Considerations in assessing germline variant pathogenicity using cosegregation analysis

Sophie Belman, MS[1,3], Michael T. Parsons, BSc Hons[2], Amanda B. Spurdle, PhD[2], David E. Goldgar, PhD[1] and Bing-Jian Feng, PhD [1]

Purpose: The American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) have developed guidelines for classifying germline variants as pathogenic or benign to interpret genetic testing results. Cosegregation analysis is an important component of the guidelines. There are two main approaches for cosegregation analysis: meiosis counting and Bayes factor–based quantitative methods. Of these, the ACMG/AMP guidelines employ only meiosis counting. The accuracy of either approach has not been sufficiently addressed in previous works.

Methods: We analyzed hypothetical, simulated, and real-life data to evaluate the accuracy of each approach for cancer-associated genes.

Results: We demonstrate that meiosis counting can provide incorrect classifications when the underlying genetic basis of the disease departs from simple Mendelian situations. Some Bayes factor approaches are currently implemented with inappropriate penetrance. We propose an improved penetrance model and describe several critical considerations, including the accuracy of cosegregation for moderate-risk genes and the impact of pleiotropy, population, and birth year. We highlight a webserver, COOL (Co-segregation Online, http://BJFengLab.org/), that implements an accurate Bayes factor cosegregation analysis.

Conclusion: An appropriate penetrance model improves the accuracy of Bayes factor cosegregation analysis for high-penetrant variants, and is a better choice than meiosis counting whenever feasible.

INTRODUCTION

Classification of germline variants within disease-causing genes as pathogenic or benign is crucial for the translation of clinical genetic testing results. The American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) classification standards and guidelines have been developed to interpret germline variants and are now widely used.[1] In the guidelines, a strength level (supporting, moderate, strong, or very strong) is assigned to each line of evidence for or against variant pathogenicity. These strength levels are then summed up within a scoring system to classify the variant as pathogenic, benign, likely pathogenic, likely benign, or variant of uncertain significance (VUS).

Cosegregation analysis, measuring how often an allele and a disease are inherited together in a pedigree, is a component of the ACMG/AMP approach for assessing germline variant pathogenicity.[1] Cosegregation can be assigned a base strength level of pathogenic supporting (code name PP1) when cosegregation with the disease is observed, or benign strong (code name BS4) when cosegregation is lacking. There is an allowance for PP1 to be up-weighted to PP1_moderate or PP1_strong with increasing evidence. It is noteworthy that cosegregation data alone are not sufficient evidence for pathogenicity, irrespective of the magnitude of evidence in favor of pathogenicity, because the variant of interest may be in linkage disequilibrium with the true pathogenic variant. That is, variant-specific data are required to support causality. Cosegregation analysis requires designation of a proband in each pedigree to adjust for ascertainment bias. Proband is defined here as the first person to be tested positive for the variant.

There are two main approaches for cosegregation analysis of variant causality. The first one is meiosis counting,[2] which is easy to calculate without using any computer software. Use of meiosis counting has been denoted in gene-specific adaptations of the ACMG/AMP criteria for *MYH7*,[3] *CDH1*,[4] *PTEN*,[5] *RUNX1*,[6] and *TP53*[7] (Table S1). All these genes associate with a complex disease or diseases characterized by variable age at onset, phenocopies, and incomplete penetrance. It is important to note that the original meiosis counting method considers both presence of variant in affected individuals and absence of variant in unaffected individuals,[2] while the above ACMG/AMP rules consider only

affected individuals.[1] Throughout this paper, meiosis counting refers to the ACMG/AMP approach for PP1.

Another method is the Bayes factor–based approach, which measures the likelihood that cosegregation patterns represent a gene–disease penetrance model. Applications of this approach include full-likelihood Bayes (FLB)[8] and cosegregation likelihood ratio (CSLR).[9] This method considers variant allele frequency, background disease incidence in the population, expected risk elevation conferred by a pathogenic variant, and age and sex of each individual. These are important factors for the analysis of complex diseases such as cancers. This approach produces a Bayes factor, which can be integrated with other lines of evidence within a quantitative multifactorial likelihood variant classification scheme.[10,11] Further, given the recent publication demonstrating that ACMG/AMP guidelines can be modeled as a Bayesian framework, the Bayes factor from one or more families with the same variant can also be transformed into strength categories to be used in the application of the ACMG/AMP guidelines.[12] Overall, Bayes factor–based approach has superior power, accuracy, and flexibility compared with meiosis counting, although at the cost of requiring parameters that may only be accessible for well-studied genes.

Penetrance parameters for Bayes factor cosegregation analysis are values corresponding to the probability of developing the disease(s) of interest among individuals with zero, one, or two copies of the putative risk allele, respectively. This definition is relatively straightforward for Mendelian diseases, but more complicated for complex diseases due to the consideration of age, sex, and multiple phenotypes associated with pathogenic variants in a given gene. For example, most cancer susceptibility genes increase cancer risk at multiple sites (Table S2), and both background incidence

and risk elevation conferred by germline pathogenic variants vary by age, sex, and cancer site (Fig. S1).[13–15]

To address this issue, ideally one would provide a penetrance value for each individual (or group of individuals) based on age, sex, disease type, and other relevant factors. Each distinct penetrance value is called a liability class. For phenotypes that are age-dependent, the approach is to model disease incidence using a survival analysis approach, whereby individuals who are diagnosed at age t years are considered unaffected until age $t–1$ and then become affected at age t (Eq. 2), hereafter referred to as the "survival model." This approach has been used to estimate risks for many adult-onset diseases, most recently for breast and ovarian cancer in individuals carrying a *PALB2* loss-of-function variant (hereafter "carrying" denotes being heterozygous for the variant predisposing to a dominantly inherited disease).[16] However, various implementations of cosegregation analysis tend to just use cumulative risks, hereafter referred to as the "cumulative risk model," or to use incidence rate for affected and cumulative risk for unaffected individuals, hereafter referred to as the "incidence rate model." An example of the former is the Bayes factor cosegregation analysis methods implemented in Analyze-MyVariant,[17] while the latter was used in both linkage studies[18,19] and in the assessment of cosegregation in most previous studies of VUS classification.[8,20,21] Difference in these approaches will manifest themselves primarily in the contribution to the cosegregation likelihood ratio of older individuals who carry the variant.

MATERIALS AND METHODS

We analyzed a hypothetical pedigree (Fig. 1) to compare different cosegregation analysis approaches for *BRCA1*. We
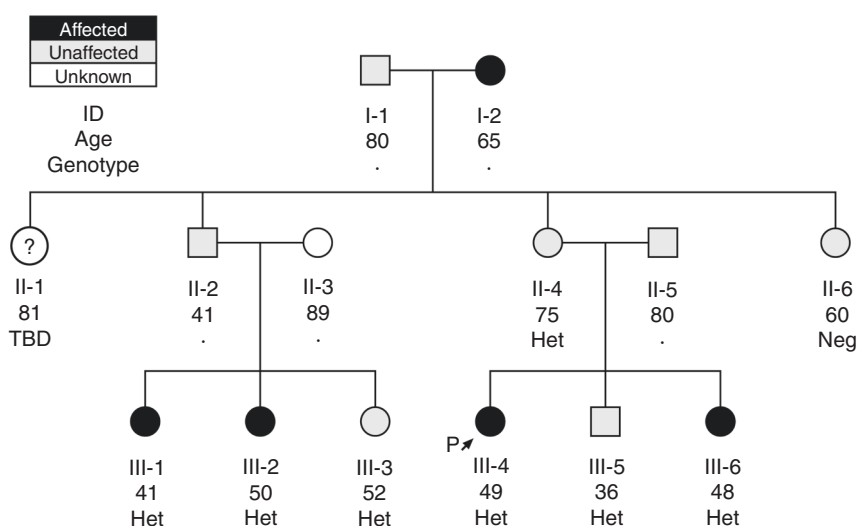

Fig. 1 A hypothetical pedigree used to compare different cosegregation methods. We varied the status of individual II-1 in multiple ways to assess the impact of affection status and genotype on cosegregation analysis. See Table 1 and Table S3 for results from these comparisons. *Affected* affected with breast cancer, *Age* age of onset for affected and age of the last examination for unaffected individuals, *Het* heterozygous for the allele of interest, *Neg* the individual is known to not have the allele of interest, *P and arrow* proband, *question mark and TBD* both affection status and genotype are to be determined.

chose this gene because reliable age-specific relative risk estimation is available (Table S2).[13–15] Affection status in the hypothetical pedigree was selected to refer to breast cancer only, since breast cancer has the highest incidence among *BRCA1*-associated cancers, the relative risk shows a decreasing trend with age, and breast cancer is the most frequently observed in individuals harboring a *BRCA1* pathogenic variant. Age-dependent cancer incidence rate in the general population was obtained from the Cancer Incidence in Five Continents (CI5),[22] United Kingdom population, year range 2008–2012. Incidence in individuals carrying a germline pathogenic variant was calculated as the product of incidence without a pathogenic variant and the relative risk for the disease in each age group. We calculated the incidence rates so that the combined incidence, based on an estimated allele frequency, was equal to the observed population incidence in CI5. The cumulative risk was calculated as the probability of having any one of these diseases at any age before the last examination:

$$\text{Cumulative Risk}_t = 1 - e^{-\sum_s \sum_{i=1}^t I_{s,i}} \quad (1)$$

where t is age in years, s denotes an associated disease, I is incidence rate, and e is 2.71828. The survival penetrance was calculated as follows:

$$\text{Survival Penetrance}_t =$$
$$\begin{cases} \text{affected with diseases: } (1 - \text{Cumulative Risk}_{t-1}) \times I_{s,t} \\ \text{unaffected: Cumulative Risk}_t \end{cases} \quad (2)$$

where t is age of diagnosis for affected and age of last exam for unaffected individuals. For affected individuals, this penetrance corresponds to the probability of not developing any one of the relevant diseases before age t multiplied by the probability of developing the observed disease at age t. Meiosis counting was performed manually based on affection status disregarding age of onset. Two software tools were used to carry out the Bayes factor–based approaches: (1) an in-house pipeline implementing the full-likelihood Bayes that is described further in Appendix S1; and (2) the AnalyzeMyVariant website (http://analyze.myvariant.org/), which permits segregation analysis using full-likelihood Bayes or cosegregation likelihood ratio.[17] Note that the website originally set up to perform cosegregation likelihood ratio calculations[9] (https://www.msbi.nl/cosegregation/) was not accessible at the time of this analysis.

Real-life pedigrees (Fig. S2) with a *PALB2* variant NM_024675.3(PALB2):c.3113G>A (NP_078951.2:p.Trp1038Ter) were obtained from the University of Cambridge (Cambridge, UK) and the Huntsman Cancer Institute (Salt Lake City, UT, USA) after removing protected health information. Pedigrees with *ATM* variants have been published previously.[23] Breast, ovarian, and pancreatic cancer risks associated with germline *PALB2* pathogenic variants were obtained from Yang et al.[16] Risks of female breast cancer and pancreatic cancer conferred by *ATM* pathogenic variants were obtained from Thompson et al.[24]

To evaluate the accuracy of cosegregation methods, we simulated pedigrees with pathogenic or benign germline variants by SLINK.[25,26] This program took the abovementioned *ATM* pedigrees as input, then simulated genotypes based on a provided penetrance model (Appendix S2). We then calculated a full-likelihood Bayes factor using the same penetrance as that in the simulation. Based on the Bayes factor to ACMG/AMP strength level conversion method,[12] we calculated that PP1_moderate was met if the Bayes factor was ≥4.33. BS4_moderate was met if the Bayes factor was ≤0.23, the reciprocal of the corresponding PP1_moderate Bayes factor. We also tested the performance of using a more stringent threshold, 18.7 for PP1_strong and 0.053 for BS4_strong.[12] We did 100,000 simulations to calculate the probability of meeting either PP1 or BS4 by Bayes factor cosegregation as a function of relative risk and number of pedigrees.

RESULTS

Limitations of meiosis counting for complex diseases

In the current ACMG/AMP guidelines, evidence against cosegregation is not considered by the PP1 rules (Table S1). Moderate-amount evidence against pathogenicity for a variant is ignored by both PP1 and BS4 rules, for example, when the variant is present in one or two unaffected individuals. This will lead to a systematic bias toward an inflated type I error when pedigrees are small. Also, meiosis counting may falsely consider nonexistent evidence for pathogenicity if the age of onset is not handled properly. It has been known that relative risks for breast cancer conferred by *BRCA1* and *BRCA2* are high at a young age and consistently decrease with age.[13–15] For people in their 80s, germline variants have little to no effect on disease risk (Fig. S1). Therefore, diagnosis with cancer at 80 years of age provides little evidence *for* pathogenicity. On the other hand, absence of cancer before 80 years of age in individuals carrying a variant provides evidence *against* pathogenicity. Take the hypothetical pedigree from Fig. 1 as an example: if individual II-1 is affected and carries the variant, and if we do not consider her age of onset, meiosis counting incorrectly upgrades cosegregation from 5 meioses to 6 meioses (Table 1). A simple solution to provide more realistic weights is to count only early-onset cases, by a stringent definition of early-onset disease in which pathogenic variants would confer a high relative risk. However, this approach would fail to recognize the evidence for pathogenicity, albeit it lower, provided by individuals affected at a slightly older age than the early-onset cutoff. More importantly, it still does not solve the problem of ignoring moderate evidence against pathogenicity. Note that diagnosis at old age is not treated as unaffected by the current ACMG/AMP rules and hence will not be counted toward BS4. Furthermore, meiosis counting does not consider sex, phenocopy, incomplete penetrance, allele frequency, and multiple diseases associated with the same gene (see below). Therefore, meiosis counting is not suitable for complex diseases like cancers.

Table 1 Analysis results of the hypothetical pedigree in Fig. 1.

Affection status and genotype of individual II-1 in Fig. 1	Meiosis counting[a] Number of meioses	Full-likelihood Bayes factor following different Bayesian analytical models		
		Cumulative risk	Incidence rate	Survival
Affected, variant present	6	46.5	21.8	3.85
Unaffected, variant present	5	8.50	6.66	3.85
Affected, variant absent	5	7.78	21.8	19.7
Unaffected, variant absent	5	43.5	34.0	19.7
Affected, variant untested	5	27.1	21.8	11.8

Cumulative risk, incidence rate, and survival refer to the different liability class models (see text). Affection status refers to breast cancer only. Untested: genetic test not conducted.

[a]Meiosis counting is based on the American College of Medical Genetics and Genomics/Association for Molecular Pathology (ACMG/AMP) rules, in which only affected individuals are considered, while absence of the variant of interest in unaffected individuals is not treated as informative meioses.[1]

Limitations of the cumulative risk and incidence rate models

We analyzed a hypothetical pedigree (Fig. 1) to demonstrate the performance of different penetrance models for Bayes factor cosegregation analysis. We varied cancer and genotype status for individual II-1, aged 81. Table 1 summarizes results from the analysis of the pedigree considering different genotype and affection status for individual II-1, using full-likelihood Bayes under the cumulative risk, incidence rate, and survival models (Fig. S3). We set the relative risk to 1 for age group 80–90. Therefore, if a model is correct, we expect the following: (1) Bayes factors are the same within each genotype stratum (variant presence or absence) irrespective of affection status, because affection status of individual II-1 does not provide evidence for or against pathogenicity; (2) the variant-present stratum has a Bayes factor lower than the variant-absent stratum, because presence of the variant in unaffected individuals is evidence against pathogenicity, while absence of the variant in unaffected individuals is complementary evidence that the variant cosegregates with the disease. As shown in Table 1, these characteristics were observed only from the survival model.

Using the incidence rate model for *BRCA1*, the Bayes factor is 6.66 if the individual carries the variant and has no cancer before 81, but 21.8 if the individual develops breast cancer at age 81 years. The difference in Bayes factor can be explained by the fact that in the first scenario, this unaffected individual provides evidence against pathogenicity, while in the second scenario, she does not provide any evidence against or for pathogenicity. To demonstrate this point, note that the Bayes factor (21.8) when this individual has an unknown genotype is the same as that from the second scenario. This problem stems from the lack of consideration of hidden data in the pedigree. The definition of likelihood is the probability of observed *data* given a hypothesis. In the case of individual II-1 when she is affected, the data are (1) she is affected at 81 years old, and (2) she is not affected *before* 81 years old. The incidence rate model only considers the first part, leaving out the second. The fact is, the second part of the data is evidence against variant pathogenicity and should be considered in a causality analysis.

Results also demonstrate a concerning limitation of the cumulative risk model. Not only does this model fail to recognize the evidence against pathogenicity provided by the fact that II-1 did not have cancer before 81 years old, but it also treats an affected II-1 as evidence for variant pathogenicity even though there is no such evidence in the data. Therefore, when individual II-1 is affected and carries the variant, the cumulative risk model yielded a considerably higher Bayes factor (46.5) than the survival model (3.85), even higher than that observed for the incidence rate model (21.8). It should be noted that the difference between Bayes factors of 3.85 and either 21.8 or 46.5 is considerable; this would translate to a change of strength category from PP1 to PP1_strong following the criteria proposed by Tavtigian et al.[12]

Because the full-likelihood Bayes and cosegregation likelihood ratio methods implemented via AnalyzeMyVariant use a liability class model based on cumulative risk, they showed similar behaviors as the cumulative risk model calculated by our in-house program (Table S3). Slight differences in Bayes factor values are probably due to the differences in relative risk and background incidence rate used. Note also that the AnalyzeMyVariant implementation of these two methods does not allow cancer site–specific liability classes, which is another concern in cosegregation analysis (see below).

Consideration of pleiotropy, population, and birth year in Bayes factor cosegregation

Pathogenic variants in some hereditary cancer genes are associated with increased risks of cancer at multiple sites (Table S2), a phenomenon known as pleiotropy.[27] Both background disease risk in the general population and the risk elevation conferred by genes vary by cancer site (Fig. S1). Accordingly, liability classes should be further divided into subgroups by cancer sites, in addition to age, sex, and affection status. Therefore, any tool that applies one penetrance model for multiple cancer sites is losing information, as well as potentially assigning incorrect weights for the data provided. The extent to which this incorrect strategy will affect variant classification will differ depending on individual pedigree structures.
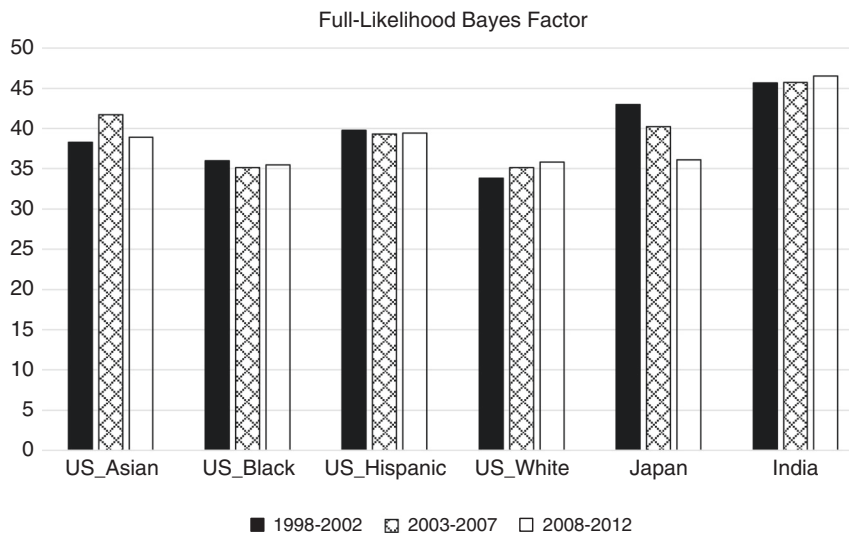
Fig. 2 Cosegregation results based on different population incidence.

To demonstrate this issue, we analyzed two real-life pedigrees with a *PALB2* variant NM_024675.3(PALB2): c.3113G>A (NP_078951.2:p.Trp1038Ter) (Fig. S2). One of these pedigrees had a pancreatic cancer case who carried the variant (Pedigree B). With a survival model that includes separate liability classes for breast, ovarian, and pancreatic cancer, using the full-likelihood Bayes approach, we obtained a Bayes factor of 5.25 combining two pedigrees (PP1_moderate). When we used pancreatic cancer liability classes only, the Bayes factor reduced to 3.99 (PP1_supporting). This is expected because the relative risk for breast cancer is much higher than pancreatic cancer. When we used breast cancer liability classes only, the Bayes factor was overestimated at 11.7. Although still falling in the moderate evidence category, there is potential for such overestimation to compound with the analysis of multiple pedigrees. Therefore, using a liability class model limited to one cancer site is not recommended as it can incorrectly estimate the strength of evidence to be applied for variant classification.

Another issue to consider is that cancer incidence rates differ between races and countries, and they change over time. Using incidence rates from one population for the analysis of pedigrees from another population will lead to systematic bias. Although the difference may be small, systematic bias accrues with an increasing number of pedigrees included in analysis. This is an emerging problem as genetic testing is becoming available to a broader range of populations. To solve this issue, we developed a program to automatically generate a liability class model based on a user's choices of population, cohort, and relative risk. Population incidence rate is obtained from the Cancer Incidence in Five Continents (CI5),[22] which has data from all cancer registries in the world at five-year intervals. To demonstrate the importance of using appropriate cancer incidence rates, we analyzed the hypothetical pedigree in Fig. 1 (setting individual II-1 as unaffected and untested for the variant) based on differing population incidence rates. We observed that the Bayes factor for this pedigree varied from 33.8 to 46.6 between populations, a non-negligible difference for variant classification (Fig. 2). The differences among US_White, Japan, and India are substantial, suggesting the importance of using population-specific incidence. Also, the declining Bayes factor for Japan corresponds to the increasing incidence in Japanese, demonstrating the importance of considering cohort in some populations.

Limitations of cosegregation analysis for moderate-risk genes

Moderate-risk genes,[28] such as *CHEK2* and *ATM*, are included in most multigene cancer panel tests. The smaller effect size of pathogenic variants in these genes, about a twofold increased risk of cancer, generates a challenge for cosegregation analysis.

To evaluate the impact of an effect size smaller than that conferred by *BRCA1*, *BRCA2*, or *PALB2*, we simulated data with a pathogenic variant associated with a relative risk ranging from 2 to 40. Results show that cosegregation would be helpful (>80% power) when the relative risk is around 10 and at least seven pedigrees are observed, while the power to reach PP1 is very low (23%) with seven pedigrees when the increased risk is at twofold. (Fig. 3a). Variants with a smaller effect size have a numerically higher rate of meeting BS4 than those with a higher relative risk (Fig. 3b). Bayes factor cosegregation using a lower relative risk model demonstrates a lower probability of meeting BS4 (Fig. 3c) and a higher rate of meeting PP1 (Fig. 3d) on benign variants. These results show that cosegregation analysis using a Bayes factor approach for moderate-risk genes can be both powerless and misleading. The same trend is also observed for the strong category of PP1 and BS4 (Fig. 3e–h), where the false BS4 (Fig. 3f) and false PP1 rates are below 5% (Fig. 3h).
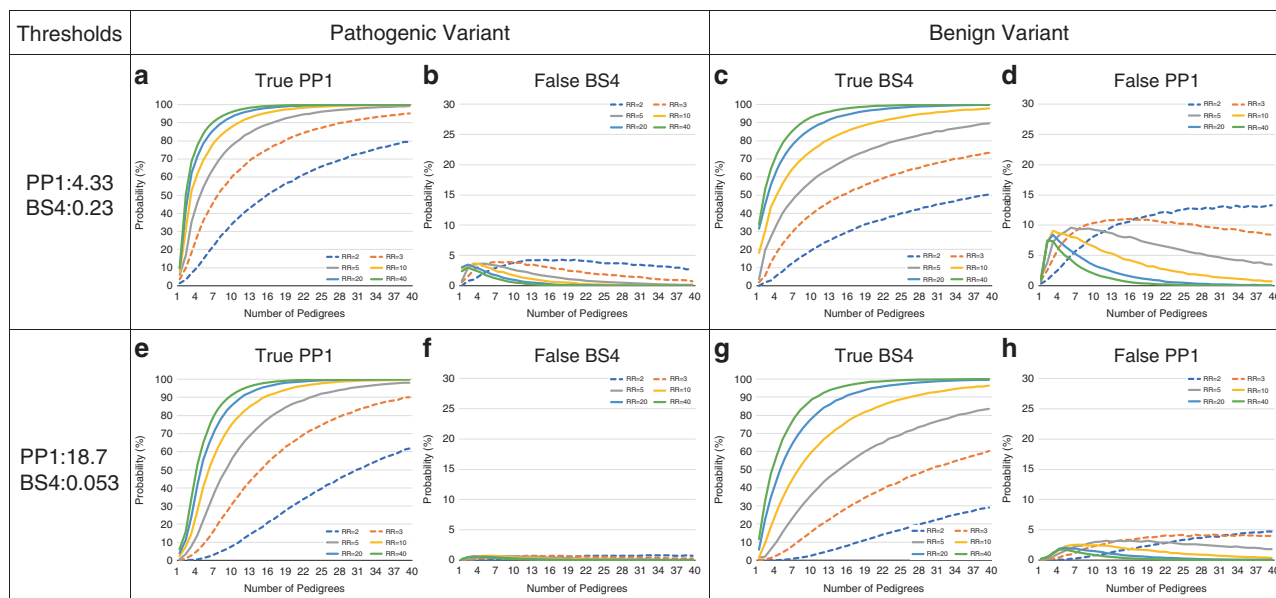
Fig. 3 Accuracy of cosegregation analysis as a function of variant effect size. Upper panels are moderate categories; lower panels are strong categories. (**a**, **b**, **e**, **f**) Probability of meeting PP1 or BS4 when variants are pathogenic. (**c**, **d**, **g**, **h**) Probability of meeting BS4 or PP1 when variants are benign. Relative risks in these four panels refer to parameters in cosegregation analysis. *BS4* benign category by cosegregation, *PP1* pathogenic category by cosegregation, *RR* relative risk, *Thresholds* Bayes factor cutoff values for meeting PP1 or BS4.

We also analyzed real-life pedigrees with moderate-risk *ATM* variants (Table S4) to assess the practicality of using cosegregation data for the classification of variants in genes like *ATM*. We excluded the variant NM_000051.3(ATM):c.7271T>G (NP_000042.2:p.Val2424Gly), which is associated with a high risk of breast cancer.[23] There were 12 pedigrees with eight unique variants, and only two variants were observed in more than one pedigree. No variant reached the PP1 supportive level based on Bayes factor. If we combined these pedigrees in one analysis as if they carry the same variant, the resulting Bayes factor was 0.51, a value that falls into the uninformative range (0.48–2.08).[12]

DISCUSSION

We demonstrate that meiosis counting is not suitable for variant classification in many common cancer-associated genes due to the violation of assumptions inherent in meiosis counting. These assumptions include complete penetrance, no phenocopies, no age or sex difference in disease risk conferred by the variant, and the same elevated risk to different cancer sites. Modifying the method to be used for complex diseases, such as relaxing nonsegregation criteria and filtering individuals by age, will lead to other problems such as systematically omitting moderate-level evidence against pathogenicity and difficulty in determining an appropriate age cutoff.

A Bayes factor approach can solve these problems when a reasonably accurate penetrance estimation is available. However, the current implementations of the Bayes factor approaches have limitations. Notably, neither the cumulative risk nor the incidence rate penetrance model accurately represents pathogenicity evidence in data. We propose a survival penetrance model that addresses these

limitations and show that it yields results consistent with expectation.

We show that the use of different penetrance models can result in significant differences when analyzing a single pedigree using a Bayes factor approach, which can be sufficient to result in different strength categories in the context of variant classification using ACMG/AMP guidelines. Logically, using different penetrance models may thus impact the final classification of individual variants and the clinical management of individuals who carry them. It is thus vital to encourage the usage of appropriate gene-specific penetrance by researchers, genetic testing laboratories, and variant curation panels. This should preferably include efforts to standardize between institutes the input information regarding age-, sex-, and site-dependent relative risk conferred by a given cancer predisposition gene.

The survival liability class model presented here is implemented as a freely available webtool named COOL version 2 (Co-segregation Online, http://BJFengLab.org/) described in the Supplementary Information (Appendix S1). The platform currently provides relative risk tables for 16 different cancer risk genes (Table S2), and permits the use of a customized relative risk table for other cancer genes, or using a customized penetrance file for noncancer genes. It should be noted that three genes, *ATM*, *CHEK2*, and *MEN1*, are annotated as inappropriate for cosegregation analysis due to current evidence indicating an association with a moderate (<5-fold) risk of cancer. We have also demonstrated the importance of selecting appropriate incidence rates based on population and birth year. We provide easy access to this information through the website. However, even with population-specific incidence, the analysis of non-European

populations assumes that relative risks are constant across populations. A large-scale study has demonstrated that the disease risks of *BRCA1*, *BRCA2*, *PALB2*, *ATM*, and *CHEK2* in Japanese were comparable with Europeans.[29] Similar studies in other populations will inform future testing and variant classification processes in non-Europeans. For complex diseases characterized by variable age at onset and pleiotropy, another crucial consideration is age of each family member. For cancer-associated genes, the proper approach may depend on the degree to which treatment for the first cancer affects risk of developing cancer at other sites. We would propose to use a model in which each family member is followed up until the diagnosis of first cancer, risk-reducing treatment, or the last observation, whichever occurred first.[16]

Lastly, we have demonstrated the difficulty of applying standard cosegregation approaches for assessing causality of moderate-penetrance alleles, with very low power to meet PP1 even when data are available from a large number of pedigrees. We also show that this approach is likely to provide misleading evidence, reaching BS4 for a true pathogenic variant. This suggests that there will be little value in allocating resources to genotyping of relatives for segregation analysis of variants of uncertain significance in moderate-penetrance genes.

In summary, we provide a method to remedy a problem that is ubiquitous in the widely used liability class model. More efforts are required to estimate the age-, sex-, site-, and population-specific relative risk conferred by cancer-associated genes to improve the accuracy of segregation analysis results.

SUPPLEMENTARY INFORMATION

The online version of this article (https://doi.org/10.1038/s41436-020-0920-4) contains supplementary material, which is available to authorized users.

ACKNOWLEDGEMENTS

DISCLOSURE

The PERCH software, of which B.-J.F. is the inventor, has been nonexclusively licensed to Ambry Genetics Corporation for their clinical genetic testing services and research. B.-J.F. reports funding and sponsorship to his institution on his behalf from Pfizer Inc., Regeneron Genetics Center LLC., and Astra Zeneca. B.-J.F. received data from Ambry Genetics for research. The other authors declare no conflicts of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17:405–423.
2. Jarvik GP, Browning BL. Consideration of cosegregation in the pathogenicity classification of genomic variants. Am J Hum Genet. 2016; 98:1077–1081.
3. Kelly MA, Caleshu C, Morales A, et al. Adaptation and validation of the ACMG/AMP variant classification framework for MYH7-associated inherited cardiomyopathies: recommendations by ClinGen's Inherited Cardiomyopathy Expert Panel. Genet Med. 2018;20:351–359.
4. Lee K, Krempely K, Roberts ME, et al. Specifications of the ACMG/AMP variant curation guidelines for the analysis of germline CDH1 sequence variants. Hum Mutat. 2018;39:1553–1568.
5. Mester JL, Ghosh R, Pesaran T, et al. Gene-specific criteria for PTEN variant curation: recommendations from the ClinGen PTEN Expert Panel. Hum Mutat. 2018;39:1581–1592.
6. Luo X, Feurstein S, Mohan S, et al. ClinGen Myeloid Malignancy Variant Curation Expert Panel recommendations for germline RUNX1 variants. Blood Adv. 2019;3:2962–2979.
7. ClinGen TP53 Expert Panel. TP53 rule specifications for the ACMG/AMP variant curation guidelines. August 2019. https://clinicalgenome.org/site/assets/files/3876/clingen_tp53_acmg_specifications_v1.pdf.
8. Thompson D, Easton DF, Goldgar DE. A full-likelihood method for the evaluation of causality of sequence variants from family data. Am J Hum Genet. 2003;73:652–655.
9. Mohammadi L, Vreeswijk MP, Oldenburg R, et al. A simple method for cosegregation analysis to evaluate the pathogenicity of unclassified variants; BRCA1 and BRCA2 as an example. BMC Cancer. 2009;9:211.
10. Goldgar DE, Easton DF, Deffenbaugh AM, et al. Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. Am J Hum Genet. 2004;75:535–544.
11. Goldgar DE, Easton DF, Byrnes GB, et al. Genetic evidence and integration of various data sources for classifying uncertain variants into a single model. Hum Mutat. 2008;29:1265–1272.
12. Tavtigian SV, Greenblatt MS, et al. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. Genet Med. 2018;20:1054–1060.
13. Kuchenbaecker KB, Hopper JL, Barnes DR, et al. Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. JAMA. 2017;317:2402.
14. Mocci E, Milne RL, Mendez-Villamil EY, et al. Risk of pancreatic cancer in breast cancer families from the breast cancer family registry. Cancer Epidemiol Biomarkers Prev. 2013;22:803–811.
15. Antoniou AC, Cunningham AP, Peto J, et al. The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. Br J Cancer. 2008;98:1457–1466.
16. Yang X, Leslie G, Doroszuk A. et al. Cancer risks associated with germline PALB2 pathogenic variants: an international study of 524 families. J Clin Oncol. 2020;38:674–685.
17. Rañola JMO, Liu Q, Rosenthal EA, Shirts BH. A comparison of cosegregation analysis methods for the clinical setting. Fam Cancer. 2018;17:295–302. 10.
18. Easton DF, Bishop DT, Ford D, Crockford GP. Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. The Breast Cancer Linkage Consortium. Am J Hum Genet. 1993;52:678–701.
19. Terwilliger JD, Ott J. Handbook of human genetic linkage. Baltimore: Johns Hopkins University Press; 1994.
20. Lovelock PK, Spurdle AB, Mok MTS, et al. Identification of BRCA1 missense substitutions that confer partial functional activity: potential moderate risk variants? Breast Cancer Res. 2007;9:R82.

ARTICLE

21. Parsons MT, Tudini E, Li H, et al. Large scale multifactorial likelihood quantitative analysis of BRCA1 and BRCA2 variants: an ENIGMA resource to support clinical variant classification. Hum Mutat. 2019;40:1557–1578.

22. Forman D, Bray F, Brewster DH, Gombe Mbalawa C, Kohler B, Piñeros M, Steliarova-Foucher E, Swaminathan R, Ferlay J, editors. Cancer incidence in five continents. Vol. X. IARC Scientific Publication No. 164. Lyon, France: International Agency for Research on Cancer; 2014. http://www.iarc.fr/en/publications/pdfs-online/epi/sp164/. Accessed July 2019.

23. Goldgar DE, Healey S, Dowty JG, et al. Rare variants in the ATM gene and risk of breast cancer. Breast Cancer Res. 2011;13:R73.

24. Thompson D, Duedal S, Kirner J, et al. Cancer risks and mortality in heterozygous ATM mutation carriers. J Natl Cancer Inst. 2005;97:813–822.

25. Ott J. Computer-simulation methods in human linkage analysis. Proc Natl Acad Sci U S A. 1989;86:4175–4178.

26. Weeks D, Ott J, Lathrop GM. SLINK: a general simulation program for linkage analysis. Am J Hum Genet. 1990;47:A204.

27. Bien SA, Peters U. Moving from one to many: insights from the growing list of pleiotropic cancer risk genes. Br J Cancer. 2019;120:1087–1089.

28. Spurdle AB, Greville-Heygate S, Antoniou AC, et al. Towards controlled terminology for reporting germline cancer susceptibility variants: an ENIGMA report. J Med Genet. 2019;56:347–357.

29. Momozawa Y, Iwasaki Y, Parsons MT, et al. Germline pathogenic variants of 11 breast cancer genes in 7,051 Japanese patients and 11,241 controls. Nat Commun. 2018;9:4083.