therefore, there is a continual reliance on observational data to assess the evidence for causation, especially and even if it is longitudinal.

We developed a powerful way to make inference about causation from family data (ICE FALCON) (Li et al., 2020). The key issues are that ICE FALCON: (i)

applies to data on pairs of related individuals who are correlated in their exposure, (ii) fits regression models, estimating within-individual and across pair coefficients independently and together and (iii) makes inference by comparing the changes in the pairs of regression coefficients.

Briefly, the data are consistent with causation if the cross-pair coefficient attenuates to the null after including a within-individual coefficient, but not vice versa. The data are consistent with familial confounding if both the cross-pair coefficient and within-individual coefficients attenuate to the same extent after including one another. Reverse causation can be tested in several ways, such as reversing the roles of the presumed cause and outcome in the regression model (Li et al., 2020). These scenarios operate irrespective of individual-specific confounders.

We have applied ICE FALCON to the atopic march (Hopper et al., 2012), mammographic density (Dite et al., 2008; Stone et al., 2012; Ye et al., 2023), bone architecture (Bui et al., 2013, 2020; Nissen et al., 2023) and DNA methylation (Li et al., 2018, 2019, 2020) and found causal evidence for these exposures. We demonstrated that ICE FALCON can make the same causal conclusion as Mendelian randomisation using much smaller data sets; for example, 66 MZ twin pairs versus 4000 individuals to find that body mass index (BMI) causes DNA methylation, not vice versa (Li et al., 2020). ICE FALCON can be applied to prospective exposure-outcome data and can be used to test if the tracking of a risk factor with time is causal and therefore justify intervention studies. This issue cannot be readily addressed using Mendelian randomisation as we found for BMI.

3.6 | ICE CRISTAL

ICE CRISTAL applies to data for unrelated individuals and so is more widely applicable. It is similar to ICE FALCON but replaces the exposure of the relative with the outcome of the relative for the family history (however defined) (Li & Hopper, 2023; Li et al., 2024). In the context of breast and colorectal cancer, ICE CRISTAL considers the extent to which family history is explained (statistically) by including a (familial) risk factor. Again, inference is made by studying the changes in the pair of regression coefficients. Observing that the exposure coefficient does not change after fitting family history, but that the family history coefficient decreases after fitting the exposure coefficient, is consistent with the exposure having a causal effect, as we observed for mammographic density and breast cancer (Martin

et al., 2010), and for polygenic risk scores and breast cancer (Li et al., 2024). Familial confounding is consistent with both the exposure and family history coefficients decreasing when estimated together. Statistical inference about the change in a regression coefficient can be made using the methodology attributed to Freedman et al. (1992) that is referenced in Martin et al. (2010).

4 | POTENTIAL APPLICATIONS

We give some examples of genetic and epidemiological studies that could utilise family or twin data to specifically research breast and bowel cancers being diagnosed at a young age.

4.1 | To discover factors in early life that likely cause breast or bowel cancer or impact on outcomes for those diagnosed with these cancers

Baseline and follow-up data from the family and twin cohorts could be used to conduct further case–control and cohort studies, including case–sibling analyses that control for shared familial factors and are less prone to participation bias (Milne et al., 2011). Analyses could focus on early-life exposures including height, weight at age 18–20, waist-to-hip measurements, physical exercise, radiation, use of medications including oral contraceptives and nonsteroid anti-inflammatory drugs and the reproductive history of females. Creative ways of combining questionnaire data to create new variables with aetiologic relevance, such as the Pike model for the cumulative effect of ovarian hormones (Pike et al., 1983) and reworked to create a risk score for age (Nguyen et al., 2020), could be explored. Interaction terms can be fitted to test if risk associations differ with age at diagnosis or family history (Hopper et al., 2018; Ye et al., 2022).

Timing of exposure could be important. For example, oral contraceptive (OC) use is generally considered to have a short-term effect on breast cancer risk that dissipates within 5 years of last use. OC use has a protective association with ovarian cancer, so underlying genetic risk is important given the high risk of ovarian cancer for *BRCA1* carriers (MacInnis, Pike, et al., 2016). Using the Breast CFR, we found strong evidence that OC use was associated with a substantial (4.5-fold; $p = 0.0001$) protective association against breast cancer diagnosed before age 40 for *BRCA1* carriers, but not for *BRCA2* carriers or noncarriers (Milne et al., 2005). The results published to date for mutation carriers are

generally null but the literature is almost entirely from studies of women over the age of 40 years; despite *BRCA1* mutation carriers having already reached their peak incidence of 2% per year in their 30s (Kuchenbaecker et al., 2017). Given our gene panel testing that includes cases, controls and relatives younger than age 40 years, these analyses could be revisited with more power and certainty to try to resolve these timing of exposure issues for all risk factors.

For outcomes, such as all-cause or disease-specific mortality, time since diagnosis would be a natural time scale at least for population-based incident samples. Figure 3 shows results from fitting flexible parametric survival analyses stratified by type of cancer and age at diagnosis, as we did for oestrogen receptor status and mortality following breast cancer (Jayasekara et al., 2019). Interaction terms for age at diagnosis could be fitted to quantify the extent to which associations depend on age at diagnosis and other measured covariates.

Opposite-sex twin pairs allow for within-pair analyses that address sex differences by controlling for age, on average 50% of autosomal genes and shared environment including shared placenta in utero. This avenue of investigation could be informative for bowel cancer.

4.2 | To identify new genetic causes and modifiers of risk for breast and bowel cancers

DEPTH (see Section 3.3) and other new analytic techniques could be systematically used to analyse the GWAS data of the Breast Cancer Association Consortium of 137,000 cases and 119,000 controls (Michailidou et al., 2017) and the Colon CFR and associated bowel cancer studies (COloRECTal Cancer Transdisciplinary study/Genetic Epidemiology of Colorectal Cancer cOhort [CORECT/GECCO]), which currently comprise 60,000 bowel cancers and 70,000 controls (Cheng et al., 2014; Lai et al., 2023; Mahmood et al., 2023). We have developed and run a pipeline to harmonise the different genome-wide array data across the world's major breast and bowel cancer GWASs to the Human Genome Version 19 (hg19/GRCh37 genome build). Findings could be integrated to allow biological contextualisation using, for example, Ingenuity Pathway Analysis software (Kramer et al., 2014). Candidate genes could be subjected to Core Analysis to provide a measure of the overlap between each gene set and three Gene Ontology terms: Molecular Function, Biological Process and Canonical.
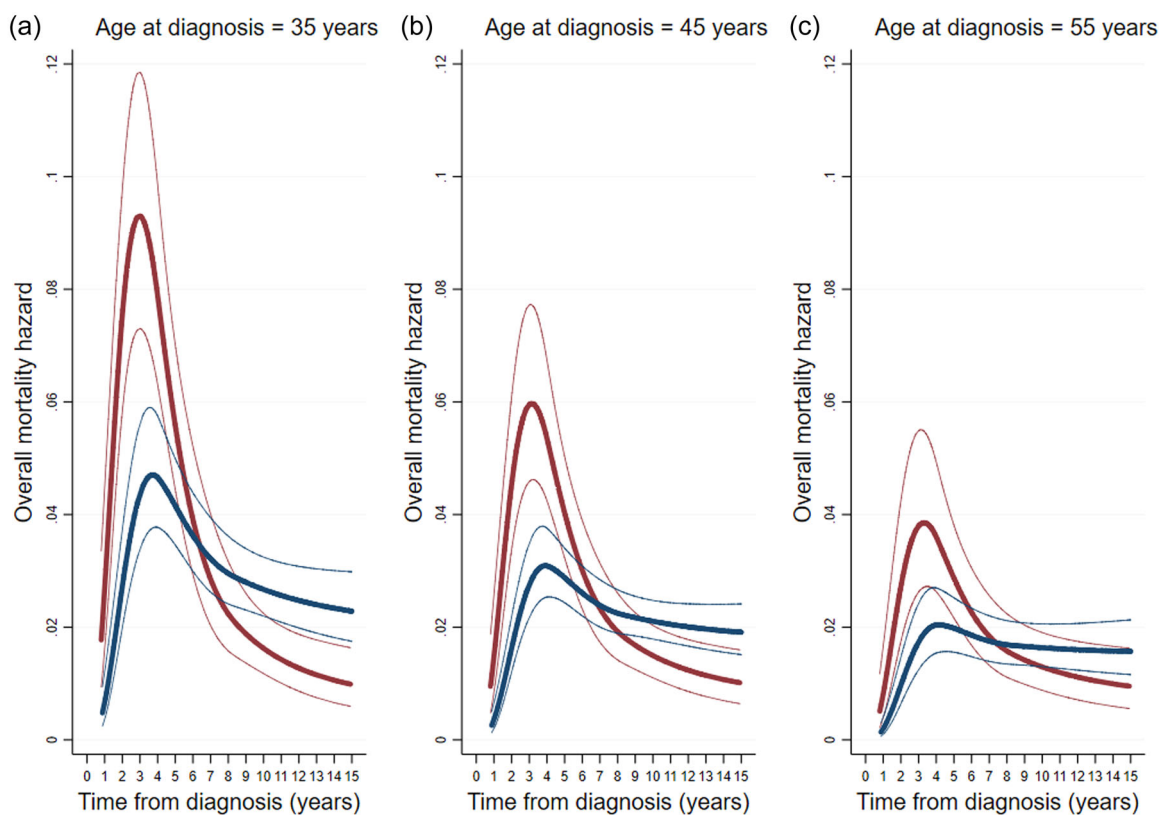

FIGURE 3 Mortality hazard after breast cancer by time from diagnosis, age at diagnosis, and oestrogen receptor status (negative red, positive blue) with 95% confidence intervals (from Jayasekara et al., 2019). (a) age at diagnosis is 35 years; (b) age at diagnosis is 45 years; (c) age at diagnosis is 55 years.

4.3 | To estimate penetrance for genes and variants as a function of characteristics

As genetic technologies advance and more genes and genetic regions are implicated in cancer risks, estimating penetrance (age-specific incidences) associated with having specific variants or classes of typically rare variants (such as those currently classified as being of 'unknown significance', will become more important. Pedigree analyses of families, once properly adjusted for their ascertainment (Li, Silvestri, et al., 2022; Win et al., 2021), can be used to estimate penetrance as a function of a variant's supposed pathogenicity predicted by different classification methods (e.g., Align-GVGD, Polyphen, Grantham matrix scores and sequence conservation in mammalian species) (Dowty et al., 2014). Quantitative estimates can then be attributed to a priori clinical categorisations, such as 'pathogenic' or 'likely pathogenic' and used to help better define, and interpret, clinical classifications.

4.4 | To determine how known and unknown genetic and other factors combine to determine, or cause, risk of breast and bowel cancers

Multivariable multiplicative logistic and Cox regression models could be fitted to determine how risk factors combine; lack of interactions on the multiplicative scale implies interactions on the absolute scale. Gene–environment interactions can be assessed by fitting interaction terms for measures of underlying familial risk profile based on pedigree models (Hopper et al., 2018) such as BOADICEA for breast cancer (Antoniou et al., 2008; Lee et al., 2019, 2022) and CRISP (Colorectal cancer RISk Prediction decision support tool) for bowel cancer (Emery et al., 2023; Walker et al., 2017, 2018). ICE FALCON could be applied to longitudinal data from siblings and twin pairs to assess if a risk factor at a young age has a causal effect on that risk factor at a later age, and if it has a casual effect on the cancers, information critical for informing prevention.

These family analyses also give insights into the role of familial confounding. For example, we found that height was a risk factor for breast cancer when using population controls but not when using siblings as controls (Milne et al., 2011). This implies that a substantial proportion of the height association is due to familial factors that cause both height and breast cancer. ICE FALCON analyses could further clarify interpretation of this observation.

4.5 | Qualitative research

Studies of families and twins allow for comparisons and observations on one another that could give insights not otherwise possible, with special relevance to early life exposures. For example, case–control questionnaire studies have asked female twin pairs to compare one another at the time of puberty using questions such as: who went through puberty first, who was tallest, heaviest and whose breast developed earlier (Hamilton & Mack, 2003; Swerdlow, 2002). Further exploration, using qualitative in-depth interviews of participants' experiences, could contribute additional insights and clarify details that questionnaire studies could not hope to answer. Qualitative studies of relative pairs discordant for cancer, for instance, could generate hypotheses about potential early life causes, as semistructured in-depth interviews allow participants to raise issues and experiences that are not necessarily apparent to researchers. Qualitative studies of relative pairs discordant for exposures could also provide more reliable evidence on potential causes, with implications for prevention. For example, we previously found evidence from statistical modelling using the regressive logistic model developed by the late George Bonney that a woman's smoking status was strongly predicted by the smoking status of her next oldest sister (Hopper, Jenkins, et al., 1999). We are currently conducting a twin study of breast and colorectal cancer screening practices with the intention of conducting qualitative research with twin pairs who differ in screening practice to help explore the reasons for their discordancy further.

5 | DISCUSSION

This paper presents some ideas on how the gaps in knowledge about the genetic and nongenetic causes of cancers, especially breast and colorectal cancers diagnosed at young adult ages for which impact is greatest yet least is known.

For example, even though having a family history is a much stronger predictor of young disease, the vast majority of these families do not have a known cause despite the latest gene panel testing. There are 'major' genes for young breast and bowel cancers yet to be discovered. Current polygenic risk scores are not making a big difference, especially for those at a young age for whom screening might not be available. There is no established polygenic risk score for Lynch syndrome (the main familial non-polyposis bowel cancer) despite wide variation in familial risk. Known nongenetic risk factors explain little variation in risk and risk discrimination on a population scale.

Advances in genetic, epigenetic, proteomic and other technologies offer great potential. Family and twin study designs are innovative and powerful, and extensive resources are already available and being enhanced, and it is important that they be extensively used. The combination of new technologies, powerful analytic tools for twin and family designs and extensive resources could provide exciting opportunities for new discoveries. It is also important that the expertise and experiences of those involved in creating and sustaining long-term studies and experience in the design and statistical analyses are utilised to avoid errors and guide interpretation.

5.1 | Limitations and future possibilities

Genetic Epidemiology was initially defined as 'a science which deals with the aetiology, distribution, and control of disease in groups of relatives and with inherited causes of disease in populations' (Morton, 1982) and traditionally based on studies of genetically or otherwise related individuals. As well as its uniqueness and strength, this familial aspect has posed limitations on research and made it more difficult to replicate findings. Collecting family data, especially on a large scale and with the epidemiological requirements of defined sampling with high responses from probands and relatives alike, is more complicated than studying isolated individuals alone. We were able to address these limitations through dedicated resource-building funding from the National Institute of Health specifically to study breast and colorectal cancers. With advances in genetic technologies, large-scale genotyping of research studies is being achieved and in doing so related participants are being accurately identified, thereby opening up new opportunities for family-based studies and analyses; see Section 2.4.

We hope that this article will attract others to join us in addressing the challenges and opportunities we have identified.

AFFILIATIONS
[1]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Carlton, Victoria, Australia

[2]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

[3]Murdoch Children's Research Institute, Royal Children's Hospital, Parkville, Victoria, Australia

[4]Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia

[5]Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, Australia

[6]Genetic Technologies Ltd., Fitzroy, Victoria, Australia

[7]Department of Data Science and AI, Faculty of Information Technology, Monash University, Melbourne, Victoria, Australia

[8]ARC Training Centre in Cognitive Computing for Medical Technologies, University of Melbourne, Carlton, Victoria, Australia

[9]The Florey Department of Neuroscience and Mental Health, The University of Melbourne, Parkville, Victoria, Australia

[10]University of Melbourne Centre for Cancer Research, Victorian Comprehensive Cancer Centre, Melbourne, Victoria, Australia

[11]Genetic Medicine, Royal Melbourne Hospital, Parkville, Victoria, Australia

[12]Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada

[13]Institute for Mental Health Policy Research, Centre for Addiction and Mental Health, Toronto, Ontario, Canada

[14]Centre for Adolescent Health, Murdoch Children's Research Institute, Parkville, Victoria, Australia

[15]School of Population and Global Health, The University of Western Australia, Perth, Western Australia, Australia

[16]Justice Health Group, Curtin School of Population Health, Curtin University, Perth, Western Australia, Australia

[17]Department of Public Health Sciences, Division of Genome and Health Big Data, Graduate School of Public Health, Seoul National University, Seoul, South Korea

[18]Genome Medicine Institute, Seoul National University, Seoul, South Korea

[19]Institute of Health and Environment, Seoul National University, Seoul, South Korea

[20]Department of Clinical Pathology, The University of Melbourne, Parkville, Victoria, Australia

[21]Department of Medicine, Royal Melbourne Hospital, The University of Melbourne, Parkville, Victoria, Australia

[22]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York, USA

CONFLICT OF INTEREST STATEMENT
G. S. D. is an employee of the Genetic Technologies Ltd. The other authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analysed in this study.

ORCID

Shuai Li http://orcid.org/0000-0002-8696-8594
Tuong L. Nguyen http://orcid.org/0000-0002-6597-8363
Zhoufeng Ye https://orcid.org/0000-0002-0884-4246
Benjamin Goudey https://orcid.org/0000-0002-2318-985X
Jennifer D. Brooks http://orcid.org/0000-0001-7574-4256
Lucas Calais-Ferreira https://orcid.org/0000-0001-6153-8080

REFERENCES

Acheampong, T., Kehm, R. D., Terry, M. B., Argov, E. L., & Tehranifar, P. (2020). Incidence trends of breast cancer molecular subtypes by age and race/ethnicity in the US from 2010 to 2016. *JAMA Network Open*, *3*(8), e2013226. https://doi.org/10.1001/jamanetworkopen.2020.13226

Antoniou, A. C., Cunningham, A. P., Peto, J., Evans, D. G., Lalloo, F., Narod, S. A., Risch, H. A., Eyfjord, J. E., Hopper, J. L., Southey, M. C., Olsson, H., Johannsson, O., Borg, A., Passini, B., Radice, P., Manoukian, S., Eccles, D. M., Tang, N., Olah, E., ... Easton, D. F. (2008). THE BOADICEA model of genetic susceptibility to breast and ovarian cancers: Updates and extensions. *British Journal of Cancer*, *98*(8), 1457–1466. https://doi.org/10.1038/sj.bjc.6604305

Araghi, M., Soerjomataram, I., Bardot, A., Ferlay, J., Cabasag, C. J., Morrison, D. S., De, P., Tervonen, H., Walsh, P. M., Bucher, O., Engholm, G., Jackson, C., McClure, C., Woods, R. R., Saint-Jacques, N., Morgan, E., Ransom, D., Thursfield, V., Møller, B., ... Arnold, M. (2019). Changes in colorectal cancer incidence in seven high-income countries: A population-based study. *The Lancet Gastroenterology & Hepatology*, *4*(7), 511–518. https://doi.org/10.1016/S2468-1253(19)30147-5

Athanasiadis, G., Meijsen, J. J., Helenius, D., Schork, A. J., Ingason, A., Thompson, W. K., Geschwind, D. H., Werge, T., & Buil, A. (2022). A comprehensive map of genetic relationships among diagnostic categories based on 48.6 million relative pairs from the Danish genealogy. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(6), e2118688119. https://doi.org/10.1073/pnas.2118688119

Brumpton, B., Sanderson, E., Heilbron, K., Hartwig, F. P., Harrison, S., Vie, G. Å., Cho, Y., Howe, L. D., Hughes, A., Boomsma, D. I., Havdahl, A., Hopper, J., Neale, M., Nivard, M. G., Pedersen, N. L., Reynolds, C. A., Tucker-Drob, E. M., Grotzinger, A., Howe, L., ... Davies, N. M. (2020). Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. *Nature Communications*, *11*(1), 3519. https://doi.org/10.1038/s41467-020-17117-4

Bui, M., Bjørnerem, Å., Ghasem-Zadeh, A., Dite, G. S., Hopper, J. L., & Seeman, E. (2013). Architecture of cortical bone determines in part its remodelling and structural decay. *Bone*, *55*(2), 353–358. https://doi.org/10.1016/j.bone.2013.04.020

Bui, M., Zebaze, R., Li, S., Hopper, J. L., & Bjørnerem, Å. (2020). Are the relationships of lean mass and fat mass with bone microarchitecture causal or due to familial confounders? A novel study of adult female twin pairs. *JBMR Plus*, *4*(9), e10386. https://doi.org/10.1002/jbm4.10386

Cheng, I., Kocarnik, J. M., Dumitrescu, L., Lindor, N. M., Chang-Claude, J., Avery, C. L., Caberto, C. P., Love, S. A., Slattery, M. L., Chan, A. T., Baron, J. A., Hindorff, L. A., Park, S. L., Schumacher, F. R., Hoffmeister, M., Kraft, P., Butler, A. M., Duggan, D. J., Hou, L., ... Peters, U. (2014). Pleiotropic effects of genetic risk variants for other cancers on colorectal cancer risk: PAGE, GECCO and CCFR consortia. *Gut*, *63*(5), 800–807. https://doi.org/10.1136/gutjnl-2013-305189

Dite, G. S., Gurrin, L. C., Byrnes, G. B., Stone, J., Gunasekara, A., McCredie, M. R. E., English, D. R., Giles, G. G., Cawson, J., Hegele, R. A., Chiarelli, A. M., Yaffe, M. J., Boyd, N. F., & Hopper, J. L. (2008). Predictors of mammographic density: Insights gained from a novel regression analysis of a twin study. *Cancer Epidemiology, Biomarkers & Prevention*, *17*(12), 3474–3481. https://doi.org/10.1158/1055-9965.EPI-07-2636

Dowty, J. G., Lee, E., McKean-Cowdin, R., Henderson, B. E., Bernstein, L., Ursin, G., & Hopper, J. L. (2014). Estimating risks for variants of unknown significance according to their predicted pathogenicity classes with application to BRCA1. *Breast Cancer Research and Treatment*, *144*(1), 171–177. https://doi.org/10.1007/s10549-014-2845-6

Dowty, J. G., Win, A. K., Buchanan, D. D., Lindor, N. M., Macrae, F. A., Clendenning, M., Antill, Y. C., Thibodeau, S. N., Casey, G., Gallinger, S., Marchand, L. L., Newcomb, P. A., Haile, R. W., Young, G. P., James, P. A., Giles, G. G., Gunawardena, S. R., Leggett, B. A., Gattas, M., ... Jenkins, M. A. (2013). Cancer risks for MLH1 and MSH2 mutation carriers. *Human Mutation*, *34*(3), 490–497. https://doi.org/10.1002/humu.22262

Easton, D. F., Ford, D., & Bishop, D. T. (1995). Breast and ovarian cancer incidence in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. *American Journal of Human Genetics*, *56*(1), 265–271. https://www.ncbi.nlm.nih.gov/pubmed/7825587

Emery, J. D., Jenkins, M. A., Saya, S., Chondros, P., Oberoi, J., Milton, S., Novy, K., Habgood, E., Karnchanachari, N., Pirotta, M., Trevena, L., Bickerstaffe, A., De Abreu Lourenço, R., Crothers, A., Ouakrim, D. A., Flander, L., Dowty, J. G., Walter, F. M., Clark, M., ... McIntosh, J. G. (2023). The Colorectal cancer RISk Prediction (CRISP) trial: A randomised controlled trial of a decision support tool for risk-stratified colorectal cancer screening. *British Journal of General Practice*, *73*(733), e556–e565. https://doi.org/10.3399/BJGP.2022.0480

Fisher, R. (1951). Limits to intensive production in animals. *British Agricultural Bulletin*, *4*, 2.

Freedman, L. S., Graubard, B. I., & Schatzkin, A. (1992). Statistical validation of intermediate endpoints for chronic diseases. *Statistics in Medicine*, *11*(2), 167–178.

Graff, R. E., Möller, S., Passarelli, M. N., Witte, J. S., Skytthe, A., Christensen, K., Tan, Q., Adami, H. O., Czene, K.,

Harris, J. R., Pukkala, E., Kaprio, J., Giovannucci, E. L., Mucci, L. A., & Hjelmborg, J. B. (2017). Familial risk and heritability of colorectal cancer in the Nordic Twin Study of Cancer. *Clinical Gastroenterology and Hepatology*, *15*(8), 1256–1264. https://doi.org/10.1016/j.cgh.2016.12.041

Hamilton, A. S., & Mack, T. M. (2003). Puberty and genetic susceptibility to breast cancer in a case–control study in twins. *New England Journal of Medicine*, *348*(23), 2313–2322. https://doi.org/10.1056/NEJMoa021293

Harris, J. R., Hjelmborg, J., Adami, H. O., Czene, K., Mucci, L., & Kaprio, J. (2019). The Nordic Twin Study on Cancer— NorTwinCan. *Twin Research and Human Genetics*, *22*(6), 817–823. https://doi.org/10.1017/thg.2019.71

Hopper, J. L. (2015). Odds per adjusted standard deviation: Comparing strengths of associations for risk factors measured on different scales and across diseases and populations. *American Journal of Epidemiology*, *182*(10), 863–867. https://doi.org/10.1093/aje/kwv193

Hopper, J. L., Bui, Q. M., Erbas, B., Matheson, M. C., Gurrin, L. C., Burgess, J. A., Lowe, A. J., Jenkins, M. A., Abramson, M. J., Walters, E. H., Giles, G. G., & Dharmage, S. C. (2012). Does eczema in infancy cause hay fever, asthma, or both in childhood? Insights from a novel regression model of sibling data. *Journal of Allergy and Clinical Immunology*, *130*(5), 1117–1122.e1. https://doi.org/10.1016/j.jaci.2012.08.003

Hopper, J. L., & Carlin, J. B. (1992). Familial aggregation of a disease consequent upon correlation between relatives in a risk factor measured on a continuous scale. *American Journal of Epidemiology*, *136*(9), 1138–1147. https://doi.org/10.1093/oxfordjournals.aje.a116580

Hopper, J. L., Dite, G. S., MacInnis, R. J., Liao, Y., Zeinomar, N., Knight, J. A., Southey, M. C., Milne, R. L., Chung, W. K., Giles, G. G., Genkinger, J. M., McLachlan, S. A., Friedlander, M. L., Antoniou, A. C., Weideman, P. C., Glendon, G., Nesci, S., Andrulis, I. L., Buys, S. S., … Terry, M. B. (2018). Age-specific breast cancer risk by body mass index and familial risk: Prospective family study cohort (ProF-SC). *Breast Cancer Research*, *20*(1), 132. https://doi.org/10.1186/s13058-018-1056-1

Hopper, J. L., Dowty, J. G., Nguyen, T. L., Li, S., Dite, G. S., MacInnis, R. J., Makalic, E., Schmidt, D. F., Bui, M., Stone, J., Sung, J., Jenkins, M. A., Giles, G. G., Southey, M. C., & Mathews, J. D. (2023). Variance of age-specific log incidence decomposition (VALID): A unifying model of measured and unmeasured genetic and non-genetic risks. *International Journal of Epidemiology*, *52*, 1557–1568. https://doi.org/10.1093/ije/dyad086

Hopper, J. L., Jenkins, M. A., Macaskill, G. T., & Giles, G. G. (1999). Analysis of familial aggregation in a binary trait by a logistic regression and the regressive logistic model. In M. C. LaBuda & E. L. Grigorenko (Eds.), *On the way to individuality: Current methodological issues in behavior genetics* (pp. 155–183). Nova Science Publishers.

Hopper, J. L., Nguyen, T. L., Schmidt, D. F., Makalic, E., Song, Y. M., Sung, J., Dite, G. S., Dowty, J. G., & Li, S. (2020). Going beyond conventional mammographic density to discover novel mammogram-based predictors of breast cancer risk. *Journal of Clinical Medicine*, *9*(3), 627. https://doi.org/10.3390/jcm9030627

Hopper, J. L., & Seeman, E. (1994). The bone density of female twins discordant for tobacco use. *New England Journal of Medicine*, *330*(6), 387–392. https://doi.org/10.1056/NEJM199402103300603

Hopper, J. L., Southey, M. C., Dite, G. S., Jolley, D. J., Giles, G. G., McCredie, M. R., Easton, D. F., & Venter, D. J. (1999). Population-based estimate of the average age-specific cumulative risk of breast cancer for a defined set of protein-truncating mutations in BRCA1 and BRCA2. Australian Breast Cancer Family Study. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, *8*(9), 741–747.

Jayasekara, H., MacInnis, R. J., Chamberlain, J. A., Dite, G. S., Leoce, N. M., Dowty, J. G., Bickerstaffe, A., Win, A. K., Milne, R. L., Giles, G. G., Terry, M. B., Eccles, D. M., Southey, M. C., & Hopper, J. L. (2019). Mortality after breast cancer as a function of time since diagnosis by estrogen receptor status and age at diagnosis. *International Journal of Cancer*, *145*(12), 3207–3217. https://doi.org/10.1002/ijc.32214

Jenkins, M. A., Buchanan, D. D., Lai, J., Makalic, E., Dite, G. S., Win, A. K., Clendenning, M., Winship, I. M., Hayes, R. B., Huyghe, J. R., Peters, U., Gallinger, S., Marchand, L. L., Figueiredo, J. C., Pai, R. K., Newcomb, P. A., Church, J. M., Casey, G., & Hopper, J. L. (2021). Assessment of a polygenic risk score for colorectal cancer to predict risk of Lynch Syndrome Colorectal Cancer. *JNCI Cancer Spectrum*, *5*(2), pkab022. https://doi.org/10.1093/jncics/pkab022

Jenkins, M. A., Win, A. K., Templeton, A. S., Angelakos, M. S., Buchanan, D. D., Cotterchio, M., Figueiredo, J. C., Thibodeau, S. N., Baron, J. A., Potter, J. D., Hopper, J. L., Casey, G., Gallinger, S., Le Marchand, L., Lindor, N. M., Newcomb, P. A., Haile, R. W., Ahnen, D., Anton, K., … Young, J. (2018). Cohort Profile: The Colon Cancer Family Registry Cohort (CCFRC). *International Journal of Epidemiology*, *47*(2), 387–388i. https://doi.org/10.1093/ije/dyy006

John, E. M., Hopper, J. L., Beck, J. C., Knight, J. A., Neuhausen, S. L., Senie, R. T., Ziogas, A., Andrulis, I. L., Anton-Culver, H., Boyd, N., Buys, S. S., Daly, M. B., O'Malley, F. P., Santella, R. M., Southey, M. C., Venne, V. L., Venter, D. J., West, D. W., Whittemore, A. S., & Seminara, D. (2004). The Breast Cancer Family Registry: An infrastructure for cooperative multinational, interdisciplinary and translational studies of the genetic epidemiology of breast cancer. *Breast Cancer Research*, *6*(4), R375–R389. https://doi.org/10.1186/bcr801

Kehm, R. D., Yang, W., Tehranifar, P., & Terry, M. B. (2019). 40 Years of change in age- and stage-specific cancer incidence rates in US women and men. *JNCI Cancer Spectrum*, *3*(3), pkz038. https://doi.org/10.1093/jncics/pkz038

Krämer, A., Green, J., Pollard, Jr., J., & Tugendreich, S. (2014). Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics*, *30*(4), 523–530. https://doi.org/10.1093/bioinformatics/btt703

Kuchenbaecker, K. B., Hopper, J. L., Barnes, D. R., Phillips, K. A., Mooij, T. M., Roos-Blom, M. J., Jervis, S., van Leeuwen, F. E., Milne, R. L., Andrieu, N., Goldgar, D. E., Terry, M. B., Rookus, M. A., Easton, D. F., Antoniou, A. C., McGuffog, L.,

Evans, D. G., Barrowdale, D., Frost, D., … Olsson, H. (2017). Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. *Journal of the American Medical Association*, *317*(23), 2402–2416. https://doi.org/10.1001/jama.2017.7112

Lai, J., Wong, C. K., Schmidt, D. F., Kapuscinski, M. K., Alpen, K., MacInnis, R. J., Buchanan, D. D., Win, A. K., Figueiredo, J. C., Chan, A. T., Harrison, T. A., Hoffmeister, M., White, E., Le Marchand, L., Pai, R. K., Peters, U., Hopper, J. L., Jenkins, M. A., & Makalic, E. (2023). Using DEPendency of association on the number of Top Hits (DEPTH) as a complementary tool to identify novel colorectal cancer susceptibility loci. *Cancer Epidemiology, Biomarkers & Prevention*, *32*, 1153–1159. https://doi.org/10.1158/1055-9965.EPI-22-1209

Lange, K. (2002). *Mathematical and statistical methods for genetic analysis* (Vol. 488). Springer.

Lee, A., Mavaddat, N., Cunningham, A., Carver, T., Ficorella, L., Archer, S., Walter, F. M., Tischkowitz, M., Roberts, J., Usher-Smith, J., Simard, J., Schmidt, M. K., Devilee, P., Zadnik, V., Jürgens, H., Mouret-Fourme, E., De Pauw, A., Rookus, M., Mooij, T. M., … Antoniou, A. C. (2022). Enhancing the BOADICEA cancer risk prediction model to incorporate new data on RAD51C, RAD51D, BARD1 updates to tumour pathology and cancer incidence. *Journal of Medical Genetics*, *59*(12), 1206–1218. https://doi.org/10.1136/jmedgenet-2022-108471

Lee, A., Mavaddat, N., Wilcox, A. N., Cunningham, A. P., Carver, T., Hartley, S., Babb de Villiers, C., Izquierdo, A., Simard, J., Schmidt, M. K., Walter, F. M., Chatterjee, N., Garcia-Closas, M., Tischkowitz, M., Pharoah, P., Easton, D. F., & Antoniou, A. C. (2019). BOADICEA: A comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genetics in Medicine*, *21*(8), 1708–1718. https://doi.org/10.1038/s41436-018-0406-9

Li, S., Bui, M., & Hopper, J. L. (2020). Inference about causation from examination of familial confounding (ICE FALCON): A model for assessing causation analogous to Mendelian randomization. *International Journal of Epidemiology*, *49*(4), 1259–1269. https://doi.org/10.1093/ije/dyaa065

Li, S., Dite, G. S., MacInnis, R. J., Bui, M., Nguyen, T. L., Esser, V. F. C., Ye, Z., Dowty, J. G., Makalic, E., Sung, J., Giles, G. G., Southey, M. C., & Hopper, J. L. (2024). Causation and familial confounding as explanations for the associations of polygenic risk scores with breast cancer: Evidence from innovative ICE FALCON and ICE CRISTAL analyses. *Genetic Epidemiology*. https://doi.org/10.1002/gepi.22556

Li, S., & Hopper, J. L. (2023). Implications of family history and polygenic risk scores for causation. *The American Journal of Human Genetics*, *110*(7), 1221–1223. https://doi.org/10.1016/j.ajhg.2023.05.015

Li, S., MacInnis, R. J., Lee, A., Nguyen-Dumont, T., Dorling, L., Carvalho, S., Dite, G. S., Shah, M., Luccarini, C., Wang, Q., Milne, R. L., Jenkins, M. A., Giles, G. G., Dunning, A. M., Pharoah, P. D. P., Southey, M. C., Easton, D. F., Hopper, J. L., & Antoniou, A. C. (2022). Segregation analysis of 17,425 population-based breast cancer families: Evidence for genetic susceptibility and risk prediction. *The American Journal of Human Genetics*, *109*(10), 1777–1788. https://doi.org/10.1016/j.ajhg.2022.09.006

Li, S., Nguyen, T. L., Nguyen-Dumont, T., Dowty, J. G., Dite, G. S., Ye, Z., Trinh, H. N., Evans, C. F., Tan, M., Sung, J., Jenkins, M. A., Giles, G. G., Hopper, J. L., & Southey, M. C. (2022). Genetic aspects of mammographic density measures associated with breast cancer risk. *Cancers*, *14*(11), 2767. https://doi.org/10.3390/cancers14112767

Li, S., Silvestri, V., Leslie, G., Rebbeck, T. R., Neuhausen, S. L., Hopper, J. L., Nielsen, H. R., Lee, A., Yang, X., McGuffog, L., Parsons, M. T., Andrulis, I. L., Arnold, N., Belotti, M., Borg, Å., Buecher, B., Buys, S. S., Caputo, S. M., Chung, W. K., … Antoniou, A. C. (2022). Cancer risks associated with BRCA1 and BRCA2 pathogenic variants. *Journal of Clinical Oncology*, *40*(14), 1529–1541. https://doi.org/10.1200/JCO.21.02112

Li, S., Wong, E. M., Bui, M., Nguyen, T. L., Joo, J. H. E., Stone, J., Dite, G. S., Dugué, P. A., Milne, R. L., Giles, G. G., Saffery, R., Southey, M. C., & Hopper, J. L. (2019). Inference about causation between body mass index and DNA methylation in blood from a twin family study. *International Journal of Obesity*, *43*(2), 243–252. https://doi.org/10.1038/s41366-018-0103-4

Li, S., Wong, E. M., Bui, M., Nguyen, T. L., Joo, J. H. E., Stone, J., Dite, G. S., Giles, G. G., Saffery, R., Southey, M. C., & Hopper, J. L. (2018). Causal effect of smoking on DNA methylation in peripheral blood: A twin and family study. *Clinical Epigenetics*, *10*(1), 18. https://doi.org/10.1186/s13148-018-0452-9

MacInnis, R. J., Pike, M. C., & Hopper, J. L. (2016). Risk-reducing surgery in hereditary breast and ovarian cancer. *New England Journal of Medicine*, *374*(24), 2403–2404. https://doi.org/10.1056/NEJMc1602861

MacInnis, R. J., Schmidt, D. F., Makalic, E., Severi, G., FitzGerald, L. M., Reumann, M., Kapuscinski, M. K., Kowalczyk, A., Zhou, Z., Goudey, B., Qian, G., Bui, Q. M., Park, D. J., Freeman, A., Southey, M. C., Al Olama, A. A., Kote-Jarai, Z., Eeles, R. A., Hopper, J. L., & Giles, G. G. (2016). Use of a novel nonparametric version of DEPTH to identify genomic regions associated with prostate cancer risk. *Cancer Epidemiology, Biomarkers & Prevention*, *25*(12), 1619–1624. https://doi.org/10.1158/1055-9965.EPI-16-0301

Mahmood, K., Thomas, M., Qu, C., Hsu, L., Buchanan, D. D., Peters, U., Wang, X., Huyghe, J. R., Joo, J. E., Georgeson, P., Arndt, V., Berndt, S. I., Bézieau, S., Bien, S. A., Bishop, D. T., Brenner, H., Brezina, S., Burnett-Hartman, A., Campbell, P. T., … Jenkins, M. A. (2023). Elucidating the risk of colorectal cancer for variants in hereditary colorectal cancer genes. *Gastroenterology*, *165*, 1070–1076.e3. https://doi.org/10.1053/j.gastro.2023.06.032

Makalic, E., Schmidt, D. F., & Hopper, J. L. (2013). DEPTH: A novel algorithm for feature ranking with application to genome-wide association studies. In *Australasian Joint Conference on Artificial Intelligence*. Springer International Publishing.

Mars, N., Lindbohm, J. V., Della Briotta Parolo, P., Widén, E., Kaprio, J., Palotie, A., & Ripatti, S. (2022). Systematic comparison of family history and polygenic risk across 24 common diseases. *The American Journal of Human Genetics*, *109*(12), 2152–2162. https://doi.org/10.1016/j.ajhg.2022.10.009

Martin, L. J., Melnichouk, O., Guo, H., Chiarelli, A. M., Hislop, T. G., Yaffe, M. J., Minkin, S., Hopper, J. L., & Boyd, N. F. (2010). Family history, mammographic density, and risk of breast cancer. *Cancer Epidemiology, Biomarkers & Prevention*, *19*(2), 456–463. https://doi.org/10.1158/1055-9965.EPI-09-0881

Mavaddat, N., Michailidou, K., Dennis, J., Lush, M., Fachal, L., Lee, A., Tyrer, J. P., Chen, T. H., Wang, Q., Bolla, M. K., Yang, X., Adank, M. A., Ahearn, T., Aittomäki, K., Allen, J., Andrulis, I. L., Anton-Culver, H., Antonenkova, N. N., Arndt, V., ... Easton, D. F. (2019). Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am J Hum Genet*, *104*(1), 21–34. https://doi.org/10.1016/j.ajhg.2018.11.002

Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., Lemaçon, A., Soucy, P., Glubb, D., Rostamianfar, A., Bolla, M. K., Wang, Q., Tyrer, J., Dicks, E., Lee, A., Wang, Z., Allen, J., Keeman, R., Eilber, U., ... Easton, D. F. (2017). Association analysis identifies 65 new breast cancer risk loci. *Nature*, *551*(7678), 92–94. https://doi.org/10.1038/nature24284

Milne, R. L., John, E. M., Knight, J. A., Dite, G. S., Southey, M. C., Giles, G. G., Apicella, C., West, D. W., Andrulis, I. L., Whittemore, A. S., & Hopper, J. L. (2011). The potential value of sibling controls compared with population controls for association studies of lifestyle-related risk factors: An example from the Breast Cancer Family Registry. *International Journal of Epidemiology*, *40*(5), 1342–1354. https://doi.org/10.1093/ije/dyr110

Milne, R. L., Knight, J. A., John, E. M., Dite, G. S., Balbuena, R., Ziogas, A., Andrulis, I. L., West, D. W., Li, F. P., Southey, M. C., Giles, G. G., McCredie, M. R. E., Hopper, J. L., & Whittemore, A. S. (2005). Oral contraceptive use and risk of early-onset breast cancer in carriers and noncarriers of BRCA1 and BRCA2 mutations. *Cancer Epidemiology, Biomarkers & Prevention*, *14*(2), 350–356. https://doi.org/10.1158/1055-9965.EPI-04-0376

Möller, S., Mucci, L. A., Harris, J. R., Scheike, T., Holst, K., Halekoh, U., Adami, H. O., Czene, K., Christensen, K., Holm, N. V., Pukkala, E., Skytthe, A., Kaprio, J., & Hjelmborg, J. B. (2016). The heritability of breast cancer among women in the Nordic Twin Study of Cancer. *Cancer Epidemiology, Biomarkers & Prevention*, *25*(1), 145–150. https://doi.org/10.1158/1055-9965.EPI-15-0913

Morton, N. E. (1982). *Outline of genetic epidemiology*. Karger Publishers.

Neuhausen, S. L., Ozcelik, H., Southey, M. C., John, E. M., Godwin, A. K., Chung, W., Iriondo-Perez, J., Miron, A., Santella, R. M., Whittemore, A., Andrulis, I. L., Buys, S. S., Daly, M. B., Hopper, J. L., Seminara, D., Senie, R. T., & Terry, M. B. (2009). BRCA1 and BRCA2 mutation carriers in the Breast Cancer Family Registry: An open resource for collaborative research. *Breast Cancer Research and Treatment*, *116*(2), 379–386. https://doi.org/10.1007/s10549-008-0153-8

Nguyen, T. L., Aung, Y. K., Evans, C. F., Dite, G. S., Stone, J., MacInnis, R. J., Dowty, J. G., Bickerstaffe, A., Aujard, K., Rommens, J. M., Song, Y. M., Sung, J., Jenkins, M. A., Southey, M. C., Giles, G. G., Apicella, C., & Hopper, J. L. (2017). Mammographic density defined by higher than conventional brightness thresholds better predicts breast cancer risk. *International Journal of Epidemiology*, *46*(2), 652–661. https://doi.org/10.1093/ije/dyw212

Nguyen, T. L., Aung, Y. K., Li, S., Trinh, N. H., Evans, C. F., Baglietto, L., Krishnan, K., Dite, G. S., Stone, J., English, D. R., Song, Y. M., Sung, J., Jenkins, M. A., Southey, M. C., Giles, G. G., & Hopper, J. L. (2018). Predicting interval and screen-detected breast cancers from mammographic density defined by different brightness thresholds. *Breast Cancer Research*, *20*(1), 152. https://doi.org/10.1186/s13058-018-1081-0

Nguyen, T. L., Li, S., Dite, G. S., Aung, Y. K., Evans, C. F., Trinh, H. N., Baglietto, L., Stone, J., Song, Y. M., Sung, J., English, D. R., Jenkins, M. A., Dugué, P. A., Milne, R. L., Southey, M. C., Giles, G. G., Pike, M. C., & Hopper, J. L. (2020). Interval breast cancer risk associations with breast density, family history and breast tissue aging. *International Journal of Cancer*, *147*(2), 375–382. https://doi.org/10.1002/ijc.32731

Nguyen, T. L., Li, S., Dowty, J. G., Dite, G. S., Ye, Z., Nguyen-Dumont, T., Trinh, H. N., Evans, C. F., Tan, M., Sung, J., Jenkins, M. A., Giles, G. G., Southey, M. C., & Hopper, J. L. (2022). Familial aspects of mammographic density measures associated with breast cancer risk. *Cancers*, *14*(6), 1483. https://doi.org/10.3390/cancers14061483

Nguyen-Dumont, T., Steen, J. A., Winship, I., Park, D. J., Pope, B. J., Hammet, F., Mahmoodi, M., Tsimiklis, H., Theys, D., Clendenning, M., Giles, G. G., Hopper, J. L., & Southey, M. C. (2020). Mismatch repair gene pathogenic germline variants in a population-based cohort of breast cancer. *Familial Cancer*, *19*(3), 197–202. https://doi.org/10.1007/s10689-020-00164-7

Nissen, F. I., Esser, V. F. C., Bui, M., Li, S., Hopper, J. L., Bjørnerem, Å., & Hansen, A. K. (2023). Is there a causal relationship between physical activity and bone microarchitecture? A study of adult female twin pairs. *Journal of Bone and Mineral Research*, *38*(7), 951–957. https://doi.org/10.1002/jbmr.4826

Pike, M. C., Krailo, M. D., Henderson, B. E., Casagrande, J. T., & Hoel, D. G. (1983). 'Hormonal' risk factors, 'breast tissue age' and the age-incidence of breast cancer. *Nature*, *303*(5920), 767–770. https://doi.org/10.1038/303767a0

Polubriaginof, F. C. G., Vanguri, R., Quinnies, K., Belbin, G. M., Yahi, A., Salmasian, H., Lorberbaum, T., Nwankwo, V., Li, L., Shervey, M. M., Glowe, P., Ionita-Laza, I., Simmerling, M., Hripcsak, G., Bakken, S., Goldstein, D., Kiryluk, K., Kenny, E. E., Dudley, J., ... Tatonetti, N. P. (2018). Disease heritability inferred from familial relationships reported in medical records. *Cell*, *173*(7), 1692–1704.e11. https://doi.org/10.1016/j.cell.2018.04.032

Stone, J., Dite, G. S., Giles, G. G., Cawson, J., English, D. R., & Hopper, J. L. (2012). Inference about causation from examination of familial confounding: Application to longitudinal twin data on mammographic density measures that predict breast cancer risk. *Cancer Epidemiology, Biomarkers & Prevention*, *21*(7), 1149–1155. https://doi.org/10.1158/1055-9965.EPI-12-0051

Swerdlow, A. J. (2002). Risk factors for breast cancer at young ages in twins: An international population-based study. *CancerSpectrum Knowledge Environment*, *94*(16), 1238–1246. https://doi.org/10.1093/jnci/94.16.1238

Terry, M. B., Phillips, K. A., Daly, M. B., John, E. M., Andrulis, I. L., Buys, S. S., Goldgar, D. E., Knight, J. A., Whittemore, A. S., Chung, W. K., Apicella, C., & Hopper, J. L. (2016). Cohort profile: The Breast Cancer Prospective Family Study Cohort (ProF-SC). *International Journal of Epidemiology*, *45*(3), 683–692. https://doi.org/10.1093/ije/dyv118

Thomas, M., Sakoda, L. C., Hoffmeister, M., Rosenthal, E. A., Lee, J. K., van Duijnhoven, F. J. B., Platz, E. A., Wu, A. H., Dampier, C. H., de la Chapelle, A., Wolk, A., Joshi, A. D., Burnett-Hartman, A., Gsur, A., Lindblom, A., Castells, A., Win, A. K., Namjou, B., Van Guelpen, B., ... Hsu, L. (2020). Genome-wide modeling of polygenic risk score in colorectal cancer risk. *The American Journal of Human Genetics*, *107*(3), 432–444. https://doi.org/10.1016/j.ajhg.2020.07.006

Walker, J. G., Bickerstaffe, A., Hewabandu, N., Maddumarachchi, S., Dowty, J. G., Jenkins, M., Pirotta, M., Walter, F. M., & Emery, J. D. (2017). The CRISP colorectal cancer risk prediction tool: An exploratory study using simulated consultations in Australian primary care. *BMC Medical Informatics and Decision Making*, *17*(1), 13. https://doi.org/10.1186/s12911-017-0407-7

Walker, J. G., Macrae, F., Winship, I., Oberoi, J., Saya, S., Milton, S., Bickerstaffe, A., Dowty, J. G., De Abreu Lourenço, R., Clark, M., Galloway, L., Fishman, G., Walter, F. M., Flander, L., Chondros, P., Ait Ouakrim, D., Pirotta, M., Trevena, L., Jenkins, M. A., & Emery, J. D. (2018). The use of a risk assessment and decision support tool (CRISP) compared with usual care in general practice to increase risk-stratified colorectal cancer screening: Study protocol for a randomised controlled trial. *Trials*, *19*(1), 397. https://doi.org/10.1186/s13063-018-2764-7

Wilcox, N., Dumont, M., Gonzalez-Neira, A., Carvalho, S., Joly Beauparlant, C., Crotti, M., Luccarini, C., Soucy, P., Dubois, S., Nunez-Torres, R., Pita, G., Gardner, E. J., Dennis, J., Alonso, M. R., Alvarez, N., Baynes, C., Collin-Deschesnes, A. C., Desjardins, S., Becher, H., ... Simard, J. (2023). Exome sequencing identifies breast cancer susceptibility genes and defines the contribution of coding variants to breast cancer risk. *Nature Genetics*, *55*, 1435–1439. https://doi.org/10.1038/s41588-023-01466-z

Win, A. K., Dowty, J. G., Reece, J. C., Lee, G., Templeton, A. S., Plazzer, J. P., Buchanan, D. D., Akagi, K., Aksoy, S., Alonso, A., Alvarez, K., Amor, D. J., Ankathil, R., Aretz, S., Arnold, J. L., Aronson, M., Austin, R., Backman, A. S., Bajwa-ten Broeke, S. W., ... Redler, S. (2021). Variation in the risk of colorectal cancer in families with Lynch syndrome: A retrospective cohort study. *The Lancet Oncology*, *22*(7), 1014–1022. https://doi.org/10.1016/S1470-2045(21)00189-3

Win, A. K., Hopper, J. L., Buchanan, D. D., Young, J. P., Tenesa, A., Dowty, J. G., Giles, G. G., Goldblatt, J., Winship, I., Boussioutas, A., Young, G. P., Parry, S., Baron, J. A., Duggan, D., Gallinger, S., Newcomb, P. A., Haile, R. W., Marchand, L. L., Lindor, N. M., & Jenkins, M. A. (2013). Are the common genetic variants associated with colorectal cancer risk for DNA mismatch repair gene mutation carriers. *European Journal of Cancer*, *49*(7), 1578–1587. https://doi.org/10.1016/j.ejca.2013.01.029

Win, A. K., Jenkins, M. A., Dowty, J. G., Antoniou, A. C., Lee, A., Giles, G. G., Buchanan, D. D., Clendenning, M., Rosty, C.,

Ahnen, D. J., Thibodeau, S. N., Casey, G., Gallinger, S., Le Marchand, L., Haile, R. W., Potter, J. D., Zheng, Y., Lindor, N. M., Newcomb, P. A., ... MacInnis, R. J. (2017). Prevalence and penetrance of major genes and polygenes for colorectal cancer. *Cancer Epidemiology, Biomarkers & Prevention*, *26*(3), 404–412. https://doi.org/10.1158/1055-9965.EPI-16-0693

Wray, N. R., Yang, J., Goddard, M. E., & Visscher, P. M. (2010). The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS Genetics*, *6*(2), e1000864. https://doi.org/10.1371/journal.pgen.1000864

Ye, Z., Li, S., Dite, G. S., Nguyen, T. L., MacInnis, R. J., Andrulis, I. L., Buys, S. S., Daly, M. B., John, E. M., Kurian, A. W., Genkinger, J. M., Chung, W. K., Phillips, K. A., Thorne, H., Thorne, H., Winship, I. M., Milne, R. L., Dugué, P. A., Southey, M. C., ... Hopper, J. L. (2022). Weight is more informative than body mass index for Predicting Postmenopausal Breast Cancer Risk: Prospective Family Study Cohort (ProF-SC). *Cancer Prevention Research*, *15*(3), 185–191. https://doi.org/10.1158/1940-6207.CAPR-21-0164

Ye, Z., Nguyen, T., Dite, G., MacInnis, R., Schmidt, D., Makalic, E., Al-Qershi, O., Bui, M., Esser, V. F. C., Dowty, J., Trinh, H., Evans, C., Tan, M., Sung, J., Jenkins, M., Giles, G., Southey, M., Hopper, J., & Li, S. (2023). Causal relationships between breast cancer risk factors based on mammogram features. *Breast Cancer Research*, *25*(1), 127. https://doi.org/10.1186/s13058-023-01733-1

Young, J. P., Win, A. K., Rosty, C., Flight, I., Roder, D., Young, G. P., Frank, O., Suthers, G. K., Hewett, P. J., Ruszkiewicz, A., Hauben, E., Adelstein, B. A., Parry, S., Townsend, A., Hardingham, J. E., & Price, T. J. (2015). Rising incidence of early-onset colorectal cancer in Australia over two decades: Report and review. *Journal of Gastroenterology and Hepatology*, *30*(1), 6–13. https://doi.org/10.1111/jgh.12792

Yu, H., & Hemminki, K. (2020). Genetic epidemiology of colorectal cancer and associated cancers. *Mutagenesis*, *35*(3), 207–219. https://doi.org/10.1093/mutage/gez022

Zheng, Y., Hua, X., Win, A. K., MacInnis, R. J., Gallinger, S., Marchand, L. L., Lindor, N. M., Baron, J. A., Hopper, J. L., Dowty, J. G., Antoniou, A. C., Zheng, J., Jenkins, M. A., & Newcomb, P. A. (2020). A new comprehensive colorectal cancer risk prediction model incorporating family history, personal characteristics, and environmental factors. *Cancer Epidemiology, Biomarkers & Prevention*, *29*(3), 549–557. https://doi.org/10.1158/1055-9965.EPI-19-0929