EXPANDING PREVENTIVE BREAST CANCER GENETICS:

FROM HIGH TO MODERATE RISK AND FROM

BREAST TO PANCREATIC CANCER

by

Erin Lynn Young

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Oncological Sciences

The University of Utah

December 2016

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Erin Lynn Young**

has been approved by the following supervisory committee members:

Sean Vahram Tavtigian	, Chair	**July 12, 2016** Date Approved
Joshua David Schiffman	, Member	**July 12, 2016** Date Approved
Deborah W. Neklason	, Member	**August 4, 2016** Date Approved
Nicola J. Camp	, Member	**July 12, 2016** Date Approved
Alun William Thomas	, Member	**July 12, 2016** Date Approved

and by **Bradley R. Cairns** , Chair/Dean of

the Department/College/School of **Oncological Sciences**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Genetic testing for pathogenic variants resulting in increased breast cancer susceptibility is a useful tool in the clinic when determining the medical management of individuals with family histories of breast cancer. Once a pathogenic variant has been identified, healthy biological relatives can be testing for the pathogenic variant, and carriers can benefit from preemptive medical or surgical risk reduction strategies. Eligibility for genetic testing can be complicated; incorporating ages of onset for cancer diagnosis (≤ 50 years, for most cancers), cancer incidence across multiple generations, or biological relatives with multiple diagnoses of primary cancers. Recent advancements in technology, notably next-generation sequencing (NGS) applications in the clinic, have allowed the detection of pathogenic variants in a selection or panel of genes. This increased clinical sensitivity has increased the ability to identify those at increased risk for cancer, as well as the need to understand the genes and variants involved.

Missense substitutions constitute an appreciable fraction of the burden of pathogenic variants, but are difficult to interpret. Many missense analysis programs have been developed as tools to estimate the deleteriousness of variants. Gene panels for cancer susceptibility include genes with differing levels of associated risk. For breast cancer panels, the high-risk genes include *BRCA1*, *BRCA2*, *PALB2*, and *TP53*. Carriers of pathogenic variants in high-risk breast cancer susceptibility genes have clear guidelines for clinicians. In order to determine the percentage of women impacted by

pathogenic variants in moderate-risk breast cancer genes, variants in nine moderate-risk breast cancer susceptibility genes from previously published studies were graded for severity via frequency, protein location, and *in silico* predictions from Align-GVGD, MAPP, CADD, and Polyphen-2. Potentially pathogenic rare variants were grouped until an acceptable stratification for variants of uncertain significance (VUS) was established. We observed that 7.5% of early-onset breast cancer cases- versus 2.4% controls, were estimated to carry a medically actionable variant in at least one of the nine moderate-risk breast cancer genes.

Genes associated with hereditary breast and ovarian cancer (HBOC) and colorectal cancer (CRC) susceptibility have been shown to play a role in pancreatic cancer susceptibility. Germline genetic testing of pancreatic cancer cases could be beneficial for at-risk relatives with pathogenic variants in established HBOC and CRC genes. In pancreatic cancer cases ascertained at the Huntsman Cancer Hospital, but unselected for family history, 7.6-8.5% carried a variant that would alter the screening recommendations for at-risk relatives. A meta-analysis including additional published studies revealed that approximately 11.9% of unselected pancreatic cancer cases carry a pathogenic variant in HBOC or CRC susceptibility genes. The inclusion of both HBOC and CRC susceptibility genes in a panel test for pancreatic cancer cases finds a high enough percentage of carriers to rationalize genetic testing of these patients for HBOC and CRC susceptibility.

TABLE OF CONTENTS

LIST OF TABLES

Tables

LIST OF FIGURES

Figures

CHAPTER 1

INTRODUCTION

Background

Genetic testing for pathogenic variants resulting in increased cancer susceptibility is a useful tool in the clinic when determining the medical management of individuals with family histories of cancer.[1–4] The results from a genetic test can be used to rationalize a change in medical management. Medical management can be altered for earlier detection of a cancer through intensifying surveillance options and/or reducing the age of screening, or lower the risk of cancer incidence through measures such as preventive surgery.[3,5] Once a pathogenic variant in a family has been identified, healthy biological relatives can be tested, and carriers can benefit from preemptive surveillance.[3,5,6] As personalized medicine continues to evolve, genetic testing may be able to guide cancer treatments for carriers that develop cancer.[7,8]

Eligibility for genetic testing is based primarily off of a family history for a specific, or a collection of specific cancers.[1–4,9–14] Eligibility, then, can be complicated incorporating ages of onset for cancer diagnosis (≤ 50 years, for most cancers), cancer incidence across multiple generations, or biological relatives with multiple diagnoses of primary cancers.[1–5,9,13,14] Because of the current criteria, inclusion is biased towards individuals who inherit pathogenic variants maternally, and complicated for those with

limited family histories, such as those from small families or those that are adopted.[15]

Recent advancements in technology, notably next generation sequencing (NGS) applications in the clinic, have allowed the detection of pathogenic variants in a selection or panel of genes, rather than just a singular gene.[16–20] This increased clinical sensitivity, i.e., the increased ability to detect more variants, has increased the ability to identify those at increased risk for cancer, but has also increased the need to understand the genes and variants involved.[20–27] The increase in the amount of the genome screened increases the number of variants of uncertain significance (VUS).[20–26] VUS are variants that have been identified in an individual, but have an unknown functional impact. VUS, then, range from benign or innocuous to pathogenic. There are six main classes of potentially deleterious variants:

1. Nonsense substitutions and frameshift indels.
2. Large indels that alter one or more exons.
3. Variants that destroy the core canonical GT-AG or rare instance or AT-AC splice junction sequences.
4. Missense substitutions that fall at evolutionarily conserved sites in the protein.
5. Nucleotide substitutions falling within the splice donor or splice acceptor. consensus sequences but outside of the core dinucleotides.
6. Exonic substitutions that create strong splice donors or acceptors.
7. Nucleotide substitutions altering the functionality of regulatory or epigenetic regions.

Missense substitutions, nonsense substitutions located in the final exon of a protein, and in-frame indels are difficult to interpret, and routine medical management may not be appropriate under uncertain circumstances for VUS carriers. Many missense

analysis programs have been developed as tools to estimate the deleteriousness of variants. Admittedly, many of these missense analysis programs have similar algorithms used in their labeling strategies.[28] Still, these missense analysis programs might be used to distinguish subsets of variants with elevated probabilities of pathogenicity. When calibrated against observational data or else independently generated variant classifications, a prior probability score can be combined with clinical data to establish whether a variant is pathogenic, and, thereby, medically actionable.[29–35]

Incidental findings are pathogenic variants identified in genes not related to the reason for ordering the sequencing, but still have medical utility. These are genes with clear enough guidelines for clinicians to influence medical management, even without a complete family history. Gene panels for cancer susceptibility include genes with differing levels of associated risk. For breast cancer panels, the standard high-risk genes *BRCA1*, *BRCA2*, and *TP53* are generally included.[3,36,37] *BRCA1* and *BRCA2* confer a 65-85% and 45-85% lifetime risk of breast cancer, respectively.[38–40] *TP53* increases risk to a wide spectrum of cancers. Carriers of pathogenic variants in high-risk breast cancer susceptibility genes are labeled by the American College of Medical Genetics (ACMG) as reportable as incidental findings.[41–45]

PALB2 confers a similar high-risk of breast cancer, but is not yet labelled as a gene with reportable incidental findings.[36,46–49] Moderate-risk breast cancer genes are those that confer only a 2-4 fold increase of risk for carriers.[37,50–55] Many of these genes are newly identified and the lifetime risks of pathogenic variants, and the identification of pathogenic variants is not as readily available. *ATM* and C*HEK2,* both moderate-risk breast cancer genes, have recently been added into NCCN's guideline of medically actionable genes with guidelines,[1] but the surrounding medical management of carriers of

pathogenic variants in other moderate-risk genes is not clear.[16,21] VUS are especially common for moderate-risk genes because missense substitutions have not been as extensively studied.[53] Although the identification and addition of new breast cancer susceptibility genes is not a problem in and of itself, translating incomplete information into the clinic setting does result in confusion and possible mismanagement.[17,20,26]

In order to determine the percentage of women impacted by pathogenic variants in moderate-risk breast cancer genes, preascertained variants in nine moderate-risk breast cancer susceptibility genes from previously published studies were graded for severity via frequency, protein location, and *in silico* predictions from Align-GVGD, MAPP, CADD, and Polyphen-2.[56–65] Common variants (> 0.1% frequency) and those too internally intronic to be predicted to impact splicing were outgrouped as lacking pathogenic potential. Potentially pathogenic rare variants were placed into one of several groups. The first group contained variants that were considered clearly pathogenic: truncating and splice junction variants that were expected to trigger nonsense mediated decay. The second group contained variants with interpretations that were not as clear: rare missense substitutions, in-frame deletions, and last-exon protein truncating variants. This second group was further divided by *in silico* analysis of missense substitutions via the score of each missense substitution with each missense analysis program. Once an acceptable stratification for VUS was established, it was applied to real populations to estimate the number of carriers of pathogenic variants in moderate-risk breast cancer susceptibility genes.

This same sample set of women were genotyped for 18 Breast Cancer Association Consortium (BCAC)-confirmed SNPs. Risk estimates from each SNP were combined to determine a final normalized polygene score (NPS) for each woman's polygenotype.[66]

These NPS were stratified to determine the subpopulation of carriers not of individual variants, but of polygenotypes that were predicted to increase risk enough to be medically actionable.

The inclusion of moderate-risk breast cancer susceptibility genes increases the number of pathogenic variants identified.[67] This increase in actionable test results could be high enough to expand the eligibility of genetic testing. Using the VUS stratification method mentioned above, unselected breast, ovarian, and pancreatic cancer cases were selected solely on their personal cancer diagnosis and family's inclusion in the Utah Population Database.[68] A panel consisting of 34 genes was used to identify known and potential medically actionable variants in order to estimate the percentage of carriers among cases alone. This was repeated with a 59 gene panel on additional pancreatic cancer cases, selected sequentially at the Huntsman Cancer Hospital.

To sequence our panel genes, we used a barcode-then-pool strategy. In short, DNA is sheared through sonication and adaptors and barcodes are ligated onto the DNA fragments in subsequent steps. The barcoded genomic libraries are then pooled for subsequence capture, and super-pooled for sequencing. Generally, 30X coverage from paired-end reads is considered sufficiently deep to detect heterozygous positions, but we were concerned about the evenness of exon representation following targeted exon capture. Our goal, therefore, was to achieve 100X coverage per nucleotide in sequence coverage. Reads were mapped to the genome using GATK best practices.[69] Both single nucleotide substitutions and short indels were extracted from the mapped reads. Since we were specifically expecting rare variants and common variants were irrelevant to our main analysis, variants were cross-referenced against dbSNP, the Thousand Genomes Project data, and ExAC.[70–72] Variants with estimated allele frequencies greater than 0.1%

in any population were set aside, and did not contribute to this study's analyses of genetic susceptibility. Many of the "found" rare variants could have been screening artifacts. Variants with an increased probability of pathogenicity, were confirmed through sanger sequencing.[73] Once the variants were confirmed in our unselected populations, they were combined into a meta-analysis with other studies to explore the proportions of the population carrying pathogenic variant.

Aims

1. Develop a method of stratification of genetic markers to estimate the number of women carrying medically actionable breast cancer susceptibility variants outside of the high-risk spectra.
    a. Isolate subgroups of missense substitutions using *in silico* methods via pathogenicity prediction from missense analysis programs.
    b. Combine risk estimates from BCAC-confirmed SNPs and determine a polygenic score that delineates subgroups of women.
2. Combine genotyping of clearly pathogenic variants with the developed stratification for VUS to estimate the number of unselected breast, ovarian, and pancreatic cancer cases with medically actionable variants identified from a panel of genes.
    a. Identify medically actionable variants in breast, ovarian, or colorectal cancer susceptibility genes in unselected breast and ovarian cancer cases.
    b. Identify medically actionable variants in breast, ovarian, or colorectal cancer susceptibility genes in unselected pancreatic cancer cases.
    c. Combine pancreatic cancer screening studies to estimate the number of

pancreatic cancer cases carrying medically actionable variants in breast, ovarian, and colorectal cancer genes.

References

1. Daly, M. B. *et al.* Genetic/familial high-risk assessment: breast and ovarian, version 2.2015. *J. Natl. Compr. Canc. Netw.* **14,** 153–62 (2016).

2. Lancaster, J. M., Powell, C. B., Chen, L.-M. M. & Richardson, D. L. Statement on risk assessment for inherited gynecologic cancer predispositions. *Gynecol. Oncol.* **136,** 3–7 (2014).

3. Hall, M. J. *et al.* Genetic testing for hereditary cancer predisposition: BRCA1/2, Lynch syndrome, and beyond. *Gynecol. Oncol.* (2016). doi:10.1016/j.ygyno.2016.01.019

4. Robson, M., Storm, C., Weitzel, J., Wollins, D. & Offit, K. American society of clinical oncology policy statement update: genetic testing for cancer susceptibility. *J. Clin. Oncol.* **21,** 2397–406 (2003).

5. Lancaster, J. M., Powell, C. B., Chen, L. M. & Richardson, D. L. Society of gynecologic oncology statement on risk assessment for inherited gynecologic cancer predispositions. *Gynecol. Oncol.* **136,** 3–7 (2015).

6. Plon, S. E. *et al.* Genetic testing and cancer risk management recommendations by physicians for at-risk relatives. *Genet. Med.* **13,** 148–54 (2011).

7. Lord, C. J., Tutt, A. N. J. & Ashworth, A. Synthetic lethality and cancer therapy: lessons learned from the development of PARP inhibitors. *Annu. Rev. Med.* **66,** 455–70 (2015).

8. Lord, C. J. & Ashworth, A. BRCAness revisited. *Nat. Rev. Cancer* 1–11 (2016). doi:10.1038/nrc.2015.21

9. American Society of Clinical Oncology. Statement of the American Society of Clinical Oncology: genetic testing for cancer susceptibility, Adopted on February 20, 1996. *J. Clin. Oncol.* **14,** 1730-6-40 (1996).

10. Lauby-Secretan, B. *et al.* Breast-cancer screening - viewpoint of the IARC working group. *N. Engl. J. Med.* **372,** 2353–2358 (2015).

11. Evaluation of genomic applications in practice and prevention (EGAPP) working group. Recommendations from the EGAPP working group: genetic testing strategies in newly diagnosed individuals with colorectal cancer aimed at reducing

morbidity and mortality from Lynch syndrome in relatives. *Genet. Med.* **11,** 35–41 (2009).

12. MacArthur, D. G. *et al.* Guidelines for investigating causality of sequence variants in human disease. *Nature* **508,** 469–76 (2014).

13. Marth, C. *et al.* AGO Austria recommendations for genetic testing of patients with ovarian cancer. *Wien. Klin. Wochenschr.* **127**, 652–4 (2015).

14. Syngal, S. *et al.* ACG clinical guideline: genetic testing and management of hereditary gastrointestinal cancer syndromes. *Am. J. Gastroenterol.* **110,** 223–262 (2015).

15. McCuaig, J. M. *et al.* Breast and ovarian cancer: the forgotten paternal contribution. *J. Genet. Couns.* **20,** 442–449 (2011).

16. Walsh, T. *et al.* Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 12629–33 (2010).

17. Easton, D. F. *et al.* Gene-panel sequencing and the prediction of breast-cancer risk. *N. Engl. J. Med.* **372,** 2243–57 (2015).

18. Hall, M. J., Forman, A. D., Pilarski, R., Wiesner, G. & Giri, V. N. Gene panel testing for inherited cancer risk. *J. Natl. Compr. Canc. Netw.* **12,** 1339–46 (2014).

19. Salo-Mullen, E. E. *et al.* Identification of germline genetic mutations in patients with pancreatic cancer. *Cancer* **121,** 4382–8 (2015).

20. Shirts, B. H. *et al.* Improving performance of multigene panels for genomic analysis of cancer predisposition. *Genet. Med.* **18**, 974-81 (2016).

21. Rainville, I. R. & Rana, H. Q. Next-generation sequencing for inherited breast cancer risk: counseling through the complexity. *Curr. Oncol. Rep.* **16,** 371 (2014).

22. Kurian, A. W. *et al.* Clinical evaluation of a multiple-gene sequencing panel for hereditary cancer risk assessment. *J. Clin. Oncol.* **32,** 2001–2009 (2014).

23. Howarth, D. R. *et al.* Initial results of multigene panel testing for hereditary breast and ovarian cancer and Lynch syndrome. *Am. Surg.* **81,** 941–4 (2015).

24. Mauer, C. B., Pirzadeh-Miller, S. M., Robinson, L. D. & Euhus, D. M. The integration of next-generation sequencing panels in the clinical cancer genetics practice: an institutional experience. *Genet. Med.* **16,** 1–6 (2013).

25. Yurgelun, M. B. *et al.* Identification of a variety of mutations in cancer predisposition genes in patients with suspected Lynch syndrome. *Gastroenterology*

149, 604e20-613e20 (2015).

26. Thompson, E. R. *et al.* Panel testing for familial breast cancer: calibrating the tension between research and clinical care. *J. Clin. Oncol.* **34**, 1455-9 (2016).

27. Desmond, A. *et al.* Clinical actionability of multigene panel testing for hereditary breast and ovarian cancer risk assessment. *JAMA Oncol.* **2114,** 1–9 (2015).

28. Tavtigian, S. V. Comparison of programs for in silico assessment of missense substitutions. *Hum. Mutat.* **32,** v (2011).

29. Vallée, M. P. *et al.* Classification of missense substitutions in the BRCA genes: a database dedicated to Ex-UVs. *Hum. Mutat.* **33,** 22–8 (2012).

30. Thompson, B. & Spurdle, A. Microsatellite instability use in mismatch repair gene sequence variant classification. *Genes (Basel).* **6,** 150–162 (2015).

31. Thompson, B. a. *et al.* A multifactorial likelihood model for MMR gene variant classification incorporating probabilities based on sequence bioinformatics and tumor characteristics: a report from the colon cancer family registry. *Hum. Mutat.* **34,** 200–9 (2013).

32. Vallée, M. P. *et al.* Adding in silico assessment of potential splice aberration to the integrated evaluation of BRCA gene unclassified variants. *Hum. Mutat.* (2016). doi:10.1002/humu.22973

33. Thompson, B. a *et al.* Calibration of multiple in silico tools for predicting pathogenicity of mismatch repair gene missense substitutions. *Hum. Mutat.* **34,** 255–65 (2013).

34. Tavtigian, S. V, Greenblatt, M. S., Lesueur, F. & Byrnes, G. B. In silico analysis of missense substitutions using sequence-alignment based methods. *Hum. Mutat.* **29,** 1327–36 (2008).

35. Tavtigian, S. V, Byrnes, G. B., Goldgar, D. E. & Thomas, A. Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. *Hum. Mutat.* **29,** 1342–54 (2008).

36. Lerner-Ellis, J., Khalouei, S., Sopik, V. & Narod, S. A. Genetic risk assessment and prevention: the role of genetic testing panels in breast cancer. *Expert Rev. Anticancer Ther.* **15,** 1315–26 (2015).

37. Stadler, Z. K., Schrader, K. a, Vijai, J., Robson, M. E. & Offit, K. Cancer genomics and inherited risk. *J. Clin. Oncol.* **32,** 687–98 (2014).

38. Ford, D. *et al.* Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. *Am. J. Hum. Genet.* **62,** 676–689 (1998).

39. Antoniou, a *et al.* Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: a combined analysis of 22 studies. *Am. J. Hum. Genet.* **72,** 1117–1130 (2003).

40. King, M.-C., Marks, J. H., Mandell, J. B. & New York Breast Cancer Study, G. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. *Science (80-.).* **302,** 643–646 (2003).

41. May, T. On the justifiability of ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *J. Law. Med. Ethics* **43,** 134–42 (2015).

42. Smith, L. a, Douglas, J., Braxton, A. a & Kramer, K. Reporting incidental findings in clinical whole exome sequencing: incorporation of the 2013 ACMG recommendations into current practices of genetic counseling. *J. Genet. Couns.* **24,** 654–662 (2014).

43. Evans, B. J. Minimizing liability risks under the ACMG recommendations for reporting incidental findings in clinical exome and genome sequencing. *Genet. Med.* **15,** 915–920 (2013).

44. Green, R. C. *et al.* ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet. Med.* **15,** 565–574 (2013).

45. Allyse, M. & Michie, M. Not-so-incidental findings: the ACMG recommendations on the reporting of incidental findings in clinical whole genome and whole exome sequencing. *Trends Biotechnol.* **31,** 439–441 (2013).

46. Evans, M. K. & Longo, D. L. PALB2 mutations and breast-cancer risk. *N. Engl. J. Med.* **371,** 566–8 (2014).

47. Rahman, N. *et al.* PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat. Genet.* **39,** 165–7 (2007).

48. Southey, M. C., Teo, Z. L. & Winship, I. PALB2 and breast cancer: ready for clinical translation! *Appl. Clin. Genet.* **6,** 43–52 (2013).

49. Antoniou, A. C. *et al.* Breast-cancer risk in families with mutations in PALB2. *N. Engl. J. Med.* **371,** 497–506 (2014).

50. Apostolou, P. & Fostira, F. Hereditary breast cancer: the era of new susceptibility genes. *Biomed Res. Int.* **2013,** 747318 (2013).

51. Filippini, S. E. & Vega, A. Breast cancer genes: beyond BRCA1 and BRCA2. *Front. Biosci.* **18,** 1358–1372 (2013).

52. Kean, S. The ‘other’ breast cancer genes. *Science* **343,** 1457–9 (2014).

53. Tavtigian, S. V & Chenevix-Trench, G. Growing recognition of the role for rare missense substitutions in breast cancer susceptibility. *Biomark. Med.* **8,** 589–603 (2014).

54. Byrnes, G. B., Southey, M. C. & Hopper, J. L. Are the so-called low penetrance breast cancer genes, ATM, BRIP1, PALB2 and CHEK2, high risk for women with strong family histories? *Breast Cancer Res.* **10,** 208 (2008).

55. Narod, S. A. Testing for CHEK2 in the cancer genetics clinic: ready for prime time? *Clin. Genet.* **78,** 1–7 (2010).

56. Tavtigian, S. V *et al.* Rare, evolutionarily unlikely missense substitutions in ATM confer increased risk of breast cancer. *Am. J. Hum. Genet.* **85,** 427–46 (2009).

57. Le Calvez-Kelm, F. *et al.* Rare, evolutionarily unlikely missense substitutions in CHEK2 contribute to breast cancer susceptibility: results from a breast cancer family registry case-control mutation-screening study. *Breast Cancer Res.* **13,** R6 (2011).

58. Damiola, F. *et al.* Rare key functional domain missense substitutions in MRE11A, RAD50, and NBN contribute to breast cancer susceptibility: results from a breast cancer family registry case-control mutation-screening study. *Breast Cancer Res.* **16,** R58 (2014).

59. Le Calvez-Kelm, F. *et al.* RAD51 and breast cancer susceptibility: no evidence for rare variant association in the breast cancer family registry study. *PLoS One* **7,** e52374 (2012).

60. Park, D. J. *et al.* Rare mutations in RINT1 predispose carriers to breast and Lynch syndrome-spectrum cancers. *Cancer Discov.* **4,** 804–15 (2014).

61. Park, D. J. *et al.* Rare mutations in XRCC2 increase the risk of breast cancer. *Am. J. Hum. Genet.* **90,** 734–9 (2012).

62. Tavtigian, S. V *et al.* Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J. Med. Genet.* **43,** 295–305 (2006).

63. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46,** 310–5 (2014).

64. Stone, E. a & Sidow, A. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Res.* **15,** 978–86 (2005).

65. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using Polyphen-2. *Curr. Protoc. Hum. Genet.* **Chapter**

7, Unit7.20 (2013).

66. Michailidou, K. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.* **45,** 353–61, 361–2 (2013).

67. LaDuca, H. *et al.* Utilization of multigene panels in hereditary cancer predisposition testing: analysis of more than 2,000 patients. *Genet. Med.* **16,** 830–7 (2014).

68. Skolnick, M. A. R. K. The Utah genealogical database: a resource for genetic epidemiology. *Banbury Rep.* **4,** 285–297 (1980).

69. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43,** 491–8 (2011).

70. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).

71. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29,** 308–11 (2001).

72. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* **536,** 285-91 (2015).

73. Johnston, H. R., Hu, Y. & Cutler, D. J. Population genetics identifies challenges in analyzing rare variants. *Genet. Epidemiol.* **39,** 145–148 (2015).

CHAPTER 2

MULTIGENE TESTING OF MODERATE-RISK GENES: BE MINDFUL OF THE MISSENSE

Young EL, Feng BJ, Stark AW, Damiola F, Durand G, Forey N, Francy TC, Gammon A, Kohlmann WK, Kaphingst KA, McKay-Chopin S, Nguyen-Dumont T, Oliver J, Paquette AM, Pertesi M, Robinot N, Rosenthal JS, Vallee M, Voegele C, Hopper JL, Southey MC, Andrulis IL, John EM, Hashibe M, Gertz J, Breast Cancer Family Registry, Le Calvez-Kelm F, Lesueur F, Goldgar DE, Tavtigian S V. Multigene testing of moderate-risk genes: be mindful of the missense. *J Med Genet.* 2016 Jun;53(6):366-76. doi: 10.1136/jmedgenet-2015-103398. Epub 2016 Jan 19.

ORIGINAL ARTICLE

Multigene testing of moderate-risk genes: be mindful of the missense

E L Young,[1] B J Feng,[2] A W Stark,[1] F Damiola,[3] G Durand,[4] N Forey,[4] T C Francy,[1] A Gammon,[5] W K Kohlmann,[5] K A Kaphingst,[6] S McKay-Chopin,[4] T Nguyen-Dumont,[7] J Oliver,[8] A M Paquette,[1] M Pertesi,[4] N Robinot,[4] J S Rosenthal,[1] M Vallee,[9] C Voegele,[4] J L Hopper,[10,11] M C Southey,[6] I L Andrulis,[12] E M John,[13,14] M Hashibe,[15] J Gertz,[1] Breast Cancer Family Registry,[13] F Le Calvez-Kelm,[4] F Lesueur,[16] D E Goldgar,[2] S V Tavtigian[1]

▸ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/jmedgenet-2015-103398).

For numbered affiliations see end of article.

Correspondence to
Professor Sean V Tavtigian, Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, Utah 84112, USA; sean.tavtigian@hci.utah.edu

Received 17 July 2015
Revised 17 December 2015
Accepted 18 December 2015
Published Online First 19 January 2016

To cite: Young EL, Feng BJ, Stark AW, *et al*. *J Med Genet* 2016;**53**:366–376.

ABSTRACT

Background Moderate-risk genes have not been extensively studied, and missense substitutions in them are generally returned to patients as variants of uncertain significance lacking clearly defined risk estimates. The fraction of early-onset breast cancer cases carrying moderate-risk genotypes and quantitative methods for flagging variants for further analysis have not been established.

Methods We evaluated rare missense substitutions identified from a mutation screen of *ATM*, *CHEK2*, *MRE11A*, *RAD50*, *NBN*, *RAD51*, *RINT1*, *XRCC2* and *BARD1* in 1297 cases of early-onset breast cancer and 1121 controls via scores from Align-Grantham Variation Grantham Deviation (GVGD), combined annotation dependent depletion (CADD), multivariate analysis of protein polymorphism (MAPP) and PolyPhen-2. We also evaluated subjects by polygenotype from 18 breast cancer risk SNPs. From these analyses, we estimated the fraction of cases and controls that reach a breast cancer OR≥2.5 threshold.

Results Analysis of mutation screening data from the nine genes revealed that 7.5% of cases and 2.4% of controls were carriers of at least one rare variant with an average OR≥2.5. 2.1% of cases and 1.2% of controls had a polygenotype with an average OR≥2.5.

Conclusions Among early-onset breast cancer cases, 9.6% had a genotype associated with an increased risk sufficient to affect clinical management recommendations. Over two-thirds of variants conferring this level of risk were rare missense substitutions in moderate-risk genes. Placement in the estimated OR≥2.5 group by at least two of these missense analysis programs should be used to prioritise variants for further study. Panel testing often creates more heat than light; quantitative approaches to variant prioritisation and classification may facilitate more efficient clinical classification of variants.

INTRODUCTION

For the last 15+ years, most clinical cancer genetics involving predisposition to breast (and ovarian) cancer have been driven by mutation screening of *BRCA1* and *BRCA2*. For these two genes, the ratio of truncating and splice junction variants (T+SJV) to pathogenic rare missense substitutions (rMS) is about 10:1,[1 2] fostering a view among clinical cancer geneticists that rMS are only a minor part of the spectrum of breast cancer predisposing variants.

Following the discoveries of *BRCA1* and *BRCA2*, many additional genes have been identified as breast cancer susceptibility genes. A prominent group of these are referred to as moderate-risk susceptibility genes because protein truncating variants and severely dysfunctional missense substitutions in them appear to confer, on average, twofold to fivefold increased risk of breast cancer. This magnitude of increased risk is less dramatic than that conferred by most pathogenic alleles in the high-risk genes *BRCA1*, *BRCA2* and *PALB2*, but potentially high enough to influence the medical management of carriers.[3–6] Beyond the moderate-risk genes, many common SNPs have been identified as markers for slightly increased breast cancer risk.[7–9] A challenge posed by these modest-risk SNPs is that, individually, they do not confer enough risk to influence the medical management of a carrier, but considered as an ensemble they may.

In our previously published case–control mutation screening studies of *ATM*, *CHEK2*, *MRE11A*, *NBN*, *RAD50*, *RAD51*, *RINT1* and *XRCC2*,[10–15] we repeatedly found, albeit with some variations in the methodology, that the summed frequency of predicted deleterious missense substitutions exceeded that of protein truncating variants. This study used the same ethnically diverse sample of 1297 breast cancer cases and 1121 controls, negative for pathogenic variants in *BRCA1*, *BRCA2* or *PALB2* (table 1). Here, we added *BARD1* mutation screening data to the original gene-by-gene analyses and applied consistent analytic models across the rare variants from all nine genes. Setting an OR≥2.5 as a threshold for clinical significance, we estimated the scores from the missense substitution analysis programs Align-GVGD,[16] CADD,[17] MAPP[18] and PolyPhen-2[19] required to identify a group of missense substitutions that reach an average OR≥2.5. These results were used to determine the proportion of cases and controls carrying a potential risk-conferring rMS. We also explored a combined evaluation of 18 Breast Cancer Association Consortium (BCAC)-confirmed

Table 1 Distribution of cases and controls by age, race/ethnicity and study centre

	Case–control mutation screening for rare variants in nine moderate-risk genes				Case–control SNP genotyping for 18 BCAC SNPs			
Distributions	**Case**	**%**	**Control**	**%**	**Case**	**%**	**Control**	**%**
Age range, years								
≤30	106	8.2	67	6.0	97	7.8	61	5.8
31–35	319	24.6	171	15.3	300	24.3	157	14.9
36–10	433	33.4	238	21.2	409	33.1	220	20.8
41–45	439	33.9	203	18.1	430	34.8	183	17.3
46–50	0	0	230	20.5	0	0	225	21.3
51–55	0	0	212	18.9	0	0	211	20.0
Race/ethnicity								
Caucasian	840	64.8	967	86.3	788	63.8	904	85.5
East Asian	202	15.6	71	6.3	193	15.6	70	6.6
Latina	158	12.2	47	4.2	158	12.8	47	4.5
Recent African Ancestry	97	7.5	36	3.2	97	7.9	36	3.4
Study centre								
BCFR-Australia	588	45.4	522	46.6	551	44.6	472	44.7
BCFR-Canada	299	23.1	463	41.3	284	23.0	499	42.5
BCFR-Northern California	410	31.6	136	12.1	401	32.4	136	12.9
Total	1297		1121		1236		1057	

Subjects were excluded from mutation-screening if performance was poor; percentage data are the total number of cases on control DNA in the category indicated that met the mutation-screening quality control standards.
BCAC, Breast Cancer Association Consortium; BCFR, Breast Cancer Family Registry.

modest-risk SNPs as a polygene, compared predicted to empirically observed ORs and estimated the prevalence of genotype combinations across these 18 SNPs with an average OR≥2.5. Our data set is unique in that the moderate-risk gene mutation screening and SNP genotyping were performed on the same subjects, giving us the opportunity to compare prevalence of the OR≥2.5 threshold across T+SJV and OR≥2.5 groupings of rMS or normalised polygene score (NPS).

MATERIALS AND METHODS

Subjects

Patients were selected from women systematically recruited by population-based sampling by the Australian, Northern Californian and Ontarian sites of the Breast Cancer Family Registry (BCFR). Patients were recruited between 1995 and 2005. The selection criteria for cases (N=1297) were diagnosis at or before age 45 years and self-reported race/ethnicity plus grandparents' country of origin consistent with Caucasian, East Asian, Hispanic/Latino or Recent African racial or ethnic heritage.[20] The controls (N=1121) were frequency matched to cases within each centre on racial or ethnic group, with age at selection not more than ±10 years difference from the age range at diagnosis of the patients systematically recruited from the same centre.

Mutation screening and SNP genotyping

Mutation screening was as described previously[10–15] and is included in online supplementary methods, as is SNP genotyping. The following methods focus on the analysis of missense substitutions and of the ensemble of 18 modest-risk SNPs.

Allele frequency threshold

Following our allele frequency analysis of *ATM, BRCA1, BRCA2* and *CHEK2* from Damiola *et al*,[12] we applied a minor allele frequency (q) threshold of ≤0.1%, based on exome variant server and 1000 genomes project allele frequency data that are independent of this study's mutation screening, for all variants of the eight genes in which biallelic truncating variants are often either embryonic lethal or else cause a highly deleterious phenotype from the ataxia telangiectasia/Fanconi anaemia spectrum. Biallelic *CHEK2* carriers are superficially healthy, and our analysis suggested a cut-off of q<0.32% for that gene.[12]

In silico missense substitution scoring

Align-GVGD (agvgd.iarc.fr/agvgd_input.php) and MAPP (mendel.stanford.edu/SidowLab/downloads/MAPP/index.html) require user-supplied protein multiple sequence alignments (pMSAs) to score missense substitutions; both compare the physicochemical features of the missense residue to the physicochemical range of variation at the relevant position in the pMSA to calculate their scores. Align-GVGD produces a score with seven discrete grades from C0 (most likely neutral) to C65 (most likely deleterious). MAPP, which additionally requires a phylogenetic tree detailing the evolutionary relationships and distances between the organisms with sequences represented in the pMSA, outputs a continuous variable, the MAPP score.

For programs that require a user-generated pMSA, it has been suggested that the pMSA for each gene needs enough variation to average at least three amino acid substitutions per position (3S/P).[21] For each gene, we created an initial pMSA containing the human sequence and 13 additional orthologs. To maintain harmony across the pMSAs, orthologs were sampled from a phylogenetically similar set of organisms ranging from a non-human primate (*Macaca mulatta*) to the non-chordate deuterostomate *Strongylocentrotus purpuratus* (see details in online supplementary methods).

Ortholog sequences downloaded from GenBank were aligned using the expresso extension of T-Coffee to create the initial pMSA.[22 23] The initial alignment was checked by hand in Geneious V.7.1.4 (http://www.geneious.com) for anomalies that might be attributed to gene model errors rather than actual

sequence divergence. Potential anomalies were corrected by reference to, and gene reprediction from, genomic DNA sequence available on the UCSC genome browser (http://genome.ucsc.edu).

Routines from the PHYLIP package (V.3.69),[24] constrained using the known phylogeny of the species included in our alignments, were used to estimate substitutions per position within each alignment and to calculate the distance matrices required by MAPP. The complete alignments, phylogenetic trees and distances are available upon request.

PolyPhen-2 (genetics.bwh.harvard.edu/pph2/)[25] and CADD (cadd.gs.washington.edu/)[17] operate without user-supplied alignments. PolyPhen-2 uses a combination of internally generated pMSAs, functional annotations and structural information to evaluate missense substitutions;[19] we used its output variable 'pph2_prob' as a continuous variable score. CADD uses a series of 63 gene annotations, combined through a support vector machine linear kernel, to define a PHRED-like score (their 'scaled C-score') for all possible single-nucleotide substitutions and small insertion–deletion mutations to the human genome.[17]

Although CADD has a built-in method for short indels, the other missense analysis programs do not. For Align-GVGD, MAPP and PolyPhen-2, nonsense substitutions in the final exon and non-frameshift indels received the score of the most severe missense substitution possible in the affected interval. Variant pathogenicity scores are summarised in online supplementary table S1.

Statistics

To assess evidence of risk from the case–control frequency distribution of T+SJV and rMS, we constructed a table with one entry per subject; the variants per subject; and annotations for whether the variant was in a key functional domain (see online supplementary table S2), its frequency, as well as study centre, case–control status, race/ethnicity and age for the subject. In-frame deletions (IFDs) were treated as rMS. For the subjects who carried more than one rare variant of interest, only the most deleterious score was considered. We then divided the subjects into groups: a reference group of non-carriers and carriers of common variants (only), carriers of rMS not in a functional domain, carriers of rMS in a key functional domain divided into two groups via score and carriers of T+SJVs. For each of the four rMS analysis programs, we toggled the program's severity score from a very relaxed to a very stringent value. We repeatedly estimated two ORs as the stringency increased: the OR for subjects that carried one or more rMS at or above the score (and no T+SJV), and the OR for subjects who carried an rMS that was below the score (and carried neither a T+SJV nor a higher scoring rMS). From this analysis, we determined a threshold severity score for each program at which subjects carrying an rMS at or above the threshold had an average OR≥2.5.

For the 18 SNP polygene, we created a polygenic risk score (PRS) by multiplying together the appropriate published OR estimate from each individual SNP genotype.[7 9 26 27] The geometric mean of the PRS of the controls was used to normalise the PRS into an NPS. Because risk estimates from Caucasian populations may not be applicable to women from other race/ethnicities, we gave each non-Caucasian subject a race/ethnicity-specific risk estimate derived from the population of the subject,[26 28–30] and normalised each race/ethnicity separately. Risk estimates of rs1045485 could not be found for the non-Caucasian subjects in this study, so the risk estimate based on Caucasian populations was used for all race/ethnicities. In instances where a Latina-specific risk estimate could not be found, we used the average between Caucasian and East Asian race/ethnicities. To determine the correlation between NPS and the observed OR, we grouped the subjects into a series of 10 contiguous bins based on percentile, using the central quintile (40–60 percentile) as the reference group. We treated groups outside of the reference as categorical variables for OR calculations. For the threshold analysis, we used the same reference group and adjusted the NPS threshold until the group containing scores above the set threshold had an OR≥2.5.

For the regressions, NPS was treated as the independent variable and the resultant OR of each group as the dependent variable weighted to the number of individuals in each group, excluding subjects in the middle quintile. p Values were found by testing the regression coefficient equal to 0 or 1. To combine the risk estimates from the rMS and NPS, we multiplied the NPS and OR from the rMS.

All analyses were performed using multivariable unconditional logistic regression using Stata V.12.1 software (StataCorp, College Station, Texas, USA). Adjustments were made for race/ethnicity and study centre, unless otherwise noted.

RESULTS

Initial evaluation of rare variants

From mutation screening of 1297 cases and 1121 controls, we observed 22 T+SJV, 9 IFDs and 196 rMS with minor allele frequencies <0.32% for *CHEK2* and <0.1% for the remaining genes. T+SJVs falling before the final exon were considered pathogenic and were associated with an OR of 3.32 (p=0.0023, table 2). Nonsense mutations located in the final exon were considered as IFDs.

The National Comprehensive Cancer Network (NCCN) and the American College of Radiology (ACR) recommend screening beginning at age 30 years and offering breast MRI in addition to mammograms for women with a ≥20% lifetime breast cancer risk.[31 32] The American Cancer Society (ACS) recommends breast MRI for women with a 20–25% or greater lifetime risk.[33] In the USA, the lifetime risk of a woman to develop breast cancer is estimated to be 12.3%;[34] however, this figure is an overestimate for our purposes because it includes women who are at high risk because of inherited mutations in genes such as *BRCA1* or *BRCA2*, or very strong family history. For a woman with minimal risk factors, for example, age at menarche ≥14 years, first childbirth at age ≤20 years and no family history, the Gail model[35] and Tyrer–Cuzick model[36] suggest a lifetime risk of 6.9% and 11% for developing breast cancer, respectively. If we assume that the average of these two estimates (9%) is approximately correct, carriage of a genotype conferring a 2.5-fold increase of risk, even in this low-risk population, would result in a lifetime risk estimate exceeding the NCCN, ACR and American Cancer Society (ACS) medically actionable threshold of a 20% lifetime risk. Subject to formal variant classification, carriers may then qualify, under current recommendations, for early mammography and/or enhanced screening with breast MRI. We note that threshold for intensified screening may be higher in other countries.

Considering all rMS as a group, we obtained a risk estimate that was elevated but that did not reach an OR≥2.5 threshold (OR=1.42, p=0.0091, table 2). We focused our analyses of rMSs to those that are relatively likely to impact key functions. This grouping included all of the rMS from the relatively small proteins encoded by *CHEK2*, *RAD51*, *RINT1* and *XRCC2*. Noting the structural similarity between BARD1 and BRCA1,

Table 2 Combined OR estimates from case–control mutation screening of nine moderate-risk genes

Analysis	Distinct variants	Control	%	Case	%	Adjusted OR*	CI	p Value
Non-carrier	148	998	89.0	1094	84.4	Reference		
Carrier of truncating or splice junction variant	22	9	0.8	27	2.1	3.31	1.53 to 7.16	2.36×10^{-3}
Rare missense substitution analyses								
Carrier of rare missense substitution	205	114	10.2	176	13.6	1.42	1.09 to 1.84	8.70×10^{-3}
Carrier of key domain rare missense substitution	140	65	5.8	136	10.5	1.94	1.41 to 2.67	5.11×10^{-5}
Carrier of non-key domain rare missense substitution	65	49	4.4	40	3.1	0.74	0.48 to 1.16	0.1926
Key domain rMS: MAPP								
rMS<11	63	39	3.5	58	4.5	1.47	0.95 to 2.26	0.0818
rMS≥11	77	26	2.3	78	6.0	2.63	1.65 to 4.21	5.32×10^{-5}
Key domain rMS : Align-GVGD								
rMS<C35	93	54	4.8	87	6.7	1.58	1.09 to 2.27	0.0146
rMS≥C35	47	11	1.0	49	3.8	3.62	1.84 to 7.13	2.03×10^{-4}
Key domain rMS : CADD								
rMS<23	97	50	4.5	88	6.8	1.66	1.14 to 2.41	0.0082
rMS≥23	43	15	1.3	48	3.7	2.87	1.57 to 5.26	6.40×10^{-4}
Key domain rMS : PolyPhen-2								
rMS<0.9	61	37	3.3	55	4.2	1.50	0.96 to 2.34	0.0752
rMS≥0.9	79	28	2.5	81	6.3	2.49	1.58 to 3.92	7.88×10^{-5}
Overlap of missense analysis programs								
One or more	89	34	3.0	93	7.2	2.37	1.57 to 3.60	4.56×10^{-5}
Two or more	65	18	1.6	70	5.4	3.18	1.85 to 5.46	2.68×10^{-5}
Three or more	45	14	1.2	52	4.0	3.27	1.77 to 6.04	1.51×10^{-4}
All four	19	2	0.2	20	1.5	8.61	1.96 to 37.81	4.35×10^{-3}
Total	375	1121		1297				

*Adjusted for race/ethnicity and study centre.
rMS, rare missense substitutions.

and that *BRCA1* pathogenic rMS are so far only known from the RING and BRCT domains,[37–40] we limited analyses of *BARD1* rMS to RING and BRCT domain substitutions. For the relatively large proteins encoded by *ATM, MRE11A*, *NBN* and *RAD50*, we focused analyses of rMS on the same key functional domains specified in our prior publications (see online supplementary table S2).[10 12] We observed 140 rMS and IFDs in key functional domains (OR 1.94, $p=5.1\times10^{-05}$, table 2), which still did not reach an OR≥2.5 threshold.

Grouping rMS to estimate risk and carrier rates

To define a higher-risk subset of rMS, we focused the next analyses on the three established moderate-risk genes: *ATM, CHEK2* and *NBN*.[41] There is no fully accepted method for rMS analysis. Instead of introducing a new method for variant classification, we used four existing missense analysis programs, Align-GVGD, CADD, MAPP and PolyPhen-2,[16–18 25] to assign severity scores to the key domain rMS from these genes. Align-GVGD was selected because its scores contribute to determination of prior probabilities of pathogenicity for key domain missense substitutions in *BRCA1* and *BRCA2*,[2] MAPP and PolyPhen-2 because of their strong performance in our recent analyses of mismatch repair protein missense substitutions,[42] and CADD because of its reported ability to prioritise variants across functional categories and effect sizes.[17] For variant evaluation, we adjusted our pMSAs to two depths: human through platypus (mammals only), and human through the organism required for 3S/P for each individual gene. Receiver operating characteristic (ROC) curves were generated for each method and depth (if applicable). Area under the curve (AUCs) were similar for all methods (see online supplementary table S3 and supplementary figure S1). The correlations between the missense analysis programs were highest between Polyphen-2 and CADD ($R^2=0.56$), but the R^2 for any combination of missense analysis programs was never >0.8, so none of the missense analysis programs were dropped from further analysis (see online supplementary table S4).

We then toggled the severity score for each of the four programs to find the lowest score where the OR for key domain rMS above the score reached at least 2.5 (figure 1A–D). The thresholds at which each of the four rMS analysis programs reached OR≥2.5 for key domain rMS were above a score of 11 for MAPP when using pMSAs consisting of organisms from human through 3S/P, C35 for Align-GVGD when using pMSAs consisting of organisms from human through 3S/P, 23 for CADD and 0.9 for PolyPhen-2 (see online supplementary table S5). It was interesting that neither Align-GVGD nor MAPP was able to achieve an OR≥2.5 with a pMSA that consisted only of mammals (human through platypus). It appears that these sequences, although generally more complete than those from more distant organisms, do not offer adequate variation to stratify variants. Examining the variants that were placed in the OR≥2.5 category by multiple missense analysis programs, we found that an overlap of at least two of the missense analysis programs resulted in a classification of variants with an OR≥2.5 (OR 2.59, p=0.0044; online supplementary table S6).

Applying the score thresholds determined from the *ATM-CHEK2-NBN* group to the key domain rMS observed in the remaining six less established moderate-risk genes, rMS ORs for the *BARD1-MRE11A-RAD50-RAD51-RINT1-XRCC2* group ranged from 2.41 (p=0.0078) using PolyPhen-2 to 4.86 (p=0.0129) using Align-GVGD (data not shown). We also found that concordance between at least two of the missense analysis programs resulted in a grouping of rMS with an OR≥2.5 (OR 4.90, p=0.0012; online supplementary table S6).

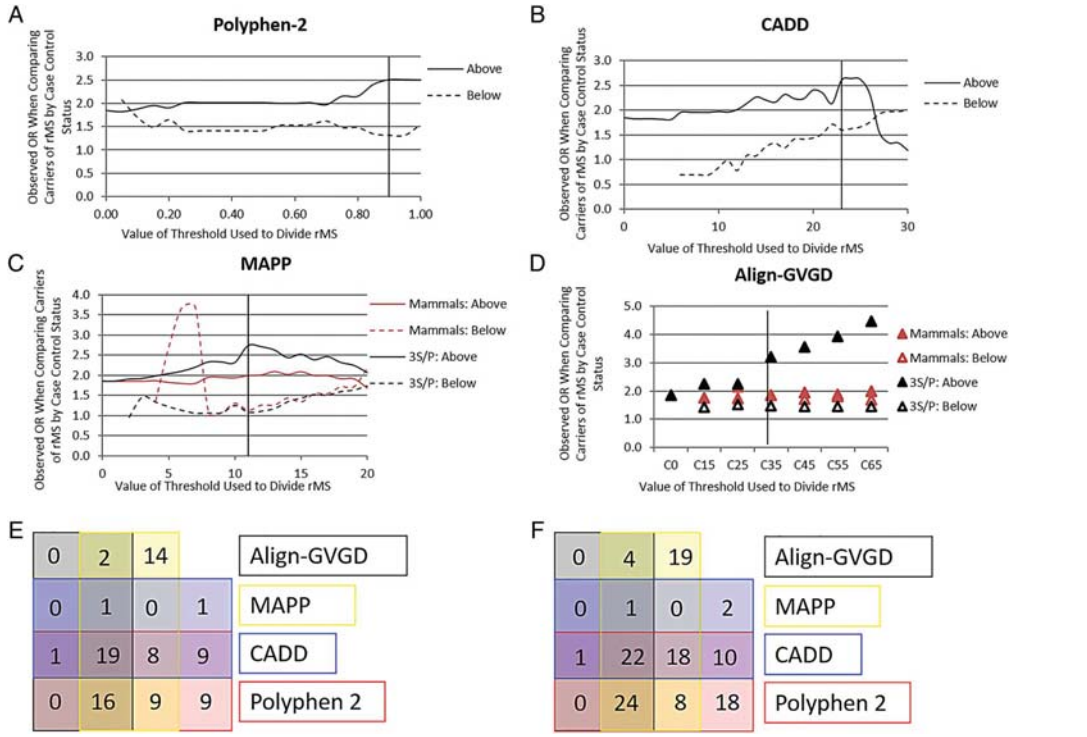


Figure 1 Observed OR and thresholds for each missense substitution analysis program. The observed OR for the carriers of key domain rare missense substitutions (rMS) of both the 'above' and 'below' groups for each severity threshold tested with (A) PolyPhen-2, (B) CADD, (C) MAPP and (D) Align-GVGD, adjusting for race/ethnicity and study centre for just the combined ensemble of ATM, CHEK2 and NBN. Vertical lines are indicative of the threshold for which the observed OR≥2.5. (E) A four-way Venn diagram detailing the number of rMS for which CADD, PolyPhen-2, Align-GVGD and MAPP placed in the 'above threshold' group for key domain rMS observed in all nine moderate-risk genes. (F) A four-way Venn diagram detailing the number of individuals for which CADD, PolyPhen-2, Align-GVGD and MAPP were placed in the 'above threshold' group for key domain rMS observed in all nine moderate-risk genes. Mammals=score obtained from using protein multiple sequence alignments (pMSA) containing sequences from human through platypus; 3S/P=scores are from gene-specific 3S/P depth pMSAs.

Having established that the thresholds identified with the *ATM-CHEK2-NBN* group were able to extract OR≥2.5 groupings from the remaining six genes, we used these thresholds to evaluate the proportions of cases and controls with above-threshold variants across the nine-gene ensemble (table 2). We found that 3.7–6.3% of cases and 1.0–2.5% of controls carried an above-threshold key domain rMS. Considering only the key domain rMS that were placed in an OR≥2.5 grouping by more than one of the missense analysis programs (figure 1E,F), concordance between two or more missense analysis programs was associated with an OR≥2.5 (OR 3.18, $p=2.68\times10^{-5}$), affecting 5.4% of cases and 1.6% of controls (table 2). These results are comparable to other studies, but include data from controls.[43–45] Comparing the proportion of above-threshold key domain rMS carriers to T+SJV carriers, rMS carriers appear to outnumber T+SJV carriers by a ratio of about 2.5:1.

Asking whether the results reported here are robust to the loss of any one gene from the less established moderate-risk gene set, we performed a series of analyses in which the genotype information of one of the genes was dropped and then the OR, rMS to T+SJV ratio, and carrier percentage was re-determined for the rMS from the remaining eight genes. We observed that, in each subset of eight genes, 3.0–5.9% of cases and 0.8–2.7% of controls were carriers of a variant from the above-threshold grouping, with a ratio of rMS to T+SJVs consistently over 1.6:1 for cases (see online supplementary table S7).

Common SNP-based polygene scores and above-threshold carrier rates

Generally, individual modest-risk SNPs do not confer enough risk to impact clinical practice. An attractive method for using SNPs in a clinical setting is to combine the risk estimates from multiple SNPs. Indeed, a recent large study combined risk estimates from 77 SNPs and found ORs≥2.5 at and above the 99th percentile of the combined scores.[9] We genotyped 18 BCAC-confirmed SNPs on the same subjects from the case–control mutation screening phase of this study (table 1). Using per-allele ORs from recent large studies,[7 26 28–30] we treated the SNPs as a polygene and created an NPS for each subject (see online supplementary table S8 and figure 2A).

To determine how closely the NPS predicted OR, we grouped the NPS scores into deciles and compared the mean NPS of each decile to its observed OR. With all subjects grouped together, the NPS correlated highly with the observed OR (coeff. =0.9232, $R^2=0.70$, p=0.0060) (figure 2B). Evaluating each race/ethnicity individually, Caucasians were the only group to achieve significance (coeff.=0.9835, $R^2=0.81$, p=0.0014) (figure 2C, data not shown), likely due to small sample sizes of the non-Caucasian groups. We also tested the alternate

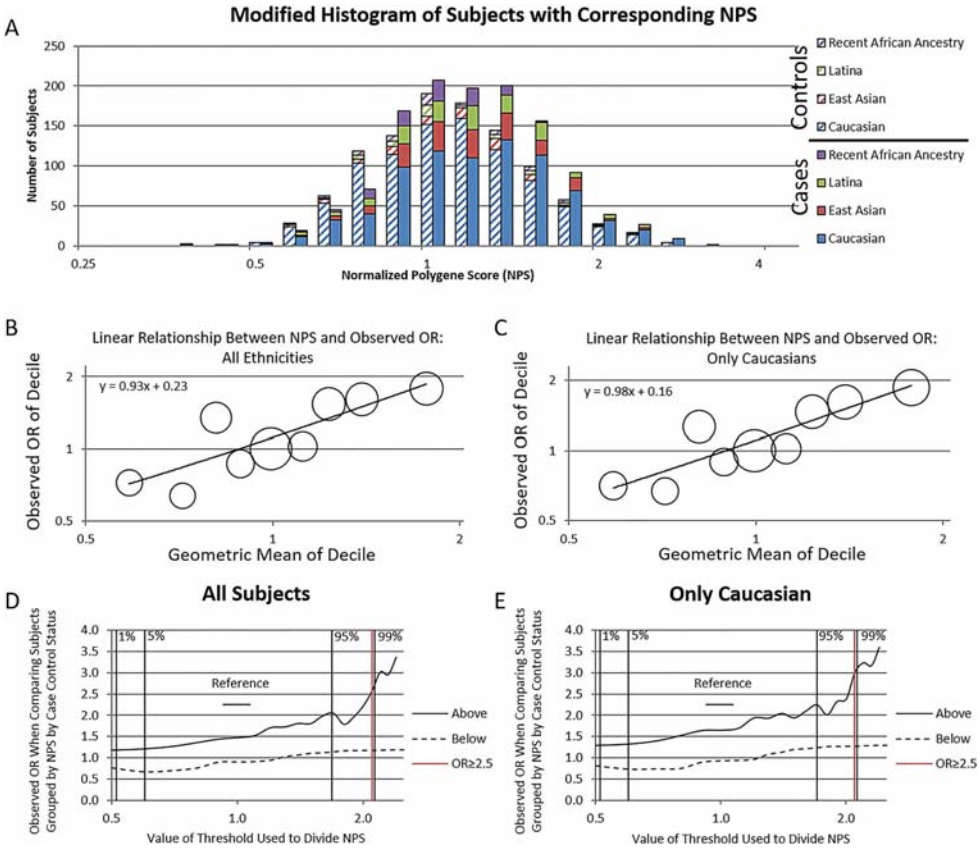


Figure 2 The normalised polygene score (NPS): distribution, NPS-OR correlation and threshold for medically actionable. (A) NPS distribution for all subjects. Comparison of observed OR and NPS for each decile for (B) all subjects and (C) only Caucasians, with the corresponding equations derived from linear regressions, excluding the central quintile. The observed OR when dividing subjects based on NPS, using the middle quintile as reference for (D) all subjects and (E) only Caucasians, adjusting for race/ethnicity and study centre. Vertical lines are indicative of the threshold for which the observed OR≥2.5, as well as the 1, 5, 95 and 99 percentiles. Bubble sizes are proportional to the number of subjects in each decile.

hypothesis, NPS=observed OR, and did not observe a significant difference (p=0.75 and 0.93 for all subjects and Caucasians, respectively).

Using the NPS, how many women are at a medically actionable risk? Using an approach analogous to that applied to the key domain rMS, we toggled the NPS to find the lowest score where the observed OR for the group of subjects with NPS above the score exceeded 2.5 (figure 2D–E). From the data, 2.1% of cases and 1.2% of controls carry a combination of SNPs such that they have a polygene score associated with an average OR≥2.5 (table 3). Limiting the analysis to Caucasians, we found that 3.2% of cases and 1.3% of controls have an NPS associated with an average OR≥2.5.

To explore the possibility of integrating gene mutation screening with SNP genotyping, we tested for interactions between carriage of an OR≥2.5 rare variant (combining T+SJVs and above-threshold rMS into a single group) and the NPS. In these tests, the interaction term never approached significance (p=0.52, 0.82, 0.51 and 0.96 for analyses with rMS scored by Align-GVGD, CADD, MAPP and PolyPhen-2, respectively). Accordingly, as the multiplicative OR model does appear to apply to combinations of rare variants from these nine genes with the NPS, we isolated the subjects who carried a rare variant in the OR≥2.5 category and multiplied the OR estimated from their rMS or T+SJV with their NPS. These combined ORs varied from ~1.0 to >5.0 (figure 3).

DISCUSSION

Neither pathogenic alleles in moderate-risk breast cancer susceptibility genes nor individual modest-risk breast cancer-associated SNPs confer the magnitude of risk of early-onset breast cancer conferred by pathogenic alleles in high-risk genes such as *BRCA1* and *BRCA2*. Nonetheless, under a generalised understanding of NCCN, ACR and ACS guidelines, a ≥2.5-fold increased risk of breast cancer is high enough to impact the medical management of otherwise healthy carriers. Across the nine moderate-risk susceptibility genes examined here, two classes of sequence variants meet or exceed this 2.5-fold risk threshold. 2.1% of cases carried a T+SJV, and these were associated with an OR of 3.32 (p=0.0023). Each of the four missense substitution analysis programs that we evaluated was able to define a set of key functional domain rMS that reached the OR≥2.5 threshold. 5.4% carried an rMS that two or more of the programs agreed was above the threshold, and this group of rMS was associated with an OR of 3.18 ($p=2.68\times10^{-5}$). In

Table 3 OR estimates and p-values for case–control screening of 18 Breast Cancer Association Consortium SNPs

	Control	%	Case	%	Adjusted OR	CI*	p>|z|
Utilizing NPS as a continuous score							
All	1057	100	1236	100			5.76×10^{-10}
Only Caucasians	904	85.5	788	63.8			3.47×10^{-10}
Excluding Caucasian	153	14.5	448	36.3			0.32
Number of subjects at risk indicated by NPS score above Threshold group							
All†	13	1.2	26	2.1	2.56	1.26 to 5.19	0.009
Only Caucasian†	12	1.3	25	3.2	3.02	1.45 to 6.29	0.003
Excluding Caucasian					Never	≥2.5	
OR of top and bottom percentiles using middle quintile as reference							
All							
0–1%	11	1.0	5	0.4	0.50	0.16 to 1.56	0.232
1–5%	42	4.0	28	2.3	0.72	0.42 to 1.24	0.239
95–99%	41	3.9	73	5.9	1.99	1.27 to 3.12	0.003
99–100%	11	1.0	22	1.8	2.74	1.27 to 5.90	0.010
Only Caucasian							
0–1%	10	1.1	3	0.4	0.48	0.13 to 1.81	0.278
1–5%	36	4.0	19	2.4	0.80	0.44 to 1.48	0.482
95–99%	36	4.0	51	6.5	2.00	1.22 to 3.26	0.006
99–100%	10	1.1	22	2.8	3.20	1.45 to 7.08	0.004

*Adjusted for study centre and race/ethnicity.
†NPS≥2.1. NPS, normalised polygene score.

addition, 2.1% of cases carried an above-threshold SNP polygene genotype.

Whether focusing on the confirmed moderate-risk genes *ATM*, *CHEK2* and *NBN* or looking at all nine genes, the ratio of carriers of T+SJV to above-threshold key domain rMS was in the range of 1:2–1:3. This finding is different than the ∼10:1 ratio observed in *BRCA1/2*; the preponderance of above-threshold rMS in these moderate-risk genes is much more reminiscent of the situation with*TP53*.[5] Because most rMS observed during clinical testing of these moderate-risk genes would be returned as variants of uncertain significance (VUS) in test reports, the relatively high proportion of above-threshold rMS reported here creates a challenge for test interpretation. During clinical counselling, to alleviate patient distress, observations of VUS rMS, especially in moderate-risk genes, are often downplayed as of minimal significance—'normalised'. However, at least for the nine genes that we examined, normalising the rMS amounts to disregarding approximately 2/3 of sequence variants with OR≥2.5 detectable by the genetic tests.

Within the logical structure of the analyses presented here, the OR≥2.5 threshold applied to rMS and SNP polygene groupings was a device used to align the analyses with current patient management standards. A consequence is that the OR point estimates reported for those groupings in tables 2 and 3 are circularly dependent on the threshold selected. Nonetheless, the following four key results are independent to the circular logic underlying those OR point estimates: (i) the a priori existence of groupings with OR≥2.5; (ii) the p values associated with those groupings; (iii) the ratios of subjects with T+SJVs, rMSs with OR≥2.5 and SNP polygene with OR≥2.5; and (iv) the frequencies among controls and early-onset cases of individuals with a genotype falling into one of these groupings. These findings all correspond to open, medically relevant questions.

Although we did not accompany this study with functional assays, a yeast complementation assay applied to 25 *CHEK2* missense substitutions included applicable Align-GVGD and PolyPhen-2 scores.[46] Among the six rMS with Align-GVGD and PolyPhen-2 scores meeting our severity criterion, the average activity was −0.062 (SD=0.027) in an assay where the internal wild-type and dysfunctional variants were given scores of 1.00 and 0.00, respectively. In contrast, the average score among the 19 rMS not meeting our concordant severity criterion was +0.472 (SD=0.388), resulting in a p value of 1.12×10^{-5} against the hypothesis that the two groups have the same mean activity. Moving forward, it will become important to develop methods that combine patient observational data with in silico and functional assay results towards clinical classification of these rMS. Such methods may leverage the Bayesian classification framework already developed for rMS in *BRCA1*, *BRCA2*, *MLH1*, *MSH2*, etc.[37 47 48]

One weakness of this study is that it focuses on early-onset cases using a data set that already contributed either to association of rMS in these genes with breast cancer susceptibility (*ATM*, *CHEK2*),[10 11] or susceptibility to breast cancer in general (*MRE11A*, *NBN*, *RAD50*, *RINT1*, *XRCC2*).[12 14 15] While the impact on the overall results of a possible false association for one or another of the genes is addressed by the leave-one-out analysis, the possibility remains that the ORs that we report are systematically inflated either because this was a study of early-onset cases or because of winner's curse.[41] These issues were partly ameliorated in two ways: (i) an OR≥2.5 grouping of rMS could be isolated by each of the four rMS analysis programs that we used, and (ii) the group of women that can benefit most from early or intensified breast cancer screening is primarily those at risk of early-onset breast cancer—largely, the group of women from which the cases used in this study are drawn. Looking forward, the ratio of the above-threshold rMS to T+SJV can be re-evaluated in case–control studies, but accurate assessment of risk will have to come from prospective cohort studies. A second weakness in our analytic strategy is that the rMS analyses in five of the genes included here—*ATM*, *BARD1*, *MRE11A*, *RAD50* and *NBN*—are somewhat dependent on our definitions of key protein functional domains. This

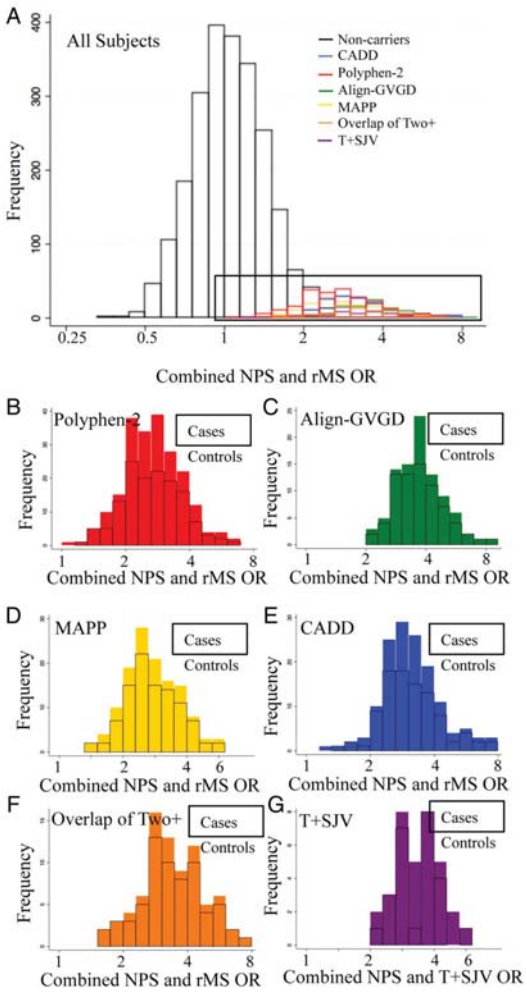


Figure 3 Combined normalised polygene score (NPS) and rare variant OR. (A) Distribution of subjects by NPS when carriers of rare missense substitutions (rMS) in the 'above threshold' group had their NPS increased by the factor of their OR. Stacked histograms of the expansion of the combined contribution of NPS and rMS for carriers with a variant conferring an average OR≥2.5 with (B) PolyPhen-2, (C) Align-GVGD, (D) MAPP, (E) CADD as well as (F) carriers of variants that were placed in the 'above threshold' group for two or more missense analysis programs and (G) truncation and splice junction variant (T+SJV) carriers.

analytic element is in need of independent evaluation and refinement.

Our analysis raises additional questions regarding standard clinical genetic testing practices using panel tests. For the established moderate-risk genes *ATM*, *CHEK2* and *NBN*, the majority of the pathogenic variants that the test can actually detect are rMS, likely to be reported to patients as VUS, and likely to be normalised during counselling. In this circumstance, how does one answer the clinical validity question, "Are the variants the test is intended to identify associated with disease risk, and are these risks well quantified?"[41] What is the impact on studies intended to explore the penetrance and tumour spectrum of pathogenic variants in these genes if the studies focus on T+SJVs even though these may represent a minority of the pathogenic variants? One path forward lies in a more nuanced use of the IARC 5-class system for variant classification and reporting to incorporate more data from ongoing research on missense substitution evaluation.[49] From work that defined the sequence analysis-based prior probabilities of pathogenicity for rMS in *BRCA1*, *BRCA2* and the mismatch repair genes, one can clearly define subsets of rMS that have relatively high probabilities of pathogenicity.[2 42] A straightforward approach for clinicians could be to make systematic efforts to enrol carriers of high probability of pathogenicity rMS in research studies, such as those coordinated through the Evidence-based Network for the Interpretation of Germline Mutant Alleles (ENIGMA) consortium,[50] while still describing these findings to patients as VUS. For *BRCA1*, *BRCA2* and the mismatch repair genes, these could be defined as rMS with prior probabilities of pathogenicity of ≥0.66 as defined at the calibrated prior probability of pathogenicity websites (priors.hci.utah.edu/PRIORS/index.php and hci-lovd.hci.utah.edu/home.php, respectively). rMS from the nine genes examined here that are placed in an OR≥2.5 grouping by two or more of the missense analysis programs similarly fall into a relatively high probability of pathogenicity subset. VUS with lower probabilities of pathogenicity could reasonably be normalised since future reclassification to a clearly pathogenic variant is rather unlikely. Such an approach would better prioritise those missense substitutions with high probabilities of pathogenicity, leading to better understanding of these VUS by clinicians and patients. This approach should empower research towards gene validation, penetrance and tumour spectrum and thereby address the question of clinical validity in the future.

Author affiliations
[1]Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, USA
[2]Department of Dermatology, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, USA
[3]Breast Cancer Genetics Group, Cancer Research Centre of Lyon, Centre Léon Bérard, Lyon, France
[4]Genetic Cancer Susceptibility group, International Agency for Research on Cancer, Lyon, France
[5]Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, USA
[6]Department of Communication and Huntsman Cancer Institute, University of Utah
[7]Genetic Epidemiology Laboratory, The University of Melbourne, Melbourne, Victoria, Australia
[8]Instituto de Ciencias Básicas y Medicina Experimental del Hospital Italiano de Buenos Aires, Buenos Aires, Argentina
[9]Cancer Genomics Laboratory, CHUQ Research Center, Quebec City, Canada
[10]Centre for Epidemiology and Biostatistics, School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia
[11]Department of Epidemiology (Genome Epidemiology Lab), Seoul National University School of Public Health, Seoul, Korea
[12]Department of Molecular Genetics, Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada
[13]Cancer Prevention Institute of California, Fremont, California, USA
[14]Department of Health Research and Policy, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, California, USA
[15]Department of Family and Preventive Medicine, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, USA
[16]Genetic Epidemiology of Cancer Team, Inserm, U900, Institut Curie, Paris, France

Twitter Follow Javier Oliver at @javiom

Acknowledgements The authors wish to thank all participants in the BCFR for their contribution to the study. The authors also appreciate the support of J. McKay and the Genetic Cancer Susceptibility group at the International Agency for Research on Cancer (IARC).

Contributors ELY and BJF were involved in data analysis, critical revision of manuscript and final approval. AWS, FD, GD, NF, TCF, AG, WKK, SMcK-C, TN-D, JO, AMP, MP, NR, JSR, MH, JG, FLC-K, FL, DEG, MV and CV were involved in data acquisition, critical revision of manuscript and final approval. KAK was involved

critical revision of manuscript and final approval. JLH, MCS, ILA and EMJ were involved in the acquisition of subjects, critical revision of manuscript and final approval. SVT was involved in the acquisition of data, critical revision of manuscript, final approval and is the corresponding author.

Funding This work was supported by the United States National Institutes of Health (NIH) National Cancer Institute (NCI) grants R01 CA121245 and R01 CA155767, by the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer programme, by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, and the Ministère de l'enseignement supérieur, de la recherche, de la science, et de la technologie du Québec through Génome Québec. The BCFR is supported by grant UM1 CA164920 from the USA National Cancer Institute.

Disclaimer The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centres in the BCFR, nor does mention of trade names, commercial products or organisations imply endorsement by the US government or the BCFR.

Competing interests None declared.

Patient consent Obtained.

Ethics approval The studies described here were approved by the institutional review board (IRB) of the International Agency for Research on Cancer (IARC), the University of Utah IRB and the local IRBs of the BCFR centres from which we received samples.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement A table of the data used for this analysis including case–control status and variant calling information is available upon contacting the corresponding author. Additionally, protein multiple sequence alignments for the nine genes are also available upon request.

REFERENCES

1 Abkevich V, Zharkikh A, Deffenbaugh AM, Frank D, Chen Y, Shattuck D, Skolnick MH, Gutin A, Tavtigian SV. Analysis of missense variation in human BRCA1 in the context of interspecific sequence variation. *J Med Genet* 2004;41:492–507.
2 Tavtigian SV, Byrnes GB, Goldgar DE, Thomas A. Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. *Hum Mutat* 2008;29:1342–54.
3 Kean S. The 'other' breast cancer genes. *Science* 2014;343:1457–9.
4 Stadler ZK, Schrader KA, Vijai J, Robson ME, Offit K. Cancer genomics and inherited risk. *J Clin Oncol* 2014;32:687–98.
5 Tavtigian SV, Chenevix-Trench G. Growing recognition of the role for rare missense substitutions in breast cancer susceptibility. *Biomark Med* 2014;8:589–603.
6 Byrnes GB, Southey MC, Hopper JL. Are the so-called low penetrance breast cancer genes, ATM, BRIP1, PALB2 and CHEK2, high risk for women with strong family histories?. *Breast Cancer Res* 2008;10:208.
7 Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, Schmidt MK, Chang-Claude J, Bojesen SE, Bolla MK, Wang Q, Dicks E, Lee A, Turnbull C, Rahman N, Fletcher O, Peto J, Gibson L, Dos Santos Silva I, Nevanlinna H, Muranen TA, Aittomäki K, Blomqvist C, Czene K, Irwanto A, Liu J, Waisfisz Q, Meijers-Heijboer H, Adank M, van der Luijt RB, Hein R, Dahmen N, Beckman L, Meindl A, Schmutzler RK, Müller-Myhsok B, Lichtner P, Hopper JL, Southey MC, Makalic E, Schmidt DF, Uitterlinden AG, Hofman A, Hunter DJ, Chanock SJ, Vincent D, Bacot F, Tessier DC, Canisius S, Wessels LFA, Haiman CA, Shah M, Luben R, Brown J, Luccarini C, Schoof N, Humphreys K, Li J, Nordestgaard BG, Nielsen SF, Flyger H, Couch FJ, Wang X, Vachon C, Stevens KN, Lambrechts D, Moisse M, Paridaens R, Christiaens M-R, Rudolph A, Nickels S, Flesch-Janys D, Johnson N, Aitken Z, Aaltonen K, Heikkinen T, Broeks A, Veer LJV, van der Schoot CE, Guénel P, Truong T, Laurent-Puig P, Menegaux F, Marme F, Schneeweiss A, Sohn C, Burwinkel B, Zamora MP, Perez JIA, Pita G, Alonso MR, Cox A, Brock IW, Cross SS, Reed MWR, Sawyer EJ, Tomlinson I, Kerin MJ, Miller N, Henderson BE, Schumacher F, Le Marchand L, Andrulis IL, Knight JA, Glendon G, Mulligan AM, Lindblom A, Margolin S, Hooning MJ, Hollestelle A, van den Ouweland AMW, Jager A, Bui QM, Stone J, Dite GS, Apicella C, Tsimiklis H, Giles GG, Severi G, Baglietto L, Fasching PA, Haeberle L, Ekici AB, Beckmann MW, Brenner H, Müller H, Arndt V, Stegmaier C, Swerdlow A, Ashworth A, Orr N, Jones M, Figueroa J, Lissowska J, Brinton L, Goldberg MS, Labrèche F, Dumont M, Winqvist R, Pylkäs K, Jukkola-Vuorinen A, Grip M, Brauch H, Hamann U, Brüning T, Radice P, Peterlongo P, Manoukian S, Bonanni B, Devilee P, Tollenaar RA, Seynaeve C, van Asperen CJ, Jakubowska A, Lubinski J, Jaworska K, Durda K, Mannermaa A, Kataja V, Kosma V-M, Hartikainen JM, Bogdanova NV, Antonenkova NN, Dörk T, Kristensen VN, Anton-Culver H, Slager S, Toland AE, Edge S, Fostira F, Kang D, Yoo K-Y, Noh D-Y, Matsuo K, Ito H, Iwata H, Sueta A, Wu AH, Tseng C-C, Van Den Berg D, Stram DO, Shu X-O, Lu W, Gao Y-T, Cai H, Teo SH, Yip CH, Phuah SY, Cornes BK, Hartman M, Miao H, Lim WY, Sng J-H, Muir K, Lophatananon A, Stewart-Brown S, Siriwanarangsan P, Shen C-Y, Hsiung C-N, Wu P-E, Ding S-L, Sangrajrang S, Gaborieau V, Brennan P, McKay J, Blot WJ, Signorello LB, Cai Q, Zheng W, Deming-Halverson S, Shrubsole M, Long J, Simard J, Garcia-Closas M, Pharoah PDP, Chenevix-Trench G, Dunning AM, Benitez J, Easton DF. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* 2013;45:353–61, 361e1–2.
8 Maxwell KN, Nathanson KL. Common breast cancer risk variants in the post-COGS era: a comprehensive review. *Breast Cancer Res* 2013;15:212.
9 Mavaddat N, Pharoah PDP, Michailidou K, Tyrer J, Brook MN, Bolla MK, Wang Q, Dennis J, Dunning AM, Shah M, Luben R, Brown J, Bojesen SE, Nordestgaard BG, Nielsen SF, Flyger H, Czene K, Darabi H, Eriksson M, Peto J, Dos-Santos-Silva I, Dudbridge F, Johnson N, Schmidt MK, Broeks A, Verhoef S, Rutgers EJ, Swerdlow A, Ashworth A, Orr N, Schoemaker MJ, Figueroa J, Chanock SJ, Brinton L, Lissowska J, Couch FJ, Olson JE, Vachon C, Pankratz VS, Lambrechts D, Wildiers H, Van Ongeval C, van Limbergen E, Kristensen V, Grenaker Alnaes G, Nord S, Borresen-Dale A-L, Nevanlinna H, Muranen TA, Aittomaki K, Blomqvist C, Chang-Claude J, Rudolph A, Seibold P, Flesch-Janys D, Fasching PA, Haeberle L, Ekici AB, Beckmann MW, Burwinkel B, Marme F, Schneeweiss A, Sohn C, Trentham-Dietz A, Newcomb P, Titus L, Egan KM, Hunter DJ, Lindstrom S, Tamimi RM, Kraft P, Rahman N, Turnbull C, Renwick A, Seal S, Li J, Liu J, Humphreys K, Benitez J, Pilar Zamora M, Arias Perez JI, Menendez P, Jakubowska A, Lubinski J, Jaworska-Bieniek K, Durda K, Bogdanova NV, Antonenkova NN, Dork T, Anton-Culver H, Neuhausen SL, Ziogas A, Bernstein L, Devilee P, Tollenaar RAEM, Seynaeve C, van Asperen CJ, Cox A, Cross SS, Reed MWR, Khusnutdinova E, Bermisheva M, Prokofyeva D, Takhirova Z, Meindl A, Schmutzler RK, Sutter C, Yang R, Schurmann P, Bremer M, Christiansen H, Park-Simon T-W, Hillemanns P, Guenel P, Truong T, Menegaux F, Sanchez M, Radice P, Peterlongo P, Manoukian S, Pensotti V, Hopper JL, Tsimiklis H, Apicella C, Southey MC, Brauch H, Bruning T, Ko Y-D, Sigurdson AJ, Doody MM, Hamann U, Torres D, Ulmer H-U, Forsti A, Sawyer EJ, Tomlinson I, Kerin MJ, Miller N, Andrulis IL, Knight JA, Glendon G, Marie Mulligan A, Chenevix-Trench G, Balleine R, Giles GG, Milne RL, McLean C, Lindblom A, Margolin S, Haiman CA, Henderson BE, Schumacher F, Le Marchand L, Eilber U, Wang-Gohrke S, Hooning MJ, Hollestelle A, van den Ouweland AMW, Koppert LB, Carpenter J, Clarke C, Scott R, Mannermaa A, Kataja V, Kosma V-M, Hartikainen JM, Brenner H, Arndt V, Stegmaier C, Karina Dieffenbach A, Winqvist R, Pylkas K, Jukkola-Vuorinen A, Grip M, Offit K, Vijai J, Robson M, Rau-Murthy R, Dwek M, Swann R, Annie Perkins K, Goldberg MS, Labreche F, Dumont M, Eccles DM, Tapper WJ, Rafiq S, John EM, Whittemore AS, Slager S, Yannoukakos D, Toland AE, Yao S, Zheng W, Halverson SL, Gonzalez-Neira A, Pita G, Rosario Alonso M, Alvarez N, Herrero D, Tessier DC, Vincent D, Bacot F, Luccarini C, Baynes C, Ahmed S, Maranian M, Healey CS, Simard J, Hall P, Easton DF, Garcia-Closas M. Prediction of Breast Cancer Risk Based on Profiling With Common Genetic Variants. *J Natl Cancer Inst* 2015;107:djv036–djv036.
10 Tavtigian SV, Oefner PJ, Babikyan D, Hartmann A, Healey S, Le Calvez-Kelm F, Lesueur F, Byrnes GB, Chuang S-C, Forey N, Feuchtinger C, Gioia L, Hall J, Hashibe M, Herte B, McKay-Chopin S, Thomas A, Vallée MP, Voegele C, Webb PM, Whiteman DC, Sangrajrang S, Hopper JL, Southey MC, Andrulis IL, John EM, Chenevix-Trench G. Rare, evolutionarily unlikely missense substitutions in ATM confer increased risk of breast cancer. *Am J Hum Genet* 2009;85:427–46.
11 Le Calvez-Kelm F, Lesueur F, Damiola F, Vallée M, Voegele C, Babikyan D, Durand G, Forey N, McKay-Chopin S, Robinot N, Nguyen-Dumont T, Thomas A, Byrnes GB, Hopper JL, Southey MC, Andrulis IL, John EM, Tavtigian SV. Rare, evolutionarily unlikely missense substitutions in CHEK2 contribute to breast cancer susceptibility: results from a breast cancer family registry case-control mutation-screening study. *Breast Cancer Res* 2011;13:R6.
12 Damiola F, Pertesi M, Oliver J, Le Calvez-Kelm F, Voegele C, Young EL, Robinot N, Forey N, Durand G, Vallée MP, Tao K, Roane TC, Williams GJ, Hopper JL, Southey MC, Andrulis IL, John EM, Goldgar DE, Lesueur F, Tavtigian SV. Rare key functional domain missense substitutions in MRE11A, RAD50, and NBN contribute to breast cancer susceptibility: results from a Breast Cancer Family Registry case-control mutation-screening study. *Breast Cancer Res* 2014;16:R58.
13 Le Calvez-Kelm F, Oliver J, Damiola F, Forey N, Robinot N, Durand G, Voegele C, Vallée MP, Byrnes G, Registry BCF, Hopper JL, Southey MC, Andrulis IL, John EM, Tavtigian SV, Lesueur F. RAD51 and Breast Cancer Susceptibility: No Evidence for Rare Variant Association in the Breast Cancer Family Registry Study. *PLoS One* 2012;7:e52374.
14 Park DJ, Tao K, Le Calvez-Kelm F, Nguyen-Dumont T, Robinot N, Hammet F, Odefrey F, Tsimiklis H, Teo ZL, Thingholm LB, Young EL, Voegele C, Lonie A, Pope BJ, Roane TC, Bell R, Hu H, Shankaracharya, Huff CD, Ellis J, Li J, Makunin IV, John EM, Andrulis IL, Terry MB, Daly M, Buys SS, Snyder C, Lynch HT,

Devilee P, Giles GG, Hopper JL, Feng B-J, Lesueur F, Tavtigian S V, Southey MC, Goldgar DE. Rare Mutations in RINT1 Predispose Carriers to Breast and Lynch Syndrome-Spectrum Cancers. *Cancer Discov* 2014;4:804–15.
15 Park DJ, Lesueur F, Nguyen-Dumont T, Pertesi M, Odefrey F, Hammet F, Neuhausen SL, John EM, Andrulis IL, Terry MB, Daly M, Buys S, Le Calvez-Kelm F, Lonie A, Pope BJ, Tsimiklis H, Voegele C, Hilbers FM, Hoogerbrugge N, Barroso A, Osorio A, Giles GG, Devilee P, Benitez J, Hopper JL, Tavtigian SV, Goldgar DE, Southey MC. Rare mutations in XRCC2 increase the risk of breast cancer. *Am J Hum Genet* 2012;90:734–9.
16 Tavtigian SV, Deffenbaugh AM, Yin L, Judkins T, Scholl T, Samollow PB, de Silva D, Zharkikh A, Thomas A. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J Med Genet* 2006;43:295–305.
17 Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;46:310–5.
18 Stone EA, Sidow A. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Res* 2005;15:978–86.
19 Adzhubei I, Jordan DM, Sunyaev SR. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Curr Protoc Hum Genet* 2013;Chapter 7: Unit7.20.
20 John EM, Hopper JL, Beck JC, Knight JA, Neuhausen SL, Senie RT, Ziogas A, Andrulis IL, Anton-Culver H, Boyd N, Buys SS, Daly MB, O'Malley FP, Santella RM, Southey MC, Venne VL, Venter DJ, West DW, Whittemore AS, Seminara D. The Breast Cancer Family Registry: an infrastructure for cooperative multinational, interdisciplinary and translational studies of the genetic epidemiology of breast cancer. *Breast Cancer Res* 2004;6:R375–89.
21 Greenblatt MS, Beaudet JG, Gump JR, Godin KS, Trombley L, Koh J, Bond JP. Detailed computational study of p53 and p16: using evolutionary sequence analysis and disease-associated mutations to predict the functional consequences of allelic variants. *Oncogene* 2003;22:1150–63.
22 Armougom F, Moretti S, Poirot O, Audic S, Dumas P, Schaeli B, Keduas V, Notredame C. Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res* 2006;34:W604–8.
23 Taly JF, Magis C, Bussotti G, Chang J-M, Di Tommaso P, Erb I, Espinosa-Carrasco J, Kemena C, Notredame C. Using the T-Coffee package to build multiple sequence alignments of protein, RNA, DNA sequences and 3D structures. *Nat Protoc* 2011;6:1669–82.
24 Felsenstein J. PHYLIP (Phylogeny Inference Package) version 3.6. *Distrib by author Dep Genome Sci Univ Washington, Seattle* 2005.
25 Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248–9.
26 Meyer KB, O'Reilly M, Michailidou K, Carlebur S, Edwards SL, French JD, Prathalingham R, Dennis J, Bolla MK, Wang Q, de Santiago I, Hopper JL, Tsimiklis H, Apicella C, Southey MC, Schmidt MK, Broeks A, Van 't Veer LJ, Hogervorst FB, Muir K, Lophatananon A, Stewart-Brown S, Siriwanarangsan P, Fasching PA, Lux MP, Ekici AB, Beckmann MW, Peto J, Dos Santos Silva I, Fletcher O, Johnson N, Sawyer EJ, Tomlinson I, Kerin MJ, Miller N, Marme F, Schneeweiss A, Sohn C, Burwinkel B, Guénel P, Truong T, Laurent-Puig P, Menegaux F, Bojesen SE, Nordestgaard BG, Nielsen SF, Flyger H, Milne RL, Zamora MP, Arias JI, Benitez J, Neuhausen S, Anton-Culver H, Ziogas A, Dur CC, Brenner H, Müller H, Arndt V, Stegmaier C, Meindl A, Schmutzler RK, Engel C, Ditsch N, Brauch H, Brüning T, Ko Y-D, Nevanlinna H, Muranen TA, Aittomäki K, Blomqvist C, Matsuo K, Ito H, Iwata H, Yatabe Y, Dörk T, Helbig S, Bogdanova NV, Lindblom A, Margolin S, Mannermaa A, Kataja V, Kosma V-M, Hartikainen JM, Chenevix-Trench G, Wu AH, Tseng C-C, Van Den Berg D, Stram DO, Lambrechts D, Thienpont B, Christiaens M-R, Smeets A, Chang-Claude J, Rudolph A, Seibold P, Flesch-Janys D, Radice P, Peterlongo P, Bonanni B, Bernard L, Couch FJ, Olson JE, Wang X, Purrington K, Giles GG, Severi G, Baglietto L, McLean C, Haiman CA, Henderson BE, Schumacher F, Le Marchand L, Simard J, Goldberg MS, Labrèche F, Dumont M, Teo S-H, Yip C-H, Phuah S-Y, Kristensen V, Grenaker Alnæs G, Børresen-Dale A-L, Zheng W, Deming-Halverson S, Shrubsole M, Long J, Winqvist R, Pylkäs K, Jukkola-Vuorinen A, Kauppila S, Andrulis IL, Knight JA, Glendon G, Tchatchou S, Devilee P, Tollenaar RAEM, Seynaeve CM, García-Closas M, Figueroa J, Chanock SJ, Lissowska J, Czene K, Darabi H, Eriksson K, Hooning MJ, Martens JWM, van den Ouweland AMW, van Deurzen CHM, Hall P, Li J, Liu J, Humphreys K, Shu X-O, Lu W, Gao Y-T, Cai H, Cox A, Reed MWR, Blot W, Signorello LB, Cai Q, Pharoah PDP, Ghoussaini M, Harrington P, Tyrer J, Kang D, Choi J-Y, Park SK, Noh D-Y, Hartman M, Hui M, Lim W-Y, Buhari SA, Hamann U, Försti A, Rüdiger T, Ulmer H-U, Jakubowska A, Lubinski J, Jaworska K, Durda K, Sangrajrang S, Gaborieau V, Brennan P, McKay J, Vachon C, Slager S, Fostira F, Pilarski R, Shen C-Y, Hsiung C-N, Wu P-E, Hou M-F, Swerdlow A, Ashworth A, Orr N, Schoemaker MJ, Ponder BAJ, Dunning AM, Easton DF. Fine-scale mapping of the FGFR2 breast cancer risk locus: putative functional variants differentially bind FOXA1 and E2F1. *Am J Hum Genet* 2013;93:1046–60.
27 Song M, Kraft P, Joshi AD, Barrdahl M, Chatterjee N. Testing calibration of risk models at extremes of disease risk. *Biostatistics* 2015;16:143–54.
28 Huo D, Zheng Y, Ogundiran TO, Adebamowo C, Nathanson KL, Domchek SM, Rebbeck TR, Simon MS, John EM, Hennis A, Nemesure B, Wu S-Y, Leske MC, Ambs S, Niu Q, Zhang J, Cox NJ, Olopade OI. Evaluation of 19 susceptibility loci of breast cancer in women of African ancestry. *Carcinogenesis* 2012;33:835–40.
29 Fejerman L, Stern MC, Ziv E, John EM, Torres-Mejia G, Hines LM, Wolff R, Wang W, Baumgartner KB, Giuliano AR, Slattery ML. Genetic ancestry modifies the association between genetic risk variants and breast cancer risk among Hispanic and non-Hispanic white women. *Carcinogenesis* 2013;34:1787–93.
30 Zheng W, Zhang B, Cai Q, Sung H, Michailidou K, Shi J, Choi J-Y, Long J, Dennis J, Humphreys MK, Wang Q, Lu W, Gao Y-T, Li C, Cai H, Park SK, Yoo K-Y, Noh D-Y, Han W, Dunning AM, Benitez J, Vincent D, Bacot F, Tessier D, Kim S-W, Lee MH, Lee JW, Lee J-Y, Xiang Y-B, Zheng Y, Wang W, Ji B-T, Matsuo K, Ito H, Iwata H, Tanaka H, Wu AH, Tseng C, Van Den Berg D, Stram DO, Teo SH, Yip CH, Kang IN, Wong TY, Shen C-Y, Yu J-C, Huang C-S, Hou M-F, Hartman M, Miao H, Lee SC, Putti TC, Muir K, Lophatananon A, Stewart-Brown S, Siriwanarangsan P, Sangrajrang S, Shen H, Chen K, Wu P-E, Ren Z, Haiman CA, Sueta A, Kim MK, Khoo US, Iwasaki M, Pharoah PDP, Wen W, Hall P, Shu X-O, Easton DF, Kang D. Common genetic determinants of breast-cancer risk in East Asian women: a collaborative study of 23 637 breast cancer cases and 25 579 controls. *Hum Mol Genet* 2013;22:2539–50.
31 Lee CH, Dershaw DD, Kopans D, Evans P, Monsees B, Monticciolo D, Brenner RJ, Bassett L, Berg W, Feig S, Hendrick E, Mendelson E, D'Orsi C, Sickles E, Burhenne LW. Breast cancer screening with imaging: recommendations from the Society of Breast Imaging and the ACR on the use of mammography, breast MRI, breast ultrasound, and other technologies for the detection of clinically occult breast cancer. *J Am Coll Radiol* 2010;7:18–27.
32 Daly M, Pilarski R, Axilbund J, Buys S, Crawford B, Friedman S, Garber J, Horton C, Kaklamani V, Klein C, Kohlmann W, Kurian A, Litton J, Madlensky L, Marcom P, Merajver S, Offit K, Pal T, Pasche B, Reiser G, Shannon K, Swisher E, Voian N, Weitzel J, Whelan A, Wiesner G, Dwyer M, Kumar R, National comprehensive cancer Network. Genetic/familial high-risk assessment: breast and ovarian, version 1.2014. *J Natl Compr Canc Netw* 2014;12:1326–38.
33 Saslow D, Boetes C, Burke W, Harms S, Leach MO, Lehman CD, Morris E, Pisano E, Schnall M, Sener S, Smith RA, Warner E, Yaffe M, Andrews KS, Russell CA. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA Cancer J Clin* 2007;57:75–89.
34 SEER Cancer Statistics Factsheets: Breast Cancer. *Natl Cancer Institute Bethesda, MD*. 2014. http//seer.cancer.gov/statfacts/html/breast.html
35 Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, Mulvihill JJ. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst* 1989;81:1879–86.
36 Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat Med* 2004;23:1111–30.
37 Vallée MP, Francy TC, Judkins MK, Babikyan D, Lesueur F, Gammon A, Goldgar DE, Couch FJ, Tavtigian SV. Classification of missense substitutions in the BRCA genes: a database dedicated to Ex-UVs. *Hum Mutat* 2012;33:22–8.
38 Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 1994;266:66–71.
39 Castilla LH, Couch FJ, Erdos MR, Hoskins KF, Calzone K, Garber JE, Boyd J, Lubin MB, Deshano ML, Brody LC. Mutations in the BRCA1 gene in families with early-onset breast and ovarian cancer. *Nat Genet* 1994;8:387–91.
40 Futreal PA, Liu Q, Shattuck-Eidens D, Cochran C, Harshman K, Tavtigian S, Bennett LM, Haugen-Strano A, Swensen J, Miki Y. BRCA1 mutations in primary breast and ovarian carcinomas. *Science* 1994;266:120–2.
41 Easton DF, Pharoah PDP, Antoniou AC, Tischkowitz M, Tavtigian SV, Nathanson KL, Devilee P, Meindl A, Couch FJ, Southey M, Goldgar DE, Evans DGR, Trench GC, Rahman N, Robson M, Domchek SM, Foulkes WD. Gene-panel sequencing and the prediction of breast-cancer risk. *N Engl J Med* 2015;372:2243–57.
42 Thompson BA, Greenblatt MS, Vallee MP, Herkert JC, Tessereau C, Young EL, Adzhubey IA, Li B, Bell R, Feng B, Mooney SD, Radivojac P, Sunyaev SR, Frebourg T, Hofstra RMW, Sijmons RH, Boucher K, Thomas A, Goldgar DE, Spurdle AB, Tavtigian SV. Calibration of multiple in silico tools for predicting pathogenicity of mismatch repair gene missense substitutions. *Hum Mutat* 2013;34:255–65.
43 Desmond A, Kurian AW, Gabree M, Mills MA, Anderson MJ, Kobayashi Y, Horick N, Yang S, Shannon KM, Tung N, Ford JM, Lincoln SE, Ellisen LW. Clinical Actionability of Multigene Panel Testing for Hereditary Breast and Ovarian Cancer Risk Assessment. *JAMA Oncol* 2015;02114:1–9.
44 Maxwell KN, Wubbenhorst B, D'Andrea K, Garman B, Long JM, Powers J, Rathbun K, Stopfer JE, Zhu J, Bradbury AR, Simon MS, DeMichele A, Domchek SM, Nathanson KL. Prevalence of mutations in a panel of breast cancer susceptibility genes in BRCA1/2-negative patients with early-onset breast cancer. *Genet Med* 2015;17:630–8.
45 Tung N, Battelli C, Allen B, Kaldate R, Bhatnagar S, Bowles K, Timms K, Garber JE, Herold C, Ellisen L, Krejdovsky J, DeLeonardis K, Sedgwick K, Soltis K, Roa B,

Wenstrup RJ, Hartman A-R. Frequency of mutations in individuals with breast cancer referred for BRCA1 and BRCA2 testing using next-generation sequencing with a 25-gene panel. *Cancer* 2015;121:25–33.
46 Roeb W, Higgins J, King MC. Response to DNA Damage of CHEK2 Missense Mutations in Familial Breast Cancer. *Hum Mol Genet* 2012;21:2738–44.
47 Thompson BA, Goldgar DE, Paterson C, Clendenning M, Walters R, Arnold S, Parsons MT, Walsh MD, Gallinger S, Haile RW, Hopper JL, Jenkins MA, Lemarchand L, Lindor NM, Newcomb PA, Thibodeau SN, Young JP, Buchanan DD, Tavtigian SV, Spurdle AB, Michael DW, Michael DW, Gallinger S, Haile RW, Hopper JL, Jenkins MA, Lemarchand L, Lindor NM, Newcomb PA, Thibodeau SN, Young JP, Buchanan DD, Tavtigian SV, Spurdle AB. A multifactorial likelihood model for MMR gene variant classification incorporating probabilities based on sequence bioinformatics and tumor characteristics: a report from the Colon Cancer Family Registry. *Hum Mutat* 2013;34:200–9.
48 Lindor NM, Guidugli L, Wang X, Vallée MP, Monteiro ANA, Tavtigian S, Goldgar DE, Couch FJ. A review of a multifactorial probability-based model for classification of BRCA1 and BRCA2 variants of uncertain significance (VUS). *Hum Mutat* 2012;33:8–21.
49 Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, Hogervorst FBL, Hoogerbrugge N, Spurdle AB, Tavtigian SV. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum Mutat* 2008;29:1282–91.
50 Spurdle AB, Healey S, Devereau A, Hogervorst FBL, Monteiro ANA, Nathanson KL, Radice P, Stoppa-Lyonnet D, Tavtigian S, Wappenschmidt B, Couch FJ, Goldgar DE. ENIGMA-evidence-based network for the interpretation of germline mutant alleles: An international initiative to evaluate risk and clinical significance associated with sequence variation in BRCA1 and BRCA2 genes. *Hum Mutat* 2012;33:2–7.

CHAPTER 3

PANCREATIC CANCER AS AN INCLUSION FACTOR FOR GENETIC TESTING OF BREAST AND COLORECTAL CANCER SUSCEPTIBILITY GENES

Abstract

Genes associated with hereditary breast and ovarian cancer (HBOC) and colorectal cancer (CRC) susceptibility have been shown to play a role in pancreatic cancer susceptibility. Germline genetic testing of pancreatic cancer cases could be beneficial for at-risk relatives with pathogenic variants in established HBOC and CRC genes, but it is unclear how many pancreatic cancer cases harbor pathogenic variants in those genes. 55 breast, 25 ovarian, and 66 pancreatic cancer cases diagnosed at the Huntsman Cancer Hospital, SLC, UT, unselected for family history were sequenced on a custom 34 gene panel including known HBOC and CRC genes. Variants of uncertain significance were down-weighted based on prior probabilities from public databases if available or predictions of deleteriousness from *in silico* predictors. A second set of 95 unselected pancreatic cancer cases, the pancreatic cancer cases of the Cancer Genome Atlas, and an unselected pancreatic cancer screen from the Mayo clinic were combined in a meta-analysis to estimate the proportion of carriers with pathogenic variants. In the pancreatic cancer cases, 7.6-8.5% carried a variant that would alter the screening

recommendations for at-risk relatives, including VUS from known HBOC and CRC susceptibility genes. A meta-analysis reveals that approximately 11.9% of unselected pancreatic cancer cases carry a pathogenic variant in HBOC or CRC susceptibility genes. With the inclusion of both HBOC and CRC susceptibility genes in a panel test, pancreatic cancer cases unselected for age or family history have a high enough percentage of carriers to rationalize genetic testing for identification of variants that could be further used in cascade testing of healthy relatives for increased HBOC and CRC surveillance measures.

Introduction

Over the last few years, massively parallel sequencing converged with targeted capture using array synthesized baits to enable panel testing of most known cancer susceptibility genes.[1–4] These panel tests then replaced Sanger sequencing of limited sets of syndromic genes, revolutionizing the genetic testing landscape. Before panel testing, the personal and family history of a cancer patient had to be examined carefully to select the most likely syndromic genes. Due to atypical personal or family history combined with the nondiscreet nature of cancer susceptibility, there were probably many instances where a patient never received the correct test.[5–10]

The transition to multigene panel testing brings with it a greater burden of Variants of Uncertain Significance (VUS) than syndromic gene testing.[5,7,11–14] Most panel testing based studies of the proportion of patients with an identifiable genetic predisposition focus on clearly pathogenic variants; this analytic strategy results in clarity but underestimates the proportion of cases with genetic predisposition. Since 2008, we have invested in calibrating computational analyses of missense substitutions and splice

junction variants.[15–21] These calibration studies simultaneously validate the utility of the computational tools examined, and convert the native output of those tools into a Bayesian prior probability of pathogenicity (Prior_P).[20,21] In the context of VUS classification efforts, this Prior_P is a starting point for clinical classification of VUS. Moreover, the Prior_Ps enable probabilistic estimation of the proportion of patients with a pathogenic variant detected through panel testing.

Today, some cancer diagnoses by themselves imply a high enough probability to carry a pathogenic variant across a panel of susceptibility genes that panel-based predisposition testing is considered automatically appropriate, such as Ovarian Cancer (OC) and triple negative Breast Cancer (BC) diagnosed at less than 60 years of age.[22–24] Universal testing of Colorectal Cancer (CRC) tumors for mismatch repair deficiency, which is, in effect, a prescreen for Lynch Syndrome (LS), provides a third example.[25,26] Other cancers still require further analysis; for example, the personal and family history of patients with luminal BC remains an important determinant for whether they would be offered genetic testing.

Genetic testing of patients with pancreatic cancer (PC) is not yet routine, in part because methods for early detection of PC have limited utility.[27–30] One substantial benefit of genetic testing is application of preventive measures for carriers of disease susceptibility alleles, notably the at-risk relatives of probands. PC patients have been reported to carry pathogenic variants in a variety of cancer susceptibility genes, notably Hereditary Breast and Ovarian Cancer (HBOC) and CRC genes.[22,31–39] Indeed, systematically testing PC cases for pathogenic germline variants in HBOC and CRC genes in order to identify at-risk relatives who would benefit from preventive measures for BC, OC, and CRC would be a medically useful application of panel testing. Testing

all PC cases, however, is only beneficial if the proportion of carriers among PC cases is high enough to justify that testing. To estimate the percentage of PC cases that carry variants with potential medical management impact for at-risk relatives, we applied panel testing to independent discovery (*n = 66*) and replication (*n = 96*) sets of PC patients, both unselected for family cancer history. To demonstrate the overlap with HBOC and CRC predisposition genes, we tested unselected BC and OC cases, and familial CRC cases, in parallel with the discovery set of PC cases. To demonstrate generalizability of the results in PC cases, we performed a meta-analysis including published panel tests of unselected PC cases.

Methods

Subjects and ethics statement

This study was approved by the Institutional Review Board of the University of Utah. All participants gave written consent which included DNA sampling for molecular studies and access to medical records.

Breast (*n = 55*), ovarian (*n = 25*) and pancreatic cancer (*n = 66*) cases were selected on the minimal requirements of personal history of cancer and having at least two grandparents in the Utah Population Database (UPDB). The individual family members are then linked to statewide cancer, demographic, and medical information.[40] BC cases were also selected to equally represent the molecular subtypes including luminal (n=20), HER2+ (n=15), and triple negative (n=13) subtypes. Ages at diagnosis and family cancer history were obtained after sequencing and variant evaluation. A second set of pancreatic cancer cases (n=96) were selected on the basis of being newly diagnosed pancreatic cancer cases at the Huntsman Cancer Hospital (HCH), Salt Lake

City, UT during the 2014 and 2015 fiscal years.

UPDB resources were used to identify 51 pedigrees spanning 3-5 generations with a statistical excess of CRCs segregating in an autosomal dominant fashion. A pair of relatives diagnosed with CRC, both with a first degree relative with CRC (*n = 98*) was selected from each pedigree. The members of these "cousin pairs" were separated by 3 to 7 meoises.[41]

Next-generation sequencing library preparation and custom targeted capture

Blood-derived genomic Deoxyribonucleic acid (DNA) (100ng - 1µg) was sheared using a Covaris S2 instrument (Covaris, Woburn, MA, United States). For BC, OC, and PC cases, libraries were prepared using the Ovation Ultralow Library System (NUGEN # 0329). For DNA from OC cases extracted from grossly uninvolved tissue from FFPE blocks (*n = 11*), 2 µl of Uracil-DNA Glycosylase (UDG) (NEB #M0280S) were added to samples following adapter ligation and incubated for 37°C for 15 min, prior to ligation clean-up in order to reduce C>T artifacts.[42] For the CRC cases, libraries were prepared using the NEBNext DNA Library Prep Reagent Set for Illumina (NEB # E6000) according to the manufacturer's instructions.

Library enrichment for a 34 or 59 gene custom panel was done with the Roche SeqCap EZ Choice Library (cat# 06266339001) and the SeqCap EZ Reagent Kit Plus v2 (NimbleGen #06-953-247-001) using the manufacturer's protocol. Individual libraries were combined into pools of 6-12 prior to hybridization. Captured libraries were sequenced on a Illumina HiSeq2000 channel using the HiSeq 101 Cycle Paired-End sequencing protocol. A complete list of genes captured is included in Supplemental Table

S3.1.

Sequences from the Utah cohort with ≥ 100X coverage as determined by CoverageBED,[43] as well as 154 PC cases from the Cancer Genome Atlas (TCGA) were analyzed using the USeq (useq.sourceforge.net) in-house pipeline, according to the Genome Analysis Toolkit (GATK v.3.3-0) best practices recommendations.[44] Variants with a mapping quality score less than 20 were excluded. ANNOVAR was used for variant functional annotation followed by conversion to Human Genome Variation Society (HGVS) nomenclature using Mutalyzer.[45,46]

Sequence variant evaluation

Truncating variants not present in the final exon of a gene were considered pathogenic. The following filters were used to exclude variants from further analysis: minor allele frequency ≥ 0.1% in one or more populations from the Exome Aggregation Consortium (ExAC) database (exac.broadinstitute.org); synonymous/intronic variants with no predicted effect on splicing via MaxEntScan;[47] variants reported as probable-nonpathogenic/nonpathogenic by more than one source with no conflicting reports in ClinVar (www.ncbi.nlm.nih.gov/clinvar).

Variants of uncertain significance (VUS) were included if their estimated probabilities of pathogenicity were >0.8 based on *in silico* predictions from publically available databases for the MMR genes (hci-lovd.hci.utah.edu), or *BRCA1*/2 (http://priors.hci.utah.edu/PRIORS/). VUS of this type were weighted according to their Prior_P score. The remaining rare VUS were included if two of the following criterion were met using *in silico* predictions: C35+ from Align-GVGD, MAPP score ≥11, Polyphen-2 HUMVAR score ≥0.9, and a CADD PHRED score ≥23.[48–52] Weights were

loosely based off of the likelihood ratios identified for *BRCA1*/2,[21] with a weight of 0.66 being used for a minimal overlap of two programs, and 0.81 for an overlap of three or more. Canonical splice acceptor/donor variants predicted to impact splicing were given the weight of 0.97 if the effect of the variant had not been demonstrated experimentally. Known and likely pathogenic sequence variants were confirmed via Sanger sequencing.[53] VUS in published studies used for meta-analysis were graded with the same weights and severity, although were not confirmed via Sanger sequencing. An overview of the datasets and methods used for evaluation are shown in Figure 3.1.

Statistical analysis

STATA V.12.1 (StataCorp, College Station, Texas, USA) was used to conduct meta-analyses and calculate BC/OV/PC cohort carrier percentage 95% confidence intervals. The meta-analyses to compare the carrier frequencies between the different PC cohorts were conducted using metaprop under a random effects model, using Freeman-Tukey transformation to stabilize the variances over the studies.[54] The R package ggplot2 was used to plot the meta-analyses. Other PC case screens were identified on the basis of being recent (2014+), and including BRCA1/2 variants. VUS were weighted as described when possible. The observed incidence of variants in the unselected PC cases, and other published studies (see Figure 3.1) were compared with the expected incidence of variants in non-TCGA ExAC controls (excluding the Finnish and undescribed populations) as a Standardized Incidence Ratio (SIR).[55] Significance tests and confidence intervals were estimated based on a Poisson distribution.[56] For the meta-analysis and SIR calculation, the genes were split into subgroups of high- and moderate-risk cancer susceptibility genes. High- and moderate-risk were definted as genes with a cumulative risk at age

80>32% or between 19-32%, respectively.[2]

Results

Targeted multigene panel of 34 genes

Multigene panel testing with our custom 34-gene panel was conducted in 237 individuals (Table 3.1). We identified 15 truncating variants in 18 individuals (Table 3.2). The remaining 183 missense substitutions underwent further *in silico* analysis. Five VUS in *BRCA1*/2 and the MMR genes were considered to have an elevated probability of pathogenicity (prior probability ≥0.8). A large portion of the risk attributable to moderate-risk genes is due to missense substitutions,[51] which are generally classified as VUS. To estimate the number of carriers of pathogenic missense variants, subjects with relatively high probability of pathogenicity VUS were weighted according to the probability of pathogenicity of the VUS(s) that they carried (Supplemental Table S3.2).

Breast and ovarian cancer cases, unselected for family history

BC cases were ascertained with three different subtypes: luminal, HER2+, and triple negative. The Genetic Counselors at HCH were very active towards BC cases, which would have removed most cases with a personal or family history qualifying them for clinical testing from our pool of candidates. Nonetheless, two luminal BC cases carried pathogenic *BRCA2* variants. No pathogenic or potentially pathogenic variants were found in HER2+ BC. One triple negative BC case carried a variant of interest. After weighting carriers for the VUS, 10% of luminal and 7.7% of triple negative BC cases carried a potentially pathogenic variant (Table 3.2).

OC has been associated with BC genes, as well as LS genes.[57] As such, it was not surprising to identify pathogenic variants in the BC genes *BRCA1*, *BRCA2*, and *RAD50*, and two VUS in the LS genes *MSH2* and *MSH6* in the OC cases. Combining the proportion of carriers of pathogenic and VUS suggest 23.6% of OC cases carry a potentially medically actionable variant (Table 3.2), a much higher carrier percentage than our BC cases.

Colorectal cancer cases, selected for moderate family history

Five pathogenic variants were identified in seven CRC cases from five families (Table 3.2). These included two *MSH2* truncating variants: one found in both members of the relative pair, and the other segregating in just one branch of the pedigree. A large deletion encompassing part of *EPCAM* and *MSH2* exons 1-8 was shared by another cousin pair which met Amsterdam II criteria for LS testing, but this was not confirmed by other sequencing means. A *BRCA2* frameshift variant, c.6275_6276del, was identified in one individual with multiple instances of early onset BC, including a daughter diagnosed at 36 years of age. There were also two VUS of interest (Table 3.2). The VUS *BRCA2* p.(W31C) was only carried by one case of the target pair. The carrier was diagnosed with OC and CRC, while the noncarrier was diagnosed with BC, endometrial cancer, and CRC. There were no other instances of BC or OC recorded in the family.

It is important to note that three out of the seven of variants identified in the CRC cases were in HBOC genes. Although the large proportion of HBOC variants identified is partially attributed to prescreening many of these families for MMR variants, this is an example where multigene panel testing offers a greater benefit beyond sequencing limited to CRC risk genes.

Pancreatic cancer cases, unselected for family history

In the initial 66 pancreatic cancer cases, 4 pathogenic variants were identified in *BRCA2*, *MSH6*, *PALB2*, and *STK11*; and 2 VUS in *ATM*. After weighting, 8.5% of PC cases carry a variant with potential medical management impact for relatives (Table 3.2). To replicate this observation, 95 independent PC cases underwent a multigene panel test, updated since the initial panel test to include genes associated with hereditary predisposition to cancer. In the second set of PC cases, three pathogenic variants and six VUS were identified (Table 3.3). In addition to *in silico* predictions, *CHEK2* p.(T476M) was found to be damaging in a functional assay for *CHEK2* variants [58], and was thus weighted more strongly towards being pathogenic. After weighting carriers, we found 7.8% of the replication series of PC cases, unselected for family history carried a variant with potential medical impact.

APC p.R99W

Three unrelated PC and one CRC case carried *APC* c.295C>T p.(R99W). Although earlier studies describe this variant as possibly pathogenic,[59] missense substitutions in APC are rarely associated with disease, and a more recent study concluded that this variant is benign.[60,61] Colonoscopy at the age of 64 on two of these individuals did not find evidence of adenomatous polyposis, suggesting further that this variant is benign.

Carrier frequencies across studies

Our two sets of pancreatic cancer cases were combined with a published study of unselected PC cases from the Mayo Clinic ($n = 96$),[62] plus the PC cases from TCGA

(n=154), in a meta-analysis (Figure 3.2). Further, the gene burdens observed in the Utah, Mayo, TCGA, and other recently published PC genetic screening studies were compared to the non-TCGA ExAC (n=49,451) as a population sample (Table 3.4; Supplemental Table S3.3).[4,31,33,34,55,63–65] Among unselected PC cases, 2.9% (1.4-4.8%) carried a pathogenic variant in a high-risk HBOC susceptibility gene (*BRCA1/2* or *PALB2*; SIR 7.7 [6.5-9.0], p<0.0001). PC is part of the LS spectrum,[66] thus it was not surprising that 1.4% (0.3-3.0%) of unselected PC cases were carriers (SIR 2.4 [1.5-3.6], $p = 6.8x10^{-04}$). For the moderate-risk homologous recombination repair (HRR) BC genes *ATM*, *NBN*, and *CHEK2*, 5.0% (2.1-8.8%) of unselected PC cases were estimated to carry a pathogenic variant (SIR 2.5 [1.8-3.2], p<0.0001). Combining these genes with others included in this study, such as MRE11A, RAD50, etc., 11.9% (8.2-16.2%) of unselected PC cases carry a variant that could impact the medical management of carriers' relatives (SIR 3.30 [2.95-3.68], p<0.0001). Although, not included in the meta-analysis or further analyses there was a splice junction variant in the PC susceptibility gene *MEN1*, in TCGA.

Discussion

Through systematic panel testing of PC cases unselected for family history, we estimate that 1.4% of probands carry a pathogenic allele of a LS gene (*MLH1*, *MSH2*, *PMS2*, or *MSH6*), 2.9% carry a pathogenic allele of a high-risk HBOC gene (*BRCA1*, *BRCA2*, or *PALB2*), and an additional 5% carry a pathogenic allele of a moderate-risk homologous recombination repair pathway breast cancer gene (i.e., *ATM*, *CHEK2*, or *NBN*). Among PC cases, for each of these three groups of genes, the burden of pathogenic alleles is significantly higher than in the population sample represented by non-TCGA ExAC, supporting an argument that these are not incidental findings. Adding

other high-risk genes such as *TP53*, *CDKN2A*, and *STK11* to the mix results in >10% of PC cases with a sequence variant that would alter medical management of healthy at-risk relatives. Focusing on individual genes, the top five genes with potential medical impact for at-risk relatives were *BRCA2, ATM, BRCA1, CDKN2A*, and *CHEK2*. Here we note that *ATM* and *CHEK2* have recently been added to NCCN's list of genes with associated medical actionability for BC.[22]

Turning to CRC, pathogenic variants in *BRCA2* and *CHEK2* were observed in target cousin pairs with unexplained CRC. While it is no longer a surprise that multigene panel testing identifies variants in genes that do not match the sentinel cancer, the more important point is that from the perspective of potential cancer prevention in at-risk relatives, a panel test can deliver more actionable results than would a test limited to syndromic genes.[5–9,67]

Today, universal LS testing, beginning with an immunohistochemical (IHC) or microsatellite instability (MSI)-based prescreen of tumors followed by germline testing for indicated individuals, is already recommended for newly diagnosed CRC cases. For perspective on this universal testing guideline, Erten et al. recently reported from a pooled analysis that prescreen based universal testing finds pathogenic LS variants in about 1.2% of all CRC cases, whereas direct sequencing based universal testing finds pathogenic LS variants in about 2.5% of CRC cases.[68–75] The differences in detection rate between the two strategies was attributed largely to loss to follow-up. A health economics analysis within Erten et al also concluded that when the cost of sequencing the mismatch repair genes drops below $609, universal testing based on sequencing alone will become more cost effective than universal testing that cascades through IHC or MSI, echoing a similar finding by Gould-Suarez et al.[75,76]

Based on our results, universal testing of PC patients using a panel test would identify pathogenic LS variants in about same proportion of cases as does prescreen based universal testing of CRC patients. Moreover, the panel test's ability to identify pathogenic variants in the high-risk HBOC genes plus susceptibility genes such as *TP53*, *CDKN2A*, and *STK11*, results in a proportion of PC patients with a high-risk variant that is notably higher than direct sequencing for LS (alone) finds in CRC patients. Since two companies, Color Genomics and Invitae, already offer panel tests at less than $609, there are really two factors that stand in the way of adopting universal testing for PC patients: (1) systematic illogic in insurance coverage determinations, (2) inefficiency in classification of VUS in high-risk susceptibility genes.

For several of these high-risk susceptibility genes, solutions to the VUS problem are almost in hand. Calibrated sequence analysis-based prior probability of pathogenicity models already exist for the four LS genes, *BRCA1*, and *BRCA2*,[18,19] and there is no obvious barrier to extending these models to TP53. High quality functional assays also exist for missense substitutions in these seven genes.[77–82] Therefore, universal panel testing of PC patients followed by collation of the obviously pathogenic variants, systematic application of the Bayesian Integrated Evaluation to the high probability of pathogenicity VUS found through the testing,[83–86] and cascade testing of at-risk relatives has the potential to add years to the lives of these at-risk relatives in much the same way as does similar testing of CRC cases and BC/OC cases suspected of HBOC.

References

1. Walsh, T. *et al.* Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 12629–33 (2010).

2. Easton, D. F. *et al.* Gene-panel sequencing and the prediction of breast-cancer risk. *N. Engl. J. Med.* **372,** 2243–57 (2015).

3. Hall, M. J., Forman, A. D., Pilarski, R., Wiesner, G. & Giri, V. N. Gene panel testing for inherited cancer risk. *J. Natl. Compr. Canc. Netw.* **12,** 1339–46 (2014).

4. Salo-Mullen, E. E. *et al.* Identification of germline genetic mutations in patients with pancreatic cancer. *Cancer* **121,** 4382–8 (2015).

5. Kurian, A. W. *et al.* Clinical evaluation of a multiple-gene sequencing panel for hereditary cancer risk assessment. *J. Clin. Oncol.* **32,** 2001–2009 (2014).

6. Saam, J. *et al.* Patients tested at a laboratory for hereditary cancer syndromes show an overlap for multiple syndromes in their personal and familial cancer histories. *Oncology* **89,** 288–93 (2015).

7. Howarth, D. R. *et al.* Initial results of multigene panel testing for hereditary breast and ovarian cancer and Lynch syndrome. *Am. Surg.* **81,** 941–4 (2015).

8. Taylor, J. C. *et al.* Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nat. Genet.* **47,** 717–726 (2015).

9. Yurgelun, M. B. *et al.* Identification of a variety of mutations in cancer-predisposition genes in patients with suspected Lynch syndrome. *Gastroenterology* **149,** 604-13 (2015).

10. Kamat, A. M. Commentary on "Risks of primary extracolonic cancers following colorectal cancer in Lynch syndrome." *Urol. Oncol. Semin. Orig. Investig.* **31,** 716 (2013).

11. Rainville, I. R. & Rana, H. Q. Next-generation sequencing for inherited breast cancer risk: counseling through the complexity. *Curr. Oncol. Rep.* **16,** 371 (2014).

12. Mauer, C. B., Pirzadeh-Miller, S. M., Robinson, L. D. & Euhus, D. M. The integration of next-generation sequencing panels in the clinical cancer genetics practice: an institutional experience. *Genet. Med.* **16,** 1–6 (2013).

13. Yurgelun, M. B. *et al.* Identification of a variety of mutations in cancer predisposition genes in patients with suspected Lynch syndrome. *Gastroenterology* **149,** 604e20-613e20 (2015).

14. Thompson, E. R. *et al.* Panel Testing for Familial Breast Cancer: Calibrating the Tension Between Research and Clinical Care. *J. Clin. Oncol.* (2016). doi:10.1200/JCO.2015.63.7454

15. Vallée, M. P. *et al.* Classification of missense substitutions in the BRCA genes: a database dedicated to Ex-UVs. *Hum. Mutat.* **33,** 22–8 (2012).

16. Thompson, B. & Spurdle, A. Microsatellite instability use in mismatch repair gene sequence variant classification. *Genes (Basel).* **6,** 150–162 (2015).

17. Thompson, B. a. *et al.* A multifactorial likelihood model for MMR gene variant classification incorporating probabilities based on sequence bioinformatics and tumor characteristics: a report from the colon cancer family registry. *Hum. Mutat.* **34,** 200–9 (2013).

18. Vallée, M. P. *et al.* Adding in silico assessment of potential splice aberration to the integrated evaluation of BRCA gene unclassified variants. *Hum. Mutat.* **37,** 627-39 (2016).

19. Thompson, B. a *et al.* Calibration of multiple in silico tools for predicting pathogenicity of mismatch repair gene missense substitutions. *Hum. Mutat.* **34,** 255–65 (2013).

20. Tavtigian, S. V, Greenblatt, M. S., Lesueur, F. & Byrnes, G. B. In silico analysis of missense substitutions using sequence-alignment based methods. *Hum. Mutat.* **29,** 1327–36 (2008).

21. Tavtigian, S. V, Byrnes, G. B., Goldgar, D. E. & Thomas, A. Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. *Hum. Mutat.* **29,** 1342–54 (2008).

22. Daly, M. B. *et al.* Genetic/familial high-risk assessment: breast and ovarian, version 2.2015. *J. Natl. Compr. Canc. Netw.* **14,** 153–62 (2016).

23. Plaskocinska, I. *et al.* New paradigms for BRCA1/BRCA2 testing in women with ovarian cancer: results of the genetic testing in epithelial ovarian cancer (GTEOC) study. *J. Med. Genet.* **2013,** 1–7 (2016).

24. Lancaster, J. M., Powell, C. B., Chen, L. M. & Richardson, D. L. Society of gynecologic oncology statement on risk assessment for inherited gynecologic cancer predispositions. *Gynecol. Oncol.* **136,** 3–7 (2015).

25. Berg, A. O., *et al.* Recommendations from the EGAPP Working Group: genetic testing strategies in newly diagnosed individuals with colorectal cancer aimed at reducing morbidity and mortality from Lynch syndrome in relatives. *Genet. Med.* **11,** 35–41 (2009).

26. Provenazle, D., Gupta, S., Ahnen, D., Alsanian, H. & Al., E. NCCN clinical practice guidelines in oncology: genetic/familial high-risk assessment: colorectal. *Natl. Compr. Cancer Netw.* **11,** (2015).

27. Whitcomb, D. C., Shelton, C. A. & Brand, R. E. Genetics and genetic testing in pancreatic cancer. *Gastroenterology* **149,** 1–13 (2015).

28. Harinck, F. *et al.* Is early diagnosis of pancreatic cancer fiction? Surveillance of individuals at high risk for pancreatic cancer. *Dig. Dis.* **28,** 670–8 (2010).

29. Chari, S. T. *et al.* Early detection of sporadic pancreatic cancer: summative review. *Pancreas* **44,** 693–712 (2015).

30. Vasen, H. *et al.* Benefit of surveillance for pancreatic cancer in high-risk individuals: outcome of long-term prospective follow-up studies from three European expert centers. *J. Clin. Oncol.* **34,** 2010–9 (2016).

31. Catts, Z. A.-K. *et al.* Statewide retrospective review of familial pancreatic cancer in Delaware, and frequency of genetic mutations in pancreatic cancer kindreds. *Ann. Surg. Oncol.* **99,** (2016).

32. Kim, D. H., Crawford, B., Ziegler, J. & Beattie, M. S. Prevalence and characteristics of pancreatic cancer in families with BRCA1 and BRCA2 mutations. *Fam. Cancer* **8,** 153–8 (2009).

33. Grant, R. C. *et al.* Prevalence of germline mutations in cancer predisposition genes in patients with pancreatic cancer. *Gastroenterology* **148,** 556–564 (2015).

34. Holter, S. *et al.* Germline BRCA mutations in a large clinic-based cohort of patients with pancreatic adenocarcinoma. *J. Clin. Oncol.* **33,** 3124–3129 (2015).

35. Hahn, S. A. *et al.* BRCA2 germline mutations in familial pancreatic carcinoma. *J. Natl. Cancer Inst.* **95,** 214–21 (2003).

36. Easton, D. F., Matthews, F. E., Ford, D., Swerdlow, A. J. & Peto, J. Cancer mortality in relatives of women with ovarian cancer: the OPCS Study. Office of population censuses and surveys. *Int. J. Cancer* **65,** 284–94 (1996).

37. Lal, G. *et al.* Inherited predisposition to pancreatic adenocarcinoma: role of family history and germ-line p16, BRCA1, and BRCA2 mutations. *Cancer Res.* **60,** 409–16 (2000).

38. Humphris, J. L. *et al.* Clinical and pathologic features of familial pancreatic cancer. *Cancer* **120,** 1–7 (2014).

39. Lucas, A. L. *et al.* BRCA1 and BRCA2 germline mutations are frequently demonstrated in both high-risk pancreatic cancer screening and pancreatic cancer cohorts. *Cancer* **120,** 1–8 (2014).

40. Skolnick, M. A. R. K. The Utah genealogical database: a resource for genetic epidemiology. *Banbury Rep.* **4,** 285–297 (1980).

41. Kerber, R. A. Method for calculating risk associated with family history of a disease. *Genet. Epidemiol.* **12,** 291–301 (1995).

42. Do, H. & Dobrovic, A. Dramatic reduction of sequence artefacts from DNA isolated from formalin-fixed cancer biopsies by treatment with uracil-DNA glycosylase. *Oncotarget* **3,** 546–558 (2012).

43. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).

44. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43,** 491–8 (2011).

45. Wildeman, M., van Ophuizen, E., den Dunnen, J. T. & Taschner, P. E. M. Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker. *Hum. Mutat.* **29,** 6–13 (2008).

46. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38,** e164 (2010).

47. Yeo, G. & Burge, C. B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **11,** 377–94 (2004).

48. Tavtigian, S. V *et al.* Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J. Med. Genet.* **43,** 295–305 (2006).

49. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using Polyphen-2. *Curr. Protoc. Hum. Genet.* **Chapter 7,** Unit7.20 (2013).

50. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46,** 310–5 (2014).

51. Young, E. L. *et al.* Multigene testing of moderate-risk genes: be mindful of the missense. *J. Med. Genet.* **53,** 366–76 (2016).

52. Stone, E. a & Sidow, A. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Res.* **15,** 978–86 (2005).

53. Johnston, H. R., Hu, Y. & Cutler, D. J. Population genetics identifies challenges in analyzing rare variants. *Genet. Epidemiol.* **39,** 145–148 (2015).

54. Freeman, M. F. & Tukey, J. W. Transformations related to the angular and the square root. *Ann. Math. Stat.* **21,** 607–611 (1950).

55. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans.

Nature **536,** 285-91 (2016)

56. Rosner, B. *Fundamentals of biostatistics*. (Duxbury Press, 1986).

57. Sopik, V., Rosen, B., Giannakeas, V. & Narod, S. a. Why have ovarian cancer mortality rates declined? Part III. Prospects for the future. *Gynecol. Oncol.* **138,** 757-61 (2015).

58. Roeb, W., Higgins, J. & King, M. Response to DNA Damage of CHEK2 missense mutations in familial breast cancer. *Hum. Mol. Genet.* **21,** 2738–44 (2012).

59. Heinimann, K. *et al.* Nontruncating APC germ-line mutations and mismatch repair deficiency play a minor role in APC mutation-negative polyposis. *Cancer Res.* **61,** 7616–7622 (2001).

60. Menéndez, M. *et al.* Functional characterization of the novel APC N1026S variant associated with attenuated familial adenomatous polyposis. *Gastroenterology* **134,** 56–64 (2008).

61. Kerr, S. E., Thomas, C. B., Thibodeau, S. N., Ferber, M. J. & Halling, K. C. APC germline mutations in individuals being evaluated for familial adenomatous polyposis: a review of the mayo clinic experience with 1591 consecutive tests. *J. Mol. Diagnostics* **15,** 31–43 (2013).

62. Hu, C. *et al.* Prevalence of pathogenic mutations in cancer predisposition genes among pancreatic cancer patients. *Cancer Epidemiol. Biomarkers Prev.* **25,** 207–11 (2016).

63. Waddell, N. *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518,** 495–501 (2015).

64. Zhen, D. B. *et al.* BRCA1, BRCA2, PALB2, and CDKN2A mutations in familial pancreatic cancer: a PACGENE study. *Genet. Med.* **17,** 569–77 (2014).

65. Roberts, N. J. *et al.* Whole genome sequencing defines the genetic heterogeneity of familial pancreatic cancer. *Cancer Discov.* **6,** 166–75 (2016)

66. Lynch, H. T., Voorhees, G. J., Lanspa, S. J., McGreevy, P. S. & Lynch, J. F. Pancreatic carcinoma and hereditary nonpolyposis colorectal cancer: a family study. *Br. J. Cancer* **52,** 271–3 (1985).

67. Cragun, D. *et al.* Panel-based testing for inherited colorectal cancer: A descriptive study of clinical testing performed by a US laboratory. *Clin. Genet.* **86,** 510–520 (2014).

68. Hartman, D. J. *et al.* Lynch syndrome-associated colorectal carcinoma: frequent involvement of the left colon and rectum and late-onset presentation supports a

universal screening approach. *Hum. Pathol.* **44,** 2518–2528 (2013).

69. Heald, B. *et al.* Implementation of universal microsatellite instability and immunohistochemistry screening for diagnosing Lynch syndrome in a large academic medical center. *J. Clin. Oncol.* **31,** 1336–1340 (2013).

70. Kidambi, T. D. *et al.* Selective versus universal screening for Lynch syndrome: a six-year clinical experience. *Dig. Dis. Sci.* **60,** 2463–2469 (2014).

71. Musulén, E., Sanz, C., Muñoz-Mármol, A. M. & Ariza, A. Mismatch repair protein immunohistochemistry: a useful population screening strategy for Lynch syndrome. *Hum. Pathol.* **45,** 1388–1396 (2014).

72. Hampel, H. *et al.* Feasibility of screening for Lynch syndrome among patients with colorectal cancer. *J. Clin. Oncol.* **26,** 5783–8 (2008).

73. Chang, S. C. *et al.* Taiwan hospital-based detection of Lynch syndrome distinguishes 2 types of microsatellite instabilities in colorectal cancers. *Surgery* **147,** 720–728 (2010).

74. Hampel, H. *et al.* Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). *N. Engl. J. Med.* **352,** 1851–60 (2005).

75. Erten, M. Z. *et al.* Universal versus targeted screening for Lynch syndrome: comparing ascertainment and costs based on clinical experience. *Dig. Dis. Sci.* **61**, 2887-95 (2016).

76. Gould-Suarez, M., El-Serag, H. B., Musher, B., Franco, L. M. & Chen, G. J. Cost-effectiveness and diagnostic effectiveness analyses of multiple algorithms for the diagnosis of Lynch syndrome. *Dig. Dis. Sci.* **59,** 2913–2926 (2014).

77. Drost, M. *et al.* A cell-free assay for the functional analysis of variants of the mismatch repair protein MLH1. *Hum. Mutat.* **31,** 247–253 (2010).

78. Drost, M. *et al.* A rapid and cell-free assay to test the activity of lynch syndrome-associated MSH2 and MSH6 missense variants. *Hum. Mutat.* **33,** 488–494 (2012).

79. Drost, M., Koppejan, H. & De Wind, N. Inactivation of DNA mismatch repair by variants of uncertain significance in the PMS2 gene. *Hum. Mutat.* **34,** 1477–1480 (2013).

80. Kato, S. *et al.* Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. *Proc Natl Acad Sci U S A* **100,** 8424–8429 (2003).

81. Guidugli, L. *et al.* A classification model for BRCA2 DNA binding domain missense variants based on homology-directed repair activity. *Cancer Res.* **73,**

265–275 (2013).

82. Woods, N. T. *et al.* Functional assays provide a robust tool for the clinical annotation of genetic variants of uncertain significance. *npj Genomic Med.* **1,** 16001 (2016).

83. Goldgar, D. E. *et al.* Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. *Am. J. Hum. Genet.* **75,** 535–44 (2004).

84. Easton, D. F. *et al.* A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes. *Am. J. Hum. Genet.* **81,** 873–83 (2007).

85. Rasmussen, L. J. *et al.* Pathological assessment of mismatch repair gene variants in Lynch syndrome: past, present, and future. *Hum. Mutat.* **33,** 1617–1625 (2012).

86. Thompson, B. a *et al.* Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. *Nat. Genet.* **46,** 107–15 (2014).

Figure 3.1. Flow chart of methods

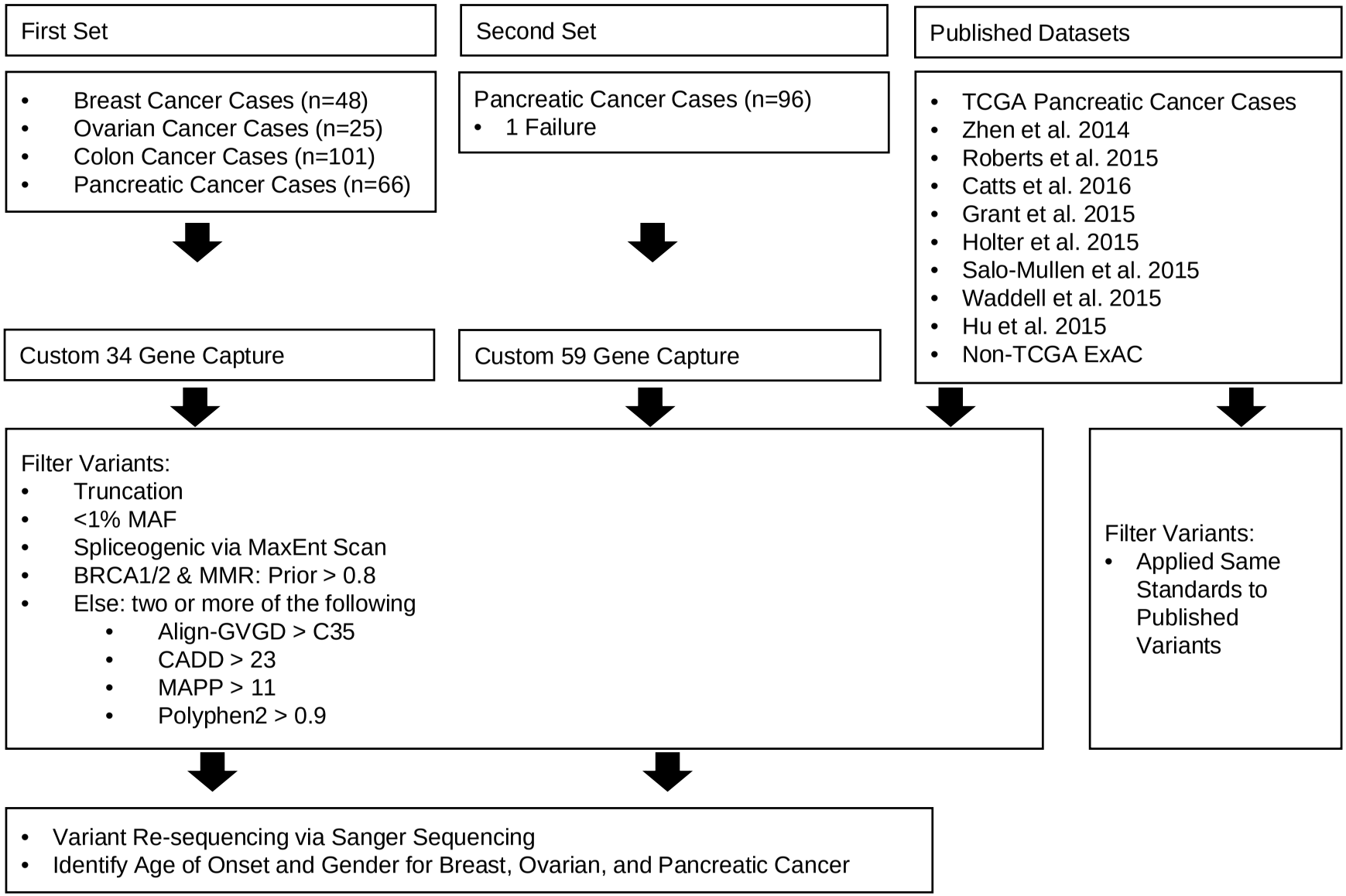

First Set
Second Set
Published Datasets
• Breast Cancer Cases (n=48)
• Ovarian Cancer Cases (n=25)
• Colon Cancer Cases (n=101)
• Pancreatic Cancer Cases (n=66)
Pancreatic Cancer Cases (n=96)
• 1 Failure
• TCGA Pancreatic Cancer Cases
• Zhen et al. 2014
• Roberts et al. 2015
• Catts et al. 2016
• Grant et al. 2015
• Holter et al. 2015
• Salo-Mullen et al. 2015
• Waddell et al. 2015
• Hu et al. 2015
• Non-TCGA ExAC
Custom 34 Gene Capture
Custom 59 Gene Capture
Filter Variants:
• Truncation
• <1% MAF
• Spliceogenic via MaxEnt Scan
• BRCA1/2 & MMR: Prior > 0.8
• Else: two or more of the following
• Align-GVGD > C35
• CADD > 23
• MAPP > 11
• Polyphen2 > 0.9
Filter Variants:
• Applied Same Standards to Published Variants
• Variant Re-sequencing via Sanger Sequencing
• Identify Age of Onset and Gender for Breast, Ovarian, and Pancreatic Cancer

Figure 3.2. Forest plots of the proportion of carriers in pancreatic cancer screens. A meta-analysis between the unselected PC cases from HCH, the Mayo Clinic, and the PC cases of The Cancer Genome Atlas (TCGA). A list of genes for each analysis is found in Supplemental Table 1. HBOC = Hereditary Breast and Ovarian Cancer; HRR = Homologous Recombination and Repair; ICR = Interstrand Crosslink Repair; OC = Ovarian Cancer; BC = Breast Cancer

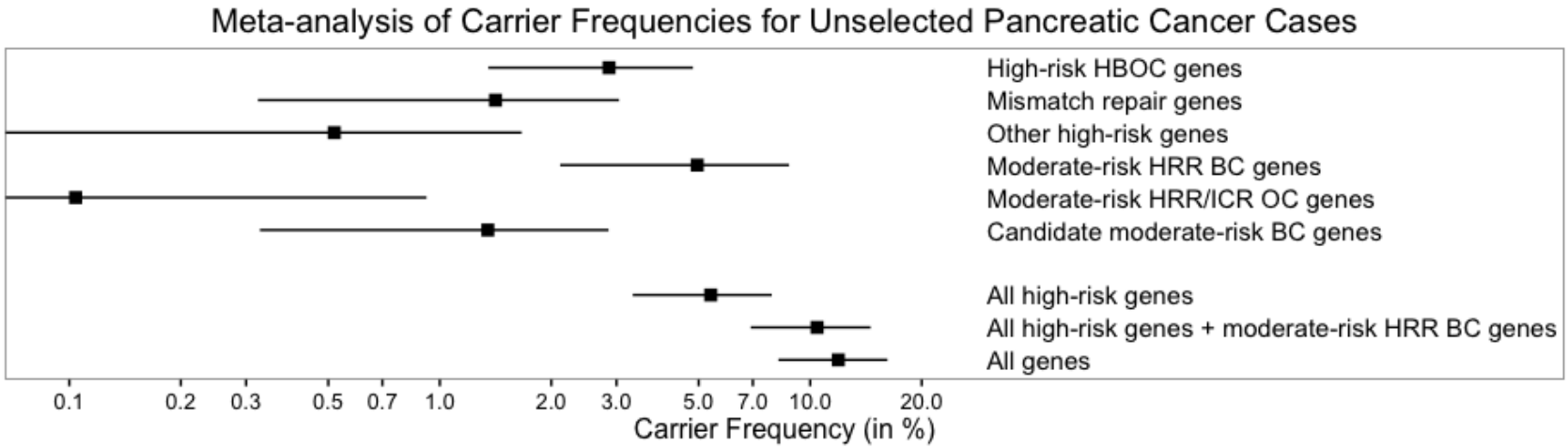

Meta-analysis of Carrier Frequencies for Unselected Pancreatic Cancer Cases
High-risk HBOC genes
Mismatch repair genes
Other high-risk genes
Moderate-risk HRR BC genes
Moderate-risk HRR/ICR OC genes
Candidate moderate-risk BC genes
All high-risk genes
All high-risk genes + moderate-risk HRR BC genes
All genes
0.1
0.2
0.3
0.5
0.7
1.0
2.0
3.0
5.0
7.0
10.0
20.0
Carrier Frequency (in %)

Table 3.1: Demographics for 34 and 59 gene screen

	Total	% Female	Average Age of Onset (low-high)
Breast Cancer Cases	48	100	51.9 (28-84)
Luminal	20	100	51.5 (30-84)
HER2+	15	100	51.5 (30-77)
Triple Negative	13	100	53.3 (28-78)
Ovarian Cancer Cases	25	100	55.1 (34-82)
Pancreatic Cancer Cases (First Set)	66	47.0	66.6 (43-91)
Pancreatic Cancer Cases (Second Set)	95	44.8	66.0 (43-86)
Colorectal Cancer Cases	98	47.0	64.8 (28-94)

Table 3.2: Pathogenic variants and variants of high probability of pathogenicity identified from a custom 34-gene panel for breast cancer, ovarian cancer, colorectal cancer, and pancreatic cancer cases

Gene	HGVS	Sex: Age of Onset (Subtype)	Weight
Breast Cancer Cases			
BRCA1	c.856_857del p.(S286*)	F: 41 (Triple Negative)	1[a]
BRCA2	c.1189_1190insTTAG p.(Q397Lfs*25)	F: 30 (Luminal)	1
BRCA2	c.1832C>A p.(S611*)	F: 37 (Luminal)	1
MRE11A	c.564G>C p.(R188S)	F: 41 (Triple Negative)	-[a]
	Carrier Frequency:		
	• Luminal	2/20=10% (1.23-31.70%)	
	• HER2+	0/15=0%	
	• Triple Negative	1/13=7.69% (0.19-36.03%)	
Ovarian Cancer Cases			
BRCA1	c.3066del p.(V1023*)	F: 59	1
BRCA2	c.6275_6276del p.(L2092Pfs*7)	F: 58	1
MSH2	c.2260A>G p.(T754A)	F: 53	0.96
MSH6	c.3674C>T p.(V1253E)	F: 41	0.94
RAD50	c.1958C>A p.(S653*)	F: 46	1
RAD51C	c.577C>T p.(R193*)	F: 55	1
	Carrier Frequency:	5.90/25=23.6% (9.36-45.13%)	
Colorectal Cancer Cases			
BRCA2	c.93G>T p.(W31C)	F: 61	0.81
BRCA2	c.6275_6276del p.(L2092Pfs*7)	F: 59	1
CHEK2	c.1100del p.(T367Mfs*15)	M: 49	1
MSH2	EPCAM & MSH2 deletion[b]	F: 28	1
		M: 30	1
MSH2	c.592G>T p.(E198*)	F: 63	1
MSH2	c.1375G>C p.(D459H)	M: 67	0.87
MSH2	c.2662del p.(L888Cfs*4)	M: 33	1
		M: 66	1
Number of Identified Variants with Possible BC Recommendations: 3/7=42.9%			
Pancreatic Cancer Cases			
ATM	c.7327C>G p.(R2443G)	M: 66	0.81
ATM	c.8734A>G p.(R2912G)	F: 65	0.81
BRCA2	c.3873del p.(Q1291Hfs*2)	M: 54	1
MSH6	c.3261dup p.(F1088Lfs*5)	F: 57	1
PALB2	c.1240C>T p.(R414*)	F: 69	1
STK11	c.738C>A p.(Y246*)	M: 45	1
	Carrier Frequency:	5.62/66=8.52% (3.41-18.7%)	

[a] The same individual carried both variants, so carrier weight was combined and only counted once. [b]Not confirmed via secondary means

Table 3.3: Pathogenic variants and variants with high probability of pathogenicity identified in a second set of 95 pancreatic cancer cases, unselected for family history

Gene	HGVS	Sex: Age of Onset	Carrier
ATM	c.8734A>G p.(R2912G)	M: 74	0.81
BRCA1	c.68_69del p.(E23Vfs*17)	M: 74	1
BRCA2	c.3974_3975insTGCT p.(T1325Cfs*4)	M: 71	1
CHEK2	c.1159A>G p.(T387A)	M: 81	0.81
CHEK2	c.1427C>T p.(T476M)	F: 50	0.99[a]
MRE11A	c.923dupT p.(M309Hfs*8)	M: 56	1
MSH6	c.3851C>T p.(T1284M)	F: 68	0.94
RAD50	c.3641G>A p.(R1214H)	F: 50	-[a]
TP53	c.847C>T p.(R283C)	M: 65	0.81
	Carrier Frequency:	7.36/95=7.75% (3.01-14.6%)	

[a]Same individual carried both variants, so weight was combined and only counted once.

Table 3.4: Standardized incidence ratios for cancer susceptibility gene groups in pancreatic cancer cases across recent literature

	Cases (weighted)	Case carrier frequency	Non-TCGA ExAC (weighted)	Non-TCGA ExAC carrier frequency	SIR	95%-CI	p-value
High-risk HBOC genes	150 (149.81)	0.056	369 (358.65)	0.007	7.56	[6.38; 8.86]	<0.0001
Mismatch repair genes	23 (22.40)	0.025	526 (484.61)	0.010	2.43	[1.52; 3.64]	0.0006
Other high-risk genes	46 (45.81)	0.027	136 (119.86)	0.003	10.46	[7.59; 13.88]	<0.0001
Moderate-risk HRR BC genes	58 (56.27)	0.052	1117 (1001.83)	0.021	2.51	[1.89; 3.24]	<0.0001
Moderate-risk HRR/ICR OC genes	6 (6)	0.007	134 (133.07)	0.003	2.71	[0.99; 5.89]	0.0515
Candidate moderate-risk BC genes	7 (6.43)	0.014	584 (488.86)	0.010	1.43	[0.49; 2.91]	0.5919
All high-risk genes	219 (218.26)	0.107	1031 (963.12)	0.020	5.34	[4.65; 6.09]	<0.0001
All high-risk & moderate-risk HRR BC genes	277 (274.53)	0.159	2148 (1964.95)	0.041	3.91	[3.46; 4.40]	<0.0001
All genes	329 (324.55)	0.211	3459 (3090.67)	0.064	3.30	[2.95; 3.68]	<0.0001

HBOC = Hereditary Breast and Ovarian cancer; HRR = Homologous Recombination and Repair; BC = Breast Cancer; ICR = Interstrand Crosslink Repair; OC = Ovarian Cancer; TCGA = The Genome Atlas; SIR = Standardized Incidence Ratios; CI = Confidence Intervals

Supplemental Table S3.1. List of genes for custom 34 and 59 gene capture and inclusion in meta-analysis and standardized incident ratio calculations

Gene	34 Gene Panel	59 Gene Panel	High-risk HBOC genes	Mismatch Repair genes	Other High-risk genes	Moderate-risk HRR BC genes	Moderate-risk HRR/ICR OC genes	Candidate moderate-risk BC genes	All high-risk genes	All high-risk & moderate-risk HRR BC genes	All genes
ALK		Y									
APC	Y	Y			Y[a]				Y[a]	Y[a]	Y[a]
ATM	Y	Y				Y[c]				Y[c]	Y[c]
AXIN2	Y	Y									
BAP1		Y									
BARD1	Y	Y						Y[d]			Y[d]
BMPR1A	Y	Y			Y[a]				Y[a]	Y[a]	Y[a]
BRCA1	Y	Y	Y						Y	Y	Y
BRCA2	Y	Y	Y						Y	Y	Y
BRIP1	Y	Y					Y[a]			Y[a]	Y[a]
CDH1	Y	Y						Y[a]			Y[a]
CDK4		Y									
CDKN2A	Y	Y			Y[a]				Y[a]	Y[a]	Y[a]
CHEK2	Y	Y				Y[b]				Y[b]	Y[b]
CTNNB1	Y	Y									
DICER1		Y									
FLCN	Y	Y									
MAX		Y									
MEN1		Y									
MLH1	Y	Y		Y					Y	Y	Y
MLH3	Y	Y									
MRE11A	Y	Y						Y[e]			Y[e]
MSH2	Y	Y		Y					Y	Y	Y
MSH6	Y	Y		Y					Y	Y	Y
MUTYH	Y	Y									
NBN	Y	Y				Y[g]				Y[g]	Y[g]
NF1		Y						Y			Y
PALB2	Y	Y	Y[a]						Y[a]	Y[a]	Y[a]
PHOX2B		Y									
PKHD1		Y									
PMS1	Y	Y									
PMS2	Y	Y		Y[b]					Y[b]	Y[b]	Y[b]
POLD1		Y									
POLE		Y									
POT1		Y									
PRSS1		Y									
PTCH1		Y									
PTEN	Y	Y			Y[a]				Y[a]	Y[a]	Y[a]
RAD50	Y	Y						Y[f]			Y[f]
RAD51B	Y	Y									
RAD51C	Y	Y					Y			Y	Y
RAD51D							Y			Y	Y
RET		Y									
RINT1		Y									
SDHA		Y									
SDHAF2		Y									

Supplemental Table S3.1. continued

Gene	34 Gene Panel	59 Gene Panel	High-risk HBOC genes	Mismatch Repair genes	Other High-risk genes	Moderate-risk HRR BC genes	Moderate-risk HRR/ICR OC genes	Candidate moderate-risk BC genes	All high-risk genes	All high-risk & moderate-risk HRR BC genes	All genes
SDHD		Y									
SMARCB1		Y									
SMAD4					Y[a]				Y[a]	Y[a]	Y[a]
STK11	Y	Y			Y[a]				Y[a]	Y[a]	Y[a]
SUFU		Y									
TCF7L2	Y	Y									
TGFBR2	Y	Y									
TMEM127		Y									
TP53	Y	Y			Y				Y	Y	Y
VHL		Y									
WT1		Y									
XRCC2	Y	Y									
XRCC3	Y	Y									

[a] Only clearly deleterious truncating and splice junction variants as well as known pathogenic missense were included; [b] *CHEK2* exons 13-16 and *PMS2* exons 1-5, 9, and 11-15 were excluded due to possible pseudo-gene contamination in ExAC[1]; [c] Only missense substitutions position 1960+ were included[2]; [d] Only missense substitutions in the RING, BRCA1 interaction, and BRCT domains were included[3]; [e] Only missense substitutions in the Phosphoesterase, RAD50 interaction, and GAR domains were included[4]; [f] Only missense substitutions in the ATPase, MRE11A interaction, and Zinc hook domains were included[4]; [g] Only missense substitutions in the GHA, BRCT, and MRE11A domains were included[4]. HBOC = Hereditary Breast and Ovarian Cancer; HRR = Homologous Recombination and Repair; BC = Breast cancer; ICR = Interstrand Crosslink Repair; OC= Ovarian Cancer.

Supplemental Table S3.2. List of rare variants with corresponding information used to identify potentially pathogenic variants and weights for patients screened from the Huntsman Cancer Hospital

Gene	Accession	DNA Change	Protein Change	Type	ExAC freq	Prior	A	M	C	P	W
ATM	NM_000051	c.6067G>A	p.(G2023R)	MS	0.0016	NA	C65	45.39	30	0.999	0
ATM	NM_000051	c.7327C>G	p.(R2443G)	MS	$8.334X10^{-6}$	NA	C65	25.9	17.45	0.984	0.81
ATM	NM_000051	c.8734A>G	p.(R2912G)	MS	0.0002883	NA	C65	25.9	20.7	1	0.81
BARD1	NM_000465	c.2216A>G	p.(K706E)	MS	5.65E-05	NA	C0	16.22	16.9	0.265	0
BMPR1A	NM_004329	c.1385C>T	p.(P462L)	MS	0	NA	C65	26.92	28.4	1	0
BRCA1	NM_007294	c.68_69del	p.(E23Vfs*17)	FS	0.0002	-	-	-	-	-	1
BRCA1	NM_007294	c.856_857del	p.(S286*)	FS	0	-	-	-	-	-	1
BRCA1	NM_007294	c.3066del	p.(V1023*)	FS	0	-	-	-	-	-	1
BRCA1	NM_007294	c.3708T>G	p.(N1236K)	MS	0.0002389	0.02	C0	-	-	0.011	0
BRCA1	NM_007294	c.4165_4166del	p.(S1389*)	FS	$1.059X10^{-5}$	-	-	-	-	-	1
BRCA2	NM_000059	c.93G>T	p.(W31C)	MS	0	0.81	C65	-	20.7	1	0.81
BRCA2	NM_000059	c.1189_1190insTTAG	p.(Q397Lfs*25)	FS	0	-	-	-	-	-	1
BRCA2	NM_000059	c.1832C>A	p.(S611*)	SG	0	-	-	-	38	-	1
BRCA2	NM_000059	c.3873del	p.(Q1291Hfs*2)	FS	0	-	-	-	-	-	1
BRCA2	NM_000059	c.3975_3978dup	p.(A1327Cfs*4)	FS	$9.678X10^{-6}$	-	-	-	-	-	1
BRCA2	NM_000059	c.6275_6276del	p.(L2092Pfs*7)	FS	0	-	-	-	-	-	1
CHEK2	NM_001005735	c.1100del	p.(T367Mfs*15)	FS	0.001818	NA	-	-		-	1
CHEK2	NM_001005735	c.1159A>G	p.(T387A)	MS	0	NA	C65	24.17	26.6	0.973	0.81
CHEK2	NM_001005735	c.1427C>T	p.(T476M)	MS	0.0004	NA	C65	23.02	19.54	0.997	0.95[b]
DICER1	NM_177438	c.5623G>A	p.(D1875N)	MS	0	NA	C15	13.97	33	0.176	0.66
EPCAM		EPCAM:ex1-9_MSH2:ex1-8		D	-	-	-	-	-	-	1
MLH1	NM_000249	c.1136A>G	p.(Y379C)	MS	3.77E-05	0.46	C15	-	16.5	0.966	0
MLH3	NM_014381	c.2221G>T	p.(V741F)	MS	0.015	NA	-	-	6.398	0.518	0
MLH3	NM_014381	c.3539G>A	p.(R1180H)	MS	0.0001	NA	C25	32	14.41	0.999	0
MLH3	NM_014381	c.716T>G	p.(I239S)	MS	0.0001	NA	C65	20.83	17.29	0.984	0
MRE11A	NM_005591	c.564G>C	p.(R188S)	MS	0	NA	C65	21.57	17.31	1	0.81
MRE11A	NM_005591	c.826C>T	p.(P276S)	MS	0	NA	C0	6.04	26.8	0.856	0
MRE11A	NM_005591	c.923dup	p.(M309Hfs*8)	FS	0	NA	-	-		-	1
MSH2	NM_000251	c.138C>G	p.(H46Q)	MS	1.89E-04	0.74	C15	14.29	26.7	1	0
MSH2	NM_000251	c.260C>G	p.(S87C)	MS	0.0003076	0.79	C15	20.41	15.8	0.961	0
MSH2	NM_000251	c.592G>T	p.(E198*)	SG	0	-	-	-	24.3	-	1
MSH2	NM_000251	c.815C>T	p.(A272V)	MS	2.17E-04	0.40	C0	19.40	32	0.944	0
MSH2	NM_000251	c.1375G>C	p.(D459H)	MS	0	0.87	C65	20.32	21.8	0.977	0.87
MSH2	NM_000251	c.2260A>G	p.(T754A)	MS	0.0000165	0.96	C65	37.99	24.7	0.991	0.96
MSH2	NM_000251	c.2662del	p.(L888Cfs*4)	FS	0	-	-	-	-	-	1
MSH6	NM_000179	c.3261dup	p.(F1088Lfs*5)	FS	0.001753	-	-	-	-	-	1
MSH6	NM_000179	c.3674C>T	p.(T1225M)	MS	0.0001156	0.73	C65	29.50	28.7	0.99	0
MSH6	NM_000179	c.3758T>A	p.(V1253E)	MS	0.0001897	0.94	C65	32.45	29.4	0.99	0.94
MSH6	NM_000179	c.3851C>T	p.(T1284M)	MS	0.0002	0.94	C65	29.50	18.23	0.999	0.94
NBN	NM_001024688	c.283G>A	p.(D95N)	MS	0.0019	NA	C0	13.92	18.24	1	0
PALB2	NM_024675	c.1240C>T	p.(R414*)	SG	0	NA	-	-	38	-	1
PMS2	NM_000535	c.86G>C	p.(G29A)	MS	0.0004763	0.89	C55	22.05	32	0.999	0[c]
PMS2	NM_000535	c.953A>G	p.(Y318C)	MS	0.0003567	0.72	C55	14.49	16.74	0.981	0
PMS2	NM_000535	c.1004A>G	p.(N335S)	MS	2.45E-04	0.61	C45	13.40	23.3	0.996	0
PMS2	NM_000535	c.1490G>A	p.(G497D)	MS	7.53E-05	0.001	C0	3.02	10.16	0.382	0
POT1	NR_003102	c.1810G>A	p.(E604K)	MS	0	NA	C55	19.98	26.5	0.907	0.81
PTEN	NM_000314	c.841C>G	p.(P281A)	MS	0	NA	C0	2.41	15.88	0.76	0
RAD50	NM_005732	c.280A>C	p.(I94L)	MS	0.003473	NA	C0	8.94	15.56	0.003	0
RAD50	NM_005732	c.980A>G	p.(R327H)	MS	0.003	NA	C0	5.64	27.7	0.742	0
RAD50	NM_005732	c.1958C>A	p.(S653*)	SG	$4.136X10^{-5}$	NA	-	-	39	-	1
RAD50	NM_005732	c.3641G>A	p.(R1214H)	MS	0	NA	C25	14.41	36	1	0.81
RAD51B	NM_002877	c.728A>G	p.(K243R)	MS	0.007347	NA	C25	-	25.4	1	0
RAD51C	NM_058216	c.577C>T	p.(R193*)	SG	$4.136X10^{-5}$	NA	-	-	22	-	1

Supplemental Table S3.2. continued

Gene	Accession	DNA Change	Protein Change	Type	ExAC freq	Prior	A	M	C	P	W
TGFBR2	NM_001024847	c.1234G>A	p.(V412M)	MS	0.001156	NA	-	-	17.28	0.954	0
TP53	NM_000546	c.847C>T	p.(R283C)	MS	0.0001	NA	C55	26.18	13.04	0.987	0.81
XRCC3	NM_005432	c.448C>T	p.(R150C)	MS	0.0002	NA	C65	29.01	17.5	0.997	0.81
XRCC3	NM_005432	c.617T>G	p.(M206R)	MS	0	NA	C0	32.16	15.77	0.10	0

[a] Because of allele frequency observed from sequencing, this variant is believed to be mosaic. [b] *In silico* predictions plus the additional information from a functional assay,[5] this variant has a higher likelihood of being pathogenic. [c] Although this variant is rare enough and has a prior probability that meets threshold, this variant was identified in homozygous state in the ExAC database, which significantly decreases its chances of being pathogenic. [d] This individual did not have any indication of Cowden's syndrome. PC= Pancreatic Cancer; OC = Ovarian Cancer; BC = Breast Cancer; CRC = Colorectal Cancer; MS = Missense; FS = Framshift; SG = Stopgain; D = Deletion; A = Align-GVGD; M = MAPP; C = CADD; P = Polyphen-2; W = Weight.

Supplemental Table S3.3. List of datasets used for meta-analysis and standardized incidence ratio calculations

Study	Sample Size	Selection Criteria	Genes Reported
HCH PC 1	66	Unselected	Supplemental Table 1
HCH PC 2	95	Unselected	Supplemental Table 1
TCGA PC Cases	154		Supplemental Table 1
Catts 2016	164	Familial PC Families	PRSS1, SPINK, STK11, CDKN2A, APC, MLH1, MSH2, MSH6, PMS2, BRCA1, BRCA2
Holter 2015	306	Unselected	BRCA1, BRCA2
Salo-Mullen 2015	159	Unselected	BRCA1, BRCA2, PALB2, MLH1, MSH2, MSH6, PMS2, FAMMM
Waddell 2015	100	Unselected	BRCA1, BRCA2, PALB2, ATM, FANCM, XRCC4, XRCC6
Hu 2015	96	Unselected	BRCA1, BRCA2, PALB2, ATM, BARD1, BRIP1, RAD51C, RAD51D, CHEK2, MRE11A, NBN, RAD50, MLH1, MSH2, MSH6, PMS2, CDH1,TP53, PTEN, STK11, FANCM
Zhen 2014	727	521/727 unrelated probands met criteria for familial pancreatic cancer	BRCA1, BRCA2, PALB2, CDKN2A
Grant 2015	290	Family History of Breast and/or Ovarian cancer, Pancreatic cancer or neither	APC, ATM, BRCA1, BRCA2, CDKN2A, MLH1, MSH2, MSH6, PALB2, PMS2, PRSS1, STK11, TP53
Roberts 2015	593	Kindreds of Familial Pancreatic Cancer	Whole Genome/87 genes
Non-TCGA ExAC	49,451	Excluding FIN and OTH subgroups	Supplemental Table 1

HCH = Huntsman Cancer Hospital, SLC, UT; PC = Pancreatic Cancer; TCGA = The Cancer Genome Atlas

APPENDIX

PATHWAY-WIDE PROTEIN MULTIPLE SEQUENCE ALIGNMENTS

Introduction

In a polygenic model, multiple variants conferring small amounts of susceptibility can be combined together to estimate risk.[1] In our previously published case-control mutation screening studies of *ATM, CHEK2, MRE11A, NBN, RAD50, RAD51, RINT1*, and *XRCC2*,[2–7] we repeatedly found that the summed frequency of bioinformatically predicted deleterious missense substitutions exceeded that of protein truncating variants. Here, we add *BARD1* mutation screening data to the analysis and apply consistent analytic models across the rare variants from all nine genes detected in an ethnically diverse sample of 1,297 breast cancer cases and 1,121 controls from Chapter 1. Given that the nine intermediate-risk genes operate together in the Homologous Recombination and Repair (HRR) pathway, we explored the combined evaluation of these genes with a concatenated reading frame of 8,611 amino acids. We estimated the scores from the missense substitution analysis programs Align-GVGD[8] and MAPP[9] required to identify a group of missense substitutions that reach an average OR ≥2.5.

We used both Align-GVGD and MAPP[8,9] to assign severity scores to the rare missense substitutions (rMS). For programs that require a user-generated protein multiple

sequence alignment (pMSA), it has been suggested that the pMSA for each gene needs to consist of high quality orthologous sequences with enough variation to average at least three amino acid substitutions per position from orthologous sequences (3S/P).[10–12] This idea, however, creates paradoxes when applied to highly conserved proteins, such as RAD51. In an effort to generalize and smooth the idea of 3S/P, we chose twenty genes from the HRR pathway to create a concatenated pathway–wide pMSA.

Methods

Protein multiple sequence alignment organisms

Orthologs from Human (*Homo sapiens*), either mouse (*Mus musculus*) or rat (*Rattus norvegicus*) from clade Murinae, either pig (*Sus scrofa*), cow (*Bos taurus*), dog (*Canis lupus*) or panda (*Ailuropoda melanoleuca*) from clade Laurasiatheria, elephant (*Loxodonta africana*), armadillo (*Dasypus novemcinctus*), either opossum (*Monodelphis domestica*) or tasmanian devil (*Sarcophilus harrisii*) from clade Metatheria, platypus (*Ornithorhynchus anatinus*), chicken (*Gallus gallus*) and lizard (*Anolis carolinensis*) or painted turtle (*Chrysemys picta bellii*) from clade Sauria, clawed frog (*Xenopus laevis* or *Xenopus tropicalis*), coelacanth (*Latimeria chalumnae*), either zebrafish (*Danio rerio*), Tetraodon (*Tetraodon nigroviridis*), or Fugu (*Takifugu rubripes*) from clade Clupeocephala, lancelet (*Branchiostoma floridae*), and sea urchin (*Strongylocentrotus purpuratus*) were included in our initial alignments. RAD50 and RAD51 did not reach three substitutions per position for their individual pMSA using the preceding organisms, so orthologous protein sequences from the model organisms *Drosophila melanogaster*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Arabidopsis thaliana* were added until three substitutions per position was achieved.

Ortholog sequences downloaded from Genbank were aligned using the expresso extension of T-Coffee to create the initial pMSA[13,14]. These initial alignments were checked by hand in Geneious v7.1.4 (http://www.geneious.com) for anomalies that might be attributed to gene model errors rather than actual sequence divergence. Potential anomalies were corrected by reference to, and gene re-prediction from, genomic DNA sequence available on the UCSC genome browser (http://genome.ucsc.edu).

Protein multiple sequence alignment depth determination

We counted substitutions in each alignment by using Protpars from PHYLIP v3.69[15] with a constrained phylogeny to make a maximum parsimony estimate of the number of substitutions that occurred for each organism in the underlying phylogeny. In order to create a distance for MAPP, we used Protdist to establish a distance matrix for protein sequences using maximum likelihood estimates with a constrained phylogeny followed by Fitch to generate a phylogenetic tree under the additive tree model. The pathway-wide concatenated pMSA's distances were calculated from the sum of substitutions and protein lengths at each node. The complete alignments, phylogenetic trees, and distances are available upon request.

Statistics

To assess evidence of risk from the case-control frequency distribution of protein-truncating variants (T), known or very likely spliceogenic splice-junction variants (SJV), and rare missense substitutions (rMS), we constructed a table with one entry per subject; the variants per subject; and annotations for whether the variant was in a key functional domain, its frequency, as well as study center, case-control status, race/ethnicity, and age

for the subject. For the subjects who carried more than one rare variant of interest, only the most deleterious score was considered. For each of the missense analysis programs, we toggled the program's severity score from a very relaxed to a very stringent value, repeatedly estimating two ORs as the stringency increased: the OR for subjects that carried one or more rMS at or above the score (and no T+SJV), and the OR for subjects who carried a rMS that was below the score (and carried neither a T+SJV nor a higher scoring rMS)(Figure A.1). From this analysis, we determined a threshold severity score for each program at which subjects carrying an rMS at or above the threshold had an average OR≥2.5.

Results

This pathway-wide pMSA reached 3.66 substitutions per position with sequences from Human through *Xenopus* (Table A.1). For variant evaluation, we then adjusted our pMSAs to three depths: Human through Platypus (mammals only), Human through *Xenopus* (pathway-wide 3S/P depth), and Human through the organism required for 3S/P for each individual gene (gene-specific 3S/P depth). Although we were able to find a score for 3S/P such that the average OR ≥2.5 for the more severe variants, we were not able to find a similar cutoff for either mammal or pathway-wide depth.

Discussion

It was interesting that neither Align-GVGD nor MAPP was able to achieve an OR≥2.5 with a pMSA that consisted of only mammals (Human through Platypus) nor the pathway-wide depth (Human through *Xenopus*), suggesting that these sequences, although generally more complete than the more distant organisms, do not offer an

adequate amount of variation to delineate variants. All of the missense analysis programs performed similarly, as determined by AUC (Table A.2, Figure A.2), this apparent discordance among severity calling is not unique to this study and has been observed by others.[11,16]

References

1. Antoniou, A. C. & Easton, D. F. Polygenic inheritance of breast cancer: Implications for design of association studies. *Genet. Epidemiol.* **25,** 190–202 (2003).

2. Tavtigian, S. V *et al.* Rare, evolutionarily unlikely missense substitutions in ATM confer increased risk of breast cancer. *Am. J. Hum. Genet.* **85,** 427–46 (2009).

3. Le Calvez-Kelm, F. *et al.* Rare, evolutionarily unlikely missense substitutions in CHEK2 contribute to breast cancer susceptibility: results from a breast cancer family registry case-control mutation-screening study. *Breast Cancer Res.* **13,** R6 (2011).

4. Damiola, F. *et al.* Rare key functional domain missense substitutions in MRE11A, RAD50, and NBN contribute to breast cancer susceptibility: results from a breast cancer family registry case-control mutation-screening study. *Breast Cancer Res.* **16,** R58 (2014).

5. Le Calvez-Kelm, F. *et al.* RAD51 and breast cancer susceptibility: no evidence for rare variant association in the breast cancer family registry study. *PLoS One* **7,** e52374 (2012).

6. Park, D. J. *et al.* Rare mutations in RINT1 predispose carriers to breast and Lynch syndrome-spectrum cancers. *Cancer Discov.* **4,** 804–15 (2014).

7. Park, D. J. *et al.* Rare mutations in XRCC2 increase the risk of breast cancer. *Am. J. Hum. Genet.* **90,** 734–9 (2012).

8. Tavtigian, S. V *et al.* Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J. Med. Genet.* **43,** 295–305 (2006).

9. Stone, E. a & Sidow, A. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Res.* **15,** 978–86 (2005).

10. Greenblatt, M. S. *et al.* Detailed computational study of p53 and p16: using evolutionary sequence analysis and disease-associated mutations to predict the functional consequences of allelic variants. *Oncogene* **22,** 1150–63 (2003).

11. Hicks, S., Wheeler, D. a., Plon, S. E. & Kimmel, M. Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. *Hum. Mutat.* **32,** 661–668 (2011).

12. Adebali, O., Reznik, A. O., Ory, D. S. & Zhulin, I. B. Establishing the precise evolutionary history of a gene improves prediction of disease-causing missense mutations. *Genet. Med.* **18,** 1029-36 (2016).

13. Armougom, F. *et al.* Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res.* **34,** W604–8 (2006).

14. Taly, J.-F. *et al.* Using the T-Coffee package to build multiple sequence alignments of protein, RNA, DNA sequences and 3D structures. *Nat. Protoc.* **6,** 1669–82 (2011).

15. Felsenstein, J. PHYLIP (Phylogeny Inference Package) version 3.6. *Distrib. by author. Dep. Genome Sci. Univ. Washington, Seattle* (2005).

16. Castellana, S. & Mazza, T. Congruency in the prediction of pathogenic missense mutations: state-of-the-art web-based tools. *Brief. Bioinform.* **14,** 448–59 (2013).

Table A.1. The average number of substitutions per position for each gene in the pathway-wide analysis for each depth of pMSA

Clade	Common Name	*ATM*	*BARD1*	*BRCA1*	*BRCA2*	*BRIP1*	*CHEK2*	*MRE11a*	*NBN*	*PALB2*	*RAD50*	*RAD51*	*RAD51B*	*RAD51C*	*RAD51D*	*RAD52*	*RBBP8*	*RINT1*	*TP53*	*XRCC2*	*XRCC3*	Across All
Metatheria[b]	Opossum/ Tasmanian Devil	0.67	1.39	2.08	2.48	1.50	1.21	0.70	1.41	2.27	0.31	0.76	1.29	0.93	1.86	1.90	1.52	0.45	1.67	0.89	1.70	1.46
Ornithorhynchus anatinus	Platypus	0.83	1.39[a]	2.08[a]	3.10	1.77	1.59	0.81	1.91	3.08	0.39	0.77	1.45	1.11	1.86[a]	2.08	2.01	0.57	2.00	1.20	2.22	1.76
Sauria	Chicken	1.11	1.97	3.13	4.24	2.29	1.87	1.12	2.62	4.17	0.61	0.80	2.04	1.34	2.23	2.59	2.46	0.73	2.72	1.43	2.71	2.38
Sauria	Lizard/Turtle	1.37	2.52	4.00	5.35	2.29[a]	2.05	1.30	3.19	5.26	0.74	0.84	2.37	1.62	2.23[a]	3.07	3.04	0.92	3.01	1.69	3.10	2.91
Xenopus	Clawed Frog	1.71	3.60	5.25	6.70	2.93	2.45	1.57	3.92	6.72	0.96	0.90	2.37[a]	1.98	2.66	3.58	3.73	1.24	3.70	2.02	3.56	3.65
Coelacanthimorpha	Coelacanth	2.00	4.19	6.31	7.93	3.60	2.89	1.89	4.35	8.20	1.12	0.93	2.67	2.29	2.94	4.24	4.53	1.47	4.48	2.33	3.88	4.33
Clupeocephala	Zebrafish/Fugu /Tetraodon	2.49	5.27	8.17	9.42	4.28	3.43	2.43	5.15	9.75	1.35	1.05	3.23	2.74	3.42	5.03	4.98	2.00	4.97	3.01	4.54	5.22
Branchiostoma floridae	Lancelet	3.34	6.34	8.45	9.42[a]	4.96	3.96	3.02	6.13	9.75[a]	1.82	1.23	3.69	3.34	3.65	7.18	4.98[a]	2.68	4.97[a]	3.73	4.54[a]	5.69
Nematostella vectensis	Purple Sea Urchin	4.01	7.44	10.58	11.36	5.38	4.80	3.42	6.96	11.55	2.22	1.41	4.28	3.82	4.36	8.80	4.98[a]	3.24	4.97[a]	4.54	5.33	6.76
Ecdysozoa	D. melanogaster										3.09	1.80										
Arabidopsis thaliana	Thale cress											3.06[c]										

[a]In instances in which a protein sequence could not be found, the number of substitutions/position from the previously identified organism was re-used; [b]The protein multiple sequence alignment includes Homo sapiens (Human), Macaca mulatta (Rhesus macaque), Callithrix jacchus (Marmoset), an organism from clade Murinae (Mouse or Rat), an organism from clade Laurasiatheria (Pig, Cow, Panda, or Dog), Loxodonta africana (African elephant), and Dasypus novemcinctus (Armadillo).[c]Because of the high level of conservation across species of RAD51, protein homologs were added for C. elegans, S. cerevisaiae, S. pombe and A. thaliana. Only the substitutions per position for Human through A. thaliana shown.

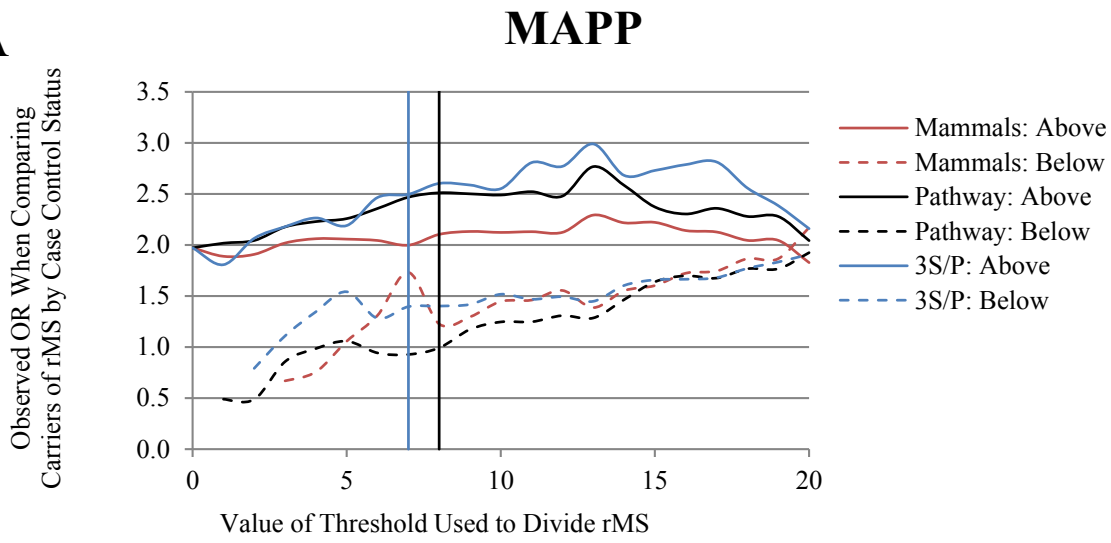
Table A.2. ROC curve estimates for each variant classifier

Variant Classifier: pMSA depth	Observation	ROC Area	Std. Err.	Asymptotic Normal 95% Confidence Interval	
Align-GVGD: M[a]	302	0.5746	0.0323	0.51135	0.63787
Align-GVGD: 3 S/P[b]	302	0.5829	0.0303	0.52355	0.64233
Align-GVGD: X[c]	302	0.5716	0.0324	0.50816	0.63512
MAPP: M[a]	302	0.5685	0.036	0.49799	0.63907
MAPP: 3 S/P[b]	302	0.5883	0.0356	0.51848	0.65809
MAPP: X[e]	302	0.5823	0.0362	0.51137	0.65329
CADD	302	0.5989	0.348	0.53073	0.66717
Polyphen2	302	0.5923	0.0337	0.52631	0.65825

[a]M=Mammals: Human through *Ornithorynchus anatinus* pMSA depth; [b]3 S/P=3 substitutions per position: Human through 3 substitutions per position pMSA depth, which was different for each gene: *ATM*, *Brachiostoma floridae*; *BARD1*, *Xenopus*; *CHEK2, Latimeria chalumnae; MRE11A, Branchiostoma floridae; NBN, Anolis carolinensis; RAD50, Drosophila melanogaster; RAD51, Arabidopsis thaliana; RINT1, Strongylocentrotus purpuratus; XRCC2, Danio rerio*; [c]X=*Xenopus*: Human through *Xenopus* pMSA depth. [d]NC=noncurated: Human through *Ornithorynchus anatinus*, without hand-curation pMSA depth; [e]X=*Xenopus*: Human through *Xenopus* pMSA depth, excluding *Anolis carolinensis* orthologs.

Figure A.1. Observed OR and thresholds for each missense substitution analysis program. The observed OR for the carriers of rMS of both the “above” and “below” groups for each severity threshold tested with A) MAPP, and B) Align-GVGD, adjusting for race/ethnicity and study center. Vertical lines are indicative of the threshold for which the observed OR≥2.5.

A


MAPP
Observed OR When Comparing Carriers of rMS by Case Control Status
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
5
10
15
20
Value of Threshold Used to Divide rMS
Mammals: Above
Mammals: Below
Pathway: Above
Pathway: Below
3S/P: Above
3S/P: Below

B

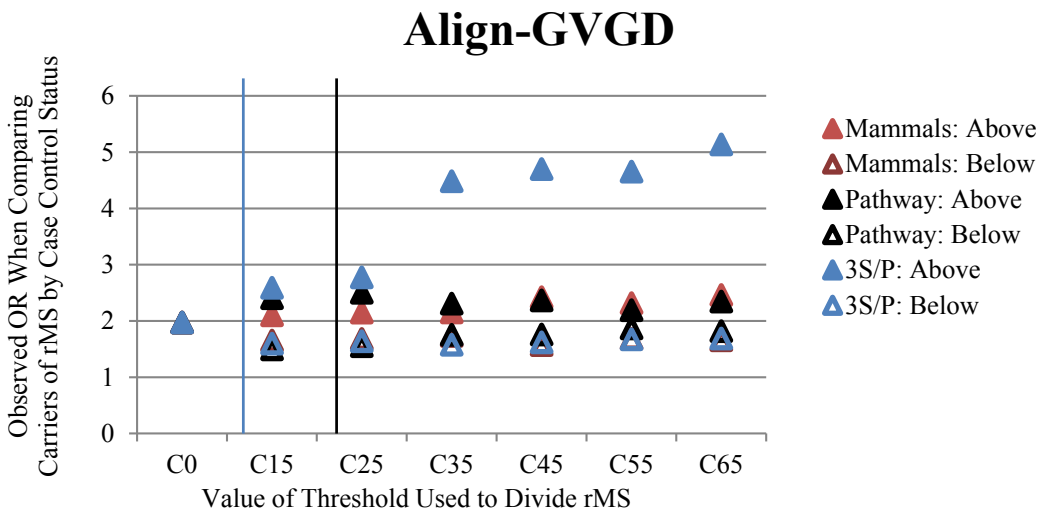
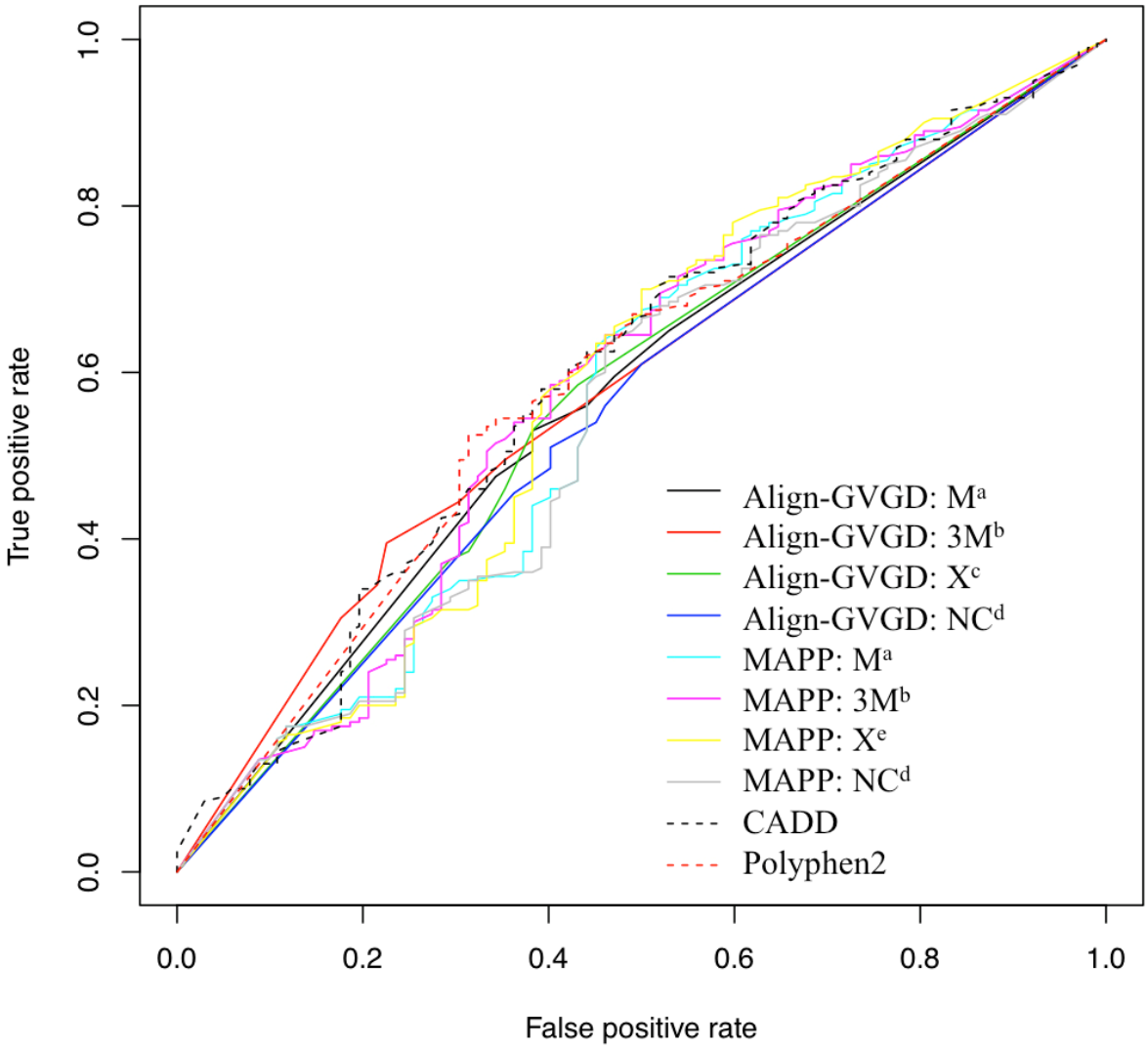

Align-GVGD
Observed OR When Comparing Carriers of rMS by Case Control Status
6
5
4
3
2
1
0
C0
C15
C25
C35
C45
C55
C65
Value of Threshold Used to Divide rMS
Mammals: Above
Mammals: Below
Pathway: Above
Pathway: Below
3S/P: Above
3S/P: Below

Figure A.2. ROC curves for each variant classifier. [a]M=Mammals: Human through *Ornithorynchus anatinus* pMSA depth; [b]3M=3 substitutions per position: Human through 3M pMSA depth, which was different for each gene: *ATM*, *Brachiostoma floridae*; *BARD1*, *Xenopus*; *CHEK2, Latimeria chalumnae; MRE11A, Branchiostoma floridae; NBN, Anolis carolinensis; RAD50, Drosophila melanogaster; RAD51, Arabidopsis thaliana; RINT1, Strongylocentrotus purpuratus; XRCC2, Danio rerio*; [c]X=*Xenopus*: Human through *Xenopus* pMSA depth. [d]NC=noncurated: Human through *Ornithorynchus anatinus*, without hand-curation pMSA depth; [e]X=*Xenopus*: Human through *Xenopus* pMSA depth, excluding *Anolis carolinensis* orthologs.


True positive rate
0.0
0.2
0.4
0.6
0.8
1.0
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
Align-GVGD: M[a]
Align-GVGD: 3M[b]
Align-GVGD: X[c]
Align-GVGD: NC[d]
MAPP: M[a]
MAPP: 3M[b]
MAPP: X[e]
MAPP: NC[d]
CADD
Polyphen2