The establishment of the Alabama Hereditary Cancer Cohort and genetic analysis

by

Madison Rose Bishop

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
December 15, 2018

Keywords: Hereditary cancer, genetics, underrepresented individuals,
targeted next-generation sequencing

Approved by

Dr. Nancy Merner, Co-chair, Assistant Professor of Pathobiology
Dr. Randall Clark, Co-chair, Gilliland Professor of Drug Discovery and Development
Dr. Raj Amin, Associate Professor of Drug Discovery and Development
Dr. Richard Curtis Bird, Professor of Molecular Biology and Cancer Genetics
Dr. Peter Panizzi, Associate Professor of Drug Discovery and Development
Dr. Miranda Reed, Associate Professor of Drug Discovery and Development
Dr. Laurie Stevison, Assistant Professor of Biological Sciences

Abstract

Inherited genetic risk factors in known breast cancer (BC) susceptibility genes increase an individual's overall lifetime risk of developing BC and are estimated to contribute towards approximately 35% of hereditary BC cases. Although next-generation sequencing (NGS) gene panels have been implemented into the clinic to improve BC survival for individuals carrying known genetic risk factors, there is a concern that clinically accredited NGS gene panels contain genes lacking clinical validity and utility. Furthermore, groups that are most susceptible to health and healthcare disparities, such as African Americans, have been underrepresented in medical research and less information is known about hereditary cancer genetics in such populations. Therefore, it is imperative to explore adaptive approaches to recruitment and enroll underrepresented individuals into research studies to assess the complete contribution of risk variants in known cancer susceptibility genes using research-based NGS gene panels. The research efforts described herein highlight the mechanisms used to establish the Alabama Hereditary Cancer Cohort (AHCC) and the subsequent genetic analysis using a research-based NGS gene panel, B.O.P. (**B**reast, **O**varian, and **P**rostate) to evaluate the genetic risk of BC, ovarian cancer (OvC) and/or prostate cancer (PC) in different ethnicities. Over a three-year period using two essential approaches, hospital and community-based recruitment (CBR), 242 individuals were enrolled into the AHCC. Although both recruitment mechanisms were instrumental, the unique trust building, educational, and traveling components of CBR dramatically facilitated the enrollment of African Americans resulting in large families for

genetic analyses. The first B.O.P screening involved 43 cancer-affected individuals from the AHCC. The purpose of this screening was to analytically assess the B.O.P. gene panel. Upon bioinformatics processing and variant filtering, called variants were validated using polymerase chain reactions (PCR) and Sanger sequencing. Subsequently, 61 of 74 variants were validated and classified as true positives (TPs). TPs had an average sequencing depth of 659X and alternate allele frequency of 51%, whereas the average false positive (FP) sequencing depth was 34X and alternate allele frequency was 33%. Although low sequencing depth was not always indicative of a FP, high sequencing depths (>100X) signified a TP. Overall, the initial B.O.P. screening enabled the establishment of criteria to alleviate future validation efforts and strongly supported the use of B.O.P. to further explain hereditary cancer susceptibility. Subsequent B.O.P. screening of 97 BC-affected individuals of African and European descent from the AHCC provided essential insight towards the variant contributions in clinically relevant cancer susceptibility genes and differences between ethnicities. Although there were no significant differences between the ethnicities regarding damaging variants, African American BC cases were more likely to have seemingly benign variants compared to European American BC cases, which could explain ethnic-specific risk. Continued B.O.P. screening will elucidate BC genetics that may explain current disparities and, ultimately, impact the clinical validity/utility of clinically-accredited NGS gene panels.

Acknowledgments

First, I want to thank my graduate co-advisors, Dr. Nancy Merner and Dr. Randall Clark, who have helped guide and inspire me throughout by graduate career. Additionally, I want to thank all faculty members of the Department of Drug Discovery and Development in the Harrison School of Pharmacy, and specifically my Graduate Advisory Committee for their help and guidance. I also want to thank the Department of Pathobiology in the College of Veterinary Medicine for their constant support and available resources as well as my laboratory team members for their assistance in these research efforts. Furthermore, these research efforts would not have been possible without the support of our study participants and community partners as well as the fruitful partnership with East Alabama Medical Center. The establishment of the AHCC could not have been accomplished without their involvement and contribution. I also want to acknowledge the Office of Information Technology at Auburn University for support and compute time on the Hopper High-Performance Computing Cluster. My research was also supported by the AURIC Graduate Fellowship Program, the AU Cellular and Molecular Biosciences (CMB) Peaks of Excellence Research Fellowship (NSF-EPS-1158862, grant G00006750). Finally, I would also like to thank my husband for his continuous support throughout my graduate career, as well as my parents for their tireless efforts to ensure that I receive a great educational foundation.

Table of Contents

List of Tables

List of Illustrations

List of Abbreviations

ABCCEDP: Alabama Breast and Cervical Cancer Early Detection Program
ACMG: American College of Medical Genetics
ADPH: Alabama Department of Public Health
AEM: Active enrollment month
AHCC: Alabama hereditary cancer cohort
Alt.: Alternate
AU: Auburn University
BAM: Binary SAM
BC: Breast cancer
B.O.P.: Breast, Ovarian and Prostate cancer gene panel
BWA: Burrows-wheeler Aligner
CBR: Community-based recruitment
CLIA: Clinical Laboratory Improvement Amendments
DHPLC: Denaturing High Performance Liquid Chromatography
EAMC: East Alabama Medical Center
EVS: Exome Variant Server
FDR: False discovery rate
FPs: False positives
FNs: False negatives
FTE: Full-time equivalent
GATK: Genome Analysis Tool Kit
GSL: Genome Services Laboratory
GWASs: Genome wide association studies
HBOC: Hereditary breast and ovarian cancer
HGVS: Human Genome Variant Society
HRM: High Resolution Melt
IRB: Institutional review board
Kb: Kilobases
LA: Linkage analysis
MAF: Minor allele frequency
NCCN: National Comprehensive Cancer Network
National Heart, Lung, and Blood Institute
NGS: Next-generation sequencing
OR: Odds ratio
OvC: Ovarian cancer
PC: Prostate cancer

PCR: Polymerase chain reactions
PI: Principal investigator
PRSs: Polygenic risk scores
QC: Quality control
RR: Relative Risk
SA: Segregation Analyses
SAM: Sequence Alignment/Mapping
SCSC: SISTAs Can Survive Coalition
SIFT: Sorting Intolerant From Tolerant
SNPs: Single nucleotide polymorphism:
TPs: True positives
TN: True negatives
VCF: Variant calling format
VUSs: Variants of unknown significance
UC: Uterine cancer
WES: Whole-exome sequencing
WGS: Whole-genome sequencing
W.H.I.P.: Working to Help Those In Pink
YBCSN: Young Breast Cancer Survivorship Network

Chapter 1: An overview of hereditary breast cancer genetics and gene panel screening.

This information was modified from a manuscript published in Human Mutation in 2016: Chandler MR, Bilgili EP, Merner ND. A review of exome sequencing efforts toward hereditary breast cancer gene discovery. Human Mutation; Accepted: May 18, 2016. PMID: 27226120.

Abstract:

Inherited genetic risk factors contribute towards BC onset. BC risk variants can be divided into three categories of penetrance (high, moderate, and low) that reflect the probability of developing the disease. Traditional BC susceptibility gene discovery approaches that searched for high and moderate risk variants in familial BC cases have had limited success; to date, these risk variants explain only ~30% of familial BC cases. The introduction of NGS technology has revolutionized human disease-associated gene discovery efforts as well as clinical genetic testing. Currently, gene panels that target BC susceptibility genes have been implemented in the clinic to determine an individual's risk of developing BC and, subsequently, develop disease management strategies specific to the genetic risk variant. Although the analytical validity for available clinical gene panels is regulated by the Clinical Laboratory Improvement Amendments, there is a concern that clinically accredited NGS gene panels contain genes lacking clinical validity and utility. Furthermore, groups that are most susceptible to health and healthcare disparities, such as African Americans, have been continuously underrepresented in medical research. As a result, less information is known about hereditary cancer genetics in such populations. The elucidation of the complete contribution in known BC susceptibility is critical for improved risk assessment and genetic counseling and to provide crucial insight towards disease mechanisms, which is necessary for more effective disease management strategies. Therefore, it is necessary to focus research

efforts on underserved and underrepresented cancer-affected individuals in known cancer susceptibility genes using a research-based NGS gene panel.

1.1 Introduction – BC genetic risk factors

An average United States woman has a 1 in 8 (12.5%) lifetime risk of developing BC. A range of risk factors contributes to the development of the disease;[1] some women inherit genetic risk factors that contribute towards disease onset.[2-4] Such risk variants are divided into broad categories of penetrance that indicate the probability of developing BC. There are three general categories that confer varying amounts of relative risk (RR). Although absolute estimates of risk are more beneficial for genetic counseling, genetic risk paired with additional risk factors (i.e., breast density, age of menstruation, and menopause) create difficulties in determining an absolute lifetime risk for individuals.[5] Therefore, the RR associated with a genetic variant(s) is generally reported. High penetrant variants are associated with a RR >4; moderate penetrant variants and low penetrant variants account for RRs between 2-4 and <1.3, respectively.[2,3,5] Overall, genes that harbor such risk variants are considered BC susceptibility genes, although many of the low-penetrant BC variants discovered by Genome Wide Analysis Studies (GWASs) are not located in genes but presumably act by affecting expression of relevant genes.

Moderate to high penetrant variants are generally rare, with a minor allele frequency (MAF) of <1%. To date, the majority of variants in these two classes have been identified within coding or splice junction sequences of BC susceptibility genes; most evidence for BC risk has been obtained for single base substitutions and small insertions or deletions that truncate the protein and are assumed to cause loss of function.[5] Missense variants, on the other hand, are typically classified as variants of unknown significance until functional studies, and/or conservation and segregation

analyses (SA) can supply further information.[5] To date, over 35 genes have been suggested to carry high and/or moderate BC risk variants;[3,6] however, only a minority (approximately 11) of those genes have variants with an established association that meet stringent statistical significance and burden testing requirements.[5] Altogether, high and moderate risk variants in known BC susceptibility genes are present in less than 30% of BC cases with a suggestive personal or family history.[3] The majority of those cases (over 80%, representing ~25% of all familial BCs) have a high penetrant variant, in one of a small number of BC susceptibility genes, associated with a specific hereditary cancer syndrome[3] These variants confer an increased risk of a number of different cancers, including BC, result in early onset cancers, and segregate in families in an autosomal dominant pattern of inheritance. Overall, the exact cancer pattern that is observed in a family depends on the mutated gene and, more specifically, the segregating risk variant,[2,5] the pedigree structure in terms of males and females, and chance. The most commonly mutated BC susceptibility genes, *BRCA1* [OMIM 113750][7] and *BRCA2* [OMIM 600185][8], which generally harbor rare high/moderate risk variants that convey a 55-85% and 35-60% lifetime risk of developing BC, respectively,[2,9-12] account for ~15% of familial BC cases.[3] The phenotype observed in *BRCA* mutation carriers is referred to as Hereditary Breast and Ovarian Cancer (HBOC) syndrome, which was first described in the early 1970s,[13-15] twenty years before it was genetically linked to the *BRCA1* locus.[16,17] Generally, BC and/or OvC is observed in multiple generations on the same side of the family; furthermore, common characteristics include one or more women diagnosed at age 45 or younger, women diagnosed with more than one primary BC or both BC and OvC, male BC, and/or an additional family history of certain other cancers, including PC, melanoma, and pancreatic cancer.[3] However, many BC cases with a family history of the disease or an early age of onset do not have a high or moderate risk variant in a known BC

susceptibility gene; in fact, the risk variants that contribute towards ~70% of familial BC cases remain unknown.

1.2 Identifying BC genetic risk factors

Two traditional BC gene discovery methods are genetic linkage analysis (LA) followed by positional cloning, and the candidate gene approach. LA uses disease families to identify high penetrant variants by genotyping markers spaced over the whole genome and predicting which marker is linked to the causative gene/mutation.[18] Generally, families are grouped together for analysis to attain the necessary power for discovery. *BRCA1* and *BRCA2* were localized to specific chromosomal regions using this approach,[16,19] but LA has had no success in identifying additional BC susceptibility genes. This is mainly thought to be due to the genetic heterogeneity of hereditary BC [20,21]. Candidate gene studies were used as an alternative approach to BC susceptibility gene discovery. Genes are selected based on function (mainly, involvement in DNA repair) and screened for mutations in cohorts of hereditary BC cases and controls. Using this candidate gene-based approach, a number of genes in the BRCA pathway have been suggested to be BC susceptibility genes that harbor moderate penetrant variants.[2] Overall, this is a limited approach since the selection of genes is biased towards previous knowledge of involvement in cancer pathways.

NGS technology has revolutionized human disease-associated gene discovery efforts.[22,23] Despite the limitation of short read lengths and accuracy, which requires optimizing alignment algorithms to properly align the reads to the reference genome,[24,25] it is relatively low in cost and involves easy sample preparation and fast sequencing.[26] This approach has been considered groundbreaking since sequencing only a select few individuals has been demonstrated to be successful in disease gene identification,[23,27] at least for rare Mendelian diseases. Whole-exome

sequencing (WES) is the most currently utilized technique to identify rare genetic variants associated with disease.[28,29] Compared to whole-genome sequencing (WGS), WES is substantially lower in cost (for instance, ~$900/sample compared to ~$3000/sample for an average of ~60X coverage) and offers a 'smaller' data set representing <2% of the genome comprised of annotated coding variants with functional predictions (through the use of many accurate *in silico* bioinformatics programs).[29] Despite overlooking the impact of non-coding variants on disease risk, WES is suggested to be a valuable approach for identifying rare disease-risk variants, since, in general, traditional disease gene discovery approaches ultimately focused on expressed (protein-coding) sequences and were successful. Furthermore, rare and coding variants that alter a protein's sequence commonly affect protein function and/or are deleterious; thus, WES provides a data set that is enriched for disease alleles.[28] Study design, filtering strategies, and segregation and validation analyses for the first 12 familial BC whole-exome sequencing efforts were reviewed and summarized in Chandler *et al.*[30] Overall, only a modest number of novel BC risk genes were identified. Two study designs were implemented in these studies, family-based and case studies, and, ultimately, 94% and 97% of the exome-sequenced families and cases, respectively, had no BC risk alleles reported through these efforts. However, the massive amount of BC WES data that has been generated through these studies can be re-analyzed to aid in future gene discoveries.

Despite extensive effort using a wide range of technology, additional BC susceptibility genes with rare and high/moderate penetrant variants have been difficult to identify. Thus, the genetic architecture of the predisposition to BC has been debated for years. Some researchers strongly believe that BC susceptibility in non-*BRCA1*/2 mutation carriers follows a polygenic risk model of inheritance for which a large number of common single nucleotide polymorphisms (SNPs; MAF >1%) contribute multiplicatively towards risk.[31] Hence, a number of GWAS that

aimed to associate common SNPs with BC under the "common-disease common-variant hypothesis" have been carried out; over 70 SNPs have been robustly associated with BC to date.[32-47] These SNPs are referred to as low penetrant variants since they confer such a small increased risk of BC individually. However, the combined effects of each of these low penetrant variants is believed to contribute to a greater BC risk; thus, recently, a large collaborative study analyzed over 30,000 BC cases and controls to generate polygenic risk scores (PRSs) that incorporated 77 SNPs in order to stratify women based on their life time risk.[48] Women in the highest 1% of the PRS had a three-fold increased risk of developing BC relative to the mean. Interestingly, individuals in the lowest 1% of the PRS had a stronger effect of family history as well as younger ages of onset (under 40 years of age), suggesting that rarer genetic variants are likely more important for such BC cases. Noteworthy, the MAF of the 77 SNPs ranged from ~0.1% to ~49.0% with odds ratios (ORs) from 0.86 to 1.36.[48,49] Thus, low penetrant variants are not necessarily common. Two (rs11571833 and rs17879961) of the 77 SNPs have MAFs of < 1%, according to Ensembl[50] and dbSNP;[51] interestingly, these two SNPs had the highest reported ORs of all 77 SNPs, 1.26 and 1.36 for all BCs,[48] and are coding variants in *BRCA2* [NM_000059.3:c.9976A>T; NP_000050:p.K3326*] and *CHEK2* [OMIM 604373; NM_007194.3:c.470T>C; NP_009125.1:p.I157T], respectively. Ultimately, it would be interesting to investigate how these two rare variants contributed to the PRSs since the analysis did not only involve common variants as the title of the manuscript indicates.[48] Furthermore, a more recent investigation into the association between *BRCA2* p.K3326* and BC has calculated ORs of 1.28 for all BCs and 1.5 for ER negative BC,[52] which is similar to that reported by Mavaddat *et al.*. Meeks *et al*. also reported an association between *BRCA2* p.K3326* and serous OvC (OR 1.46). Inherited BC genetic risk factors contribute towards familial and 'sporadic' BCs. Regarding the latter, in which case a

family history is generally not observed, certain combinations of low penetrant alleles (that are associated with a high PRS) have been shown to contribute towards disease onset.[48] However, it is important to identify additional rare BC genetic risk factors and to understand how variants collectively contribute towards BC risk.

1.3 Clinical- and research-based gene panel screening

Since the implementation of clinical gene testing in the early 1990s, these screenings have improved BC survival for individuals carrying genetic risk factors through increased clinical breast exams and mammograms as well as other risk-reducing strategies (i.e. prophylactic surgery).[53] Therefore, the National Comprehensive Cancer Network (NCCN) developed disease management guidelines for individuals with mutations in susceptibility genes (*ATM* [NM_000051], *BARD1* [NM_000465], *BRCA*1 [NM_007300], *BRCA2* [NM_000059], *CDH1* [NM_004360], *BRIP1* [NM_032043], *CHEK2* [NM_001005735], *NBN* [NM_002485], *PALB2* [NM_024675], *PTEN* [NM_000314], *RAD51C* [NM_058216], *RAD51D* [NM_001142571], *STK11* [NM_000455], *TP53* [NM_000546]) that have sufficient evidence to support an association with hereditary BC and/or OvC.[54] In the past, several mutation screening approaches, such as Sanger Sequencing,[55,56] DHPLC (Denaturing High Performance Liquid Chromatography),[57] and HRM (High Resolution Melting),[58] were used to identify mutations in known disease-associated genes. With the introduction of NGS technology,[26,59] gene panels can now be created for mutation screening. This is a very high-throughput screening method that involves designing probes that target the genes of interest, capturing the targeted genomic regions, and massively parallel sequencing the captured DNA.[5,26,59]

Currently, gene panels that target BC susceptibility genes, such as BROCA (University of Washington) and myRisk (Myriad Genetics), have been implemented into the clinic to determine an individual's risk of developing BC and, subsequently, develop disease management strategies specific to the genetic risk variant.[5,60,61] However, prior to a BC susceptibility gene panel being used in the clinic, the suitability of the panel must be determined.[5] Specifically, the analytical validity is assessed to determine the accuracy of variant detection for available clinical gene panels and is regulated by the Clinical Laboratory Improvement Amendments (CLIA).[5,62] Important measurements to assess the analytical validity of a gene panel include sensitivity, specificity and false discovery rates (FDRs). Despite such regulations for analytical validity, clinical gene panels have many challenges that have yet to be addressed. For example, thus far, neither clinical validity (the strength of the variant's association with increased disease risk) nor clinical utility (the impact on clinical management strategies) are delimited for gene panels through such an organization. Additionally, regardless of the American College of Medial Genetics and Genomics' and the Association for Molecular Pathology's recommendations for variant classification (pathogenic, likely pathogenic, variants of uncertain significance [VUSs], likely benign, or benign),[63] discrepancies in variant categorization are prevalent, and decisions on which variants to report back to patients remain convoluted. For instance, although *BRCA1* and *BRCA2* have pre-established, strong associations with a high-risk of BC as well as other cancers, not all protein-truncating variants confer the same RR despite being proximal to one another.[64,65] Additionally, since studies that do not identify an association fail to publish their findings due to a seemingly low-level impact, the majority of published reports are subject to publication bias, resulting in an inaccurate estimation of cancer risk.[5] Furthermore, in the past, the majority of association studies focused on overtly damaging variants with a presumed loss of function. As a result, a plethora of

missense variants are classified as VUSs and remain to be fully assessed for an association with BC along with their functional effects using pedigree data, tumor subtype analysis, and conservation algorithms.[5,66-68] Interestingly, these VUSs may only slightly elevate an individual's lifetime risk; however, specific combinations of VUSs may multiplicatively influence risk of developing BC and should not be disregarded.

Moreover, the clear majority of hereditary BC genetics research and, specifically, gene panel screenings, has been carried out on populations of European descent.[53] Therefore, the contribution of pathogenic mutations in known and strongly suggested genes remains relatively unknown in minority populations throughout the United States.[5,69] In fact, even in the era of genomic sequencing and precision medicine/health and despite federal mandates to include women and minorities in federally funded research,[70] ethnic minorities and underprivileged individuals remain omitted from biobanks (collections of catalogued biospecimens) and stemming genetics research studies.[71-74] If underrepresented and underprivileged individuals, such as African Americans, continue to be overlooked regarding genetics research, progress in precision medicine will be incomplete, and health disparities will be exacerbated. [71,73,75,76] Although, historically, these disparities have been mainly attributed to socioeconomic factors,[77] African American women are reported to have a higher incidence rate of BC under the age of 45 as well as a higher mortality rate at every age.[78] Furthermore, African American women are more likely to be diagnosed with triple-negative BC, a more aggressive BC sub-type with a poor short term prognosis.[78-80] Therefore, genetic risk factors are now strongly believed to contribute towards African American BC,[81,82] and it is imperative to explore adaptive approaches to recruitment and enroll underrepresented individuals into research studies and assess

the complete contribution of risk variants in know cancer susceptibility genes using research-based NGS gene panels.

1.4 Conclusions

The elucidation of hereditary BC genetics is critical for improved risk assessment and genetic counseling, and to provide crucial insight towards disease mechanisms, which is necessary for more effective disease management strategies. Therefore, it is necessary to focus research efforts on underserved and underrepresented cancer-affected individuals for all variant types in known cancer susceptibility genes using a research-based NGS gene panel. This dissertation highlights the two essential recruitment approaches, hospital and CBR, that were developed and implemented to establish the AHCC, and the subsequent genetic analysis using a research-based NGS gene panel, B.O.P. to evaluate the genetic risk of BC, OvC and/or PC in different ethnicities.

1.5 References

1 ACS. American Cancer Society: Breast Cancer Facts and Figures 2013-2014. (2014).

2 Apostolou, P. & Fostira, F. Hereditary breast cancer: the era of new susceptibility genes. *Biomed Res Int* **2013**, 747318, doi:10.1155/2013/747318 (2013).

3 Shiovitz, S. & Korde, L. A. Genetics of breast cancer: a topic in evolution. *Ann Oncol*, doi:10.1093/annonc/mdv022 (2015).

4 Berliner, J. L. & Fay, A. M. Risk assessment and genetic counseling for hereditary breast and ovarian cancer: recommendations of the National Society of Genetic Counselors. *J Genet Couns* **16**, 241-260, doi:10.1007/s10897-007-9090-7 (2007).

5 Easton, D. F. *et al.* Gene-panel sequencing and the prediction of breast-cancer risk. *N Engl J Med* **372**, 2243-2257, doi:10.1056/NEJMsr1501341 (2015).

6 OMIM®. Online Mendelian Inheritance in Man. (2015).

7 Miki, Y. *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**, 66-71 (1994).

8 Wooster, R. *et al.* Identification of the breast cancer susceptibility gene BRCA2. *Nature* **378**, 789-792, doi:10.1038/378789a0 (1995).

9 Lalloo, F. & Evans, D. G. Familial breast cancer. *Clin Genet* **82**, 105-114, doi:10.1111/j.1399-0004.2012.01859.x (2012).

10 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, doi:10.1086/375033 (2003).

11 King, M. C., Marks, J. H. & Mandell, J. B. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. *Science* **302**, 643-646, doi:10.1126/science.1088759 (2003).

12 Fackenthal, J. D. & Olopade, O. I. Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations. *Nat Rev Cancer* **7**, 937-948, doi:10.1038/nrc2054 (2007).

13 Lynch, H. T. & Krush, A. J. Carcinoma of the breast and ovary in three families. *Surg Gynecol Obstet* **133**, 644-648 (1971).

14 Lynch, H. T. *et al.* Tumor variation in families with breast cancer. *JAMA* **222**, 1631-1635 (1972).

15 Lynch, H. T. *et al.* Familial association of carcinoma of the breast and ovary. *Surg Gynecol Obstet* **138**, 717-724 (1974).

16 Hall, J. M. *et al.* Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* **250**, 1684-1689 (1990).

17 Narod, S. A. *et al.* Familial breast-ovarian cancer locus on chromosome 17q12-q23. *Lancet* **338**, 82-83 (1991).

18 Morton, N. E. Sequential tests for the detection of linkage. *Am J Hum Genet* **7**, 277-318 (1955).

19 Wooster, R. *et al.* Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science* **265**, 2088-2090 (1994).

20 Gonzalez-Neira, A. *et al.* Genomewide high-density SNP linkage analysis of non-BRCA1/2 breast cancer families identifies various candidate regions and has greater power than microsatellite studies. *BMC Genomics* **8**, 299, doi:10.1186/1471-2164-8-299 (2007).

21 Hilbers, F. S. *et al.* Exome sequencing of germline DNA from non-BRCA1/2 familial breast cancer cases selected on the basis of aCGH tumor profiling. *PLoS One* **8**, e55734, doi:10.1371/journal.pone.0055734 (2013).

22 Ng, S. B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272-276, doi:10.1038/nature08250 (2009).

23 Ng, S. B. *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* **42**, 30-35, doi:10.1038/ng.499 (2010).

24 Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* **18**, 1851-1858, doi:10.1101/gr.078212.108 (2008).

25 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

26 Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat Biotechnol* **26**, 1135-1145, doi:10.1038/nbt1486 (2008).

27 Ku, C. S., Naidoo, N. & Pawitan, Y. Revisiting Mendelian disorders through exome sequencing. *Hum Genet* **129**, 351-370, doi:10.1007/s00439-011-0964-2 (2011).

28 Bamshad, M. J. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* **12**, 745-755, doi:10.1038/nrg3031 (2011).

29 Kiezun, A. *et al.* Exome sequencing and the genetic basis of complex traits. *Nat Genet* **44**, 623-630, doi:10.1038/ng.2303 (2012).

30 Chandler, M. R., Bilgili, E. P. & **Merner, N. D.** A Review of Whole-Exome Sequencing Efforts Toward Hereditary Breast Cancer Susceptibility Gene Discovery. *Hum Mutat* **37**, 835-846, doi:10.1002/humu.23017 (2016).

31 Ponder, B. A., Antoniou, A., Dunning, A., Easton, D. F. & Pharoah, P. D. Polygenic inherited predisposition to breast cancer. *Cold Spring Harbor symposia on quantitative biology* **70**, 35-41, doi:10.1101/sqb.2005.70.029 (2005).

32 Michailidou, K. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* **45**, 353-361, 361e351-352, doi:10.1038/ng.2563 (2013).

33 Easton, D. F. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087-1093, doi:10.1038/nature05887 (2007).

34 Hunter, D. J. *et al.* A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* **39**, 870-874, doi:10.1038/ng2075 (2007).

35 Stacey, S. N. *et al.* Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* **39**, 865-869, doi:10.1038/ng2064 (2007).

36 Stacey, S. N. *et al.* Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* **40**, 703-706, doi:10.1038/ng.131 (2008).

37 Ahmed, S. *et al.* Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet* **41**, 585-590, doi:10.1038/ng.354 (2009).

38 Thomas, G. *et al.* A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nat Genet* **41**, 579-584, doi:10.1038/ng.353 (2009).

39 Zheng, W. *et al.* Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet* **41**, 324-328, doi:10.1038/ng.318 (2009).

40 Antoniou, A. C. *et al.* A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population. *Nat Genet* **42**, 885-892, doi:10.1038/ng.669 (2010).

41 Turnbull, C. *et al.* Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet* **42**, 504-507, doi:10.1038/ng.586 (2010).

42 Fletcher, O. *et al.* Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. *J Natl Cancer Inst* **103**, 425-435, doi:10.1093/jnci/djq563 (2011).

43 Haiman, C. A. *et al.* A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. *Nat Genet* **43**, 1210-1214, doi:10.1038/ng.985 (2011).

44 Ghoussaini, M. *et al.* Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nat Genet* **44**, 312-318, doi:10.1038/ng.1049 (2012).

45 Siddiq, A. *et al.* A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. *Hum Mol Genet* **21**, 5373-5384, doi:10.1093/hmg/dds381 (2012).

46 Garcia-Closas, M. *et al.* Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nat Genet* **45**, 392-398, 398e391-392, doi:10.1038/ng.2561 (2013).

47 Michailidou, K. *et al.* Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet* **47**, 373-380, doi:10.1038/ng.3242 (2015).

48 Mavaddat, N. *et al.* Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst* **107**, doi:10.1093/jnci/djv036 (2015).

49 Narod, S. A. Breast cancer prevention in the era of precision medicine. *J Natl Cancer Inst* **107**, doi:10.1093/jnci/djv078 (2015).

50 Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Res* **46**, D754-D761, doi:10.1093/nar/gkx1098 (2018).

51 Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**, 308-311 (2001).

52 Meeks, H. D. *et al.* BRCA2 Polymorphic Stop Codon K3326X and the Risk of Breast, Prostate, and Ovarian Cancers. *J Natl Cancer Inst* **108**, doi:10.1093/jnci/djv315 (2016).

53 Shah, P. D. & Nathanson, K. L. Application of Panel-Based Tests for Inherited Risk of Cancer. *Annu Rev Genomics Hum Genet* **18**, 201-227, doi:10.1146/annurev-genom-091416-035305 (2017).

54 National Comprehensive Cancer Network. *Genetic/Familial High-Risk Assessment: Breast and Ovarian (Version I.2019)*, <https://www.nccn.org/professionals/physician_gls/pdf/genetics_screening.pdf> (2019).

55 Sanger, F. & Coulson, A. R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* **94**, 441-448 (1975).

56 Huang, X. C., Quesada, M. A. & Mathies, R. A. DNA sequencing using capillary array electrophoresis. *Anal Chem* **64**, 2149-2154 (1992).

57 Fackenthal, D. L., Chen, P. X. & Das, S. Denaturing high-performance liquid chromatography for mutation detection and genotyping. *Methods Mol Biol* **311**, 73-96, doi:10.1385/1-59259-957-5:073 (2005).

58 Reed, G. H., Kent, J. O. & Wittwer, C. T. High-resolution DNA melting analysis for simple and efficient molecular diagnostics. *Pharmacogenomics* **8**, 597-608, doi:10.2217/14622416.8.6.597 (2007).

59 Shendure, J., Mitra, R. D., Varma, C. & Church, G. M. Advanced sequencing technologies: methods and goals. *Nat Rev Genet* **5**, 335-344, doi:10.1038/nrg1325 (2004).

60 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**, 12629-12633, doi:10.1073/pnas.1007983107 (2010).

61 Walsh, T. *et al.* Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. *Proc Natl Acad Sci U S A* **108**, 18032-18037, doi:10.1073/pnas.1115052108 (2011).

62 Centers for Disease Control and Prevention. *Clinical Laboratory Improvement Amendments (CLIA)*, <https://wwwn.cdc.gov/clia/> (2017).

63 Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405-424, doi:10.1038/gim.2015.30 (2015).

64 Thompson, D., Easton, D. & Breast Cancer Linkage, C. Variation in BRCA1 cancer risks by mutation position. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **11**, 329-336 (2002).

65 Thompson, D., Easton, D. & Breast Cancer Linkage, C. Variation in cancer risks, by mutation position, in BRCA2 mutation carriers. *American journal of human genetics* **68**, 410-419, doi:10.1086/318181 (2001).

66 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. *American journal of human genetics* **81**, 873-883, doi:10.1086/521032 (2007).

67 Goldgar, D. E. *et al.* Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. *American journal of human genetics* **75**, 535-544, doi:10.1086/424388 (2004).

68 Plon, S. E. *et al.* Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Human mutation* **29**, 1282-1291, doi:10.1002/humu.20880 (2008).

69 Bishop, M. R. *et al.* Establishment of the Alabama Hereditary Cancer Cohort - strategies for the inclusion of underrepresented populations in cancer genetics research. *Mol Genet Genomic Med*, doi:10.1002/mgg3.443 (2018).

70 NIH. Inclusion of Women and Minorities as Participants in Research Involving Human Subjects- Policy Implementation Page. (2015).

71 Spratt, D. E. *et al.* Racial/Ethnic Disparities in Genomic Sequencing. *JAMA Oncol* **2**, 1070-1074, doi:10.1001/jamaoncol.2016.1854 (2016).

72 Wong, S. L. Medically underserved populations: disparities in quality and outcomes. *J Oncol Pract* **11**, 193-194, doi:10.1200/JOP.2015.004259 (2015).

73 Bustamante, C. D., Burchard, E. G. & De la Vega, F. M. Genomics for the world. *Nature* **475**, 163-165, doi:10.1038/475163a (2011).

74 Ashley, E. A. Towards precision medicine. *Nat Rev Genet* **17**, 507-522, doi:10.1038/nrg.2016.86 (2016).

75 Konkel, L. Racial and Ethnic Disparities in Research Studies: The Challenge of Creating More Diverse Cohorts. *Environ Health Perspect* **123**, A297-302, doi:10.1289/ehp.123-A297 (2015).

76 Streicher, S. A. *et al.* Reasons for participating and genetic information needs among racially and ethnically diverse biobank participants: a focus group study. *Journal of community genetics* **2**, 153-163, doi:10.1007/s12687-011-0052-2 (2011).

77 Daly, B. & Olopade, O. I. Race, ethnicity, and the diagnosis of breast cancer. *Jama* **313**, 141-142, doi:10.1001/jama.2014.17323 (2015).

78 Breast Cancer Facts & Figures. *American Cancer Society* (2015).

79 Stark, A. *et al.* African ancestry and higher prevalence of triple-negative breast cancer: findings from an international study. *Cancer* **116**, 4926-4932, doi:10.1002/cncr.25276 (2010).

80 Dietze, E. C., Sistrunk, C., Miranda-Carboni, G., O'Regan, R. & Seewaldt, V. L. Triple-negative breast cancer in African-American women: disparities versus biology. *Nature reviews. Cancer* **15**, 248-254, doi:10.1038/nrc3896 (2015).

81 Bishop, M. R. *et al.* Establishment of the Alabama Hereditary Cancer Cohort - strategies for the inclusion of underrepresented populations in cancer genetics research. *Molecular Genetics & Genomic Medicine (Accepted with minor revisions. Revised manuscript (MGG3-2018-03-0049.R1) submitted on 06/06/18)*, doi:DOI:10.1002/mgg3.443 (2018).

82 Daly, B. & Olopade, O. I. A perfect storm: How tumor biology, genomics, and health care delivery patterns collide to create a racial survival disparity in breast cancer and proposed interventions for change. *CA Cancer J Clin* **65**, 221-238, doi:10.3322/caac.21271 (2015).

Chapter 2: Establishment of the Alabama Hereditary Cancer Cohort - strategies for the inclusion of underrepresented populations in cancer genetics research

This information was published in Molecular Genetics and Genomic Medicine in 2018: Bishop MR, Shah A, Shively M, Huskey ALW, Omeler SM, Bilgili EP, Jackson E, Daniell K, Stallworth E, Spina S, Shepp K, Bergstresser S, Davis A, Dean H, Gibson J, Johnson B, Merner ND. Establishment of the Alabama Hereditary Cancer Cohort - strategies for the inclusion of underrepresented populations in cancer genetics research. Molecular Genetics and Genomic Medicine; Accepted: July 1, 2018. PMID:29962060.

2.1 Abstract

Historically, groups that are most susceptible to health and healthcare disparities have been underrepresented in medical research. It is imperative to explore approaches that can facilitate the recruitment of underrepresented individuals into research studies. Two approaches, hospital and CBR, were developed and implemented over 36 months to study the genetics of hereditary BC and associated cancers in Alabama, a medically underserved state with double the national percentage of self-identifying African Americans, establishing the AHCC. Overall, 242 individuals enrolled. This included 84 cancer probands through hospital recruitment, as well as 76 probands and 82 family members through CBR. Eighty-one percent of the study participants' counties of residence are completely medically underserved. Furthermore, African Americans represent 26% of the hospital probands compared to 49% and 70% of the probands and family members who, respectively, enrolled through CBR. Although both recruitment mechanisms were instrumental, the unique trust building, educational, and traveling components of CBR facilitated the enrollment of African Americans resulting in large families for genetic analyses. The ultimate goal is to gain

insight from these rudimentary efforts in order to expand recruitment and accrue a unique resource for cancer genetics research.

Keywords: Recruitment; biobank; hereditary breast cancer; underrepresented individuals; African American

2.2 Background

Health and healthcare disparities have been an enduring and tenacious issue in the United States. Many groups are vulnerable to such disparities, including (but not limited to) ethnic minorities and individuals of lower socioeconomic status, as well as people in particular geographic locations. Since such groups are not always mutually exclusive, many subgroups of vulnerable populations exist.[72,83,84] Historically, groups that are most susceptible to health and healthcare disparities have been underrepresented in medical research.[75,85,86] Even today, in the era of genomic sequencing and precision medicine/health, this fact still holds true.[71,73,74] This is extremely unfortunate since ethnic diversity, socioeconomic status, and geography all play a role in disease susceptibility, progression, and outcomes.[71,72,75,87] If underrepresented individuals continue to be overlooked as research participants, progress in precision medicine/health will be limited, and health disparities will be exacerbated.[71,73,75,76]

Involving underrepresented individuals in research studies is not a simple task; in fact, the time and effort that must be invested for success can be greatly underappreciated.[85,86,88] It requires overcoming barriers and addressing informational, logistical, sociocultural, and attitudinal factors that could otherwise negatively influence research participation.[85,86] There are examples of the successful recruitment of individuals in rural areas[89] and collaborative efforts for the study of minority groups,[90] but these efforts are generally the exceptions. It is imperative that researchers continue to explore methods that will help facilitate the recruitment of underrepresented

individuals since the implementation of well-executed and appropriate recruitment efforts is key to true inclusion.[85]

Herein, we describe the recruitment approaches and biobanking for the AHCC to facilitate the genetic analyses of hereditary BC and associated cancers such as OvC and PC[30] in Alabama. Over 60% of the Alabama population is medically underserved; this includes the entire population of 85% of its counties, most of which are rural (Figure 2.1).[91,92] Furthermore, the percentage of the Alabama population who self-describe as being black or African American is nearly double that of the national population (26.8% versus 13.6%, respectively) with a predominantly African American population residing within the Alabama Black Belt region, an area associated with low economic status that encompasses 25% of the state's counties (Figure 2.1)[93,94] Thus, our recruitment mechanisms, which include standard hospital recruitment along with strategic and adaptive CBR, target underprivileged and minority groups in Alabama and aim to create a unique cohort of underrepresented individuals to study cancer genetics and disparities.

2.3 Methods

2.3.1 Ethical Compliance

Two different Auburn University (AU; Auburn, Alabama; Lee County; Figure 2.1) Institutional Review Board (IRB) approved protocols, hospital recruitment (AU IRB #14-232; approved January 2015) and CBR (AU IRB #15-111; approved March 2015), were designed to recruit and enroll BC, OvC, and PC affected individuals/families for the establishment of the AHCC.

Recruitment criteria and timeline: Study criteria include individuals: (i) diagnosed with BC, OvC, or PC (at any age) who have a family history of cancer, or (ii) diagnosed with BC, OvC, or PC under the age of 45 years without a family history of cancer. Furthermore, both cancer-

affected and unaffected family members of each study participant can join the study. The first cancer-affected individual to enroll in the study from each family is defined as the family proband. Recruitment and enrollment efforts were carried out over a 36-month period from January 2015 to December 2017.

2.3.2 Hospital recruitment

A general AU hospital recruitment protocol was initiated based on a partnership with the Cancer Center of East Alabama Medical Center (EAMC) and the approval of EAMC IRB protocol, 14-03-E. EAMC is located in Opelika, Alabama (Lee County) and serves six Alabama counties (Lee, Chambers, Tallapoosa, Macon, Russell, and Randolph counties; Figure 2.1). Overall, the hospital recruitment protocol was designed to allow the recruitment and enrollment of patients who fit the study criteria at current and future collaborating hospitals. The recruitment effort involves a designated hospital staff member, typically, a project-assigned research nurse, screening patients for eligibility. At EAMC, a part-time research nurse carried out this effort by screening the medical records of individuals on the cancer center's weekly schedule. Upon identification of potential study participants, the research nurse would contact those individuals to inform them of the study and schedule an enrollment appointment at the hospital, if interested. Upon consent, hospital medical records are accessed for information pertaining to the cancer diagnoses; furthermore, demographic information is recorded along with the participant's personal and family history of cancer and other cancer risk factors (i.e. number of children, breastfeeding habits, etc.). A pedigree is drawn to detail this information. Moreover, each study participant agrees to future contact for additional sampling and/or information regarding cancer risk and updates. Additionally, a blood sample is provided for DNA extraction and subsequent genetic analysis (AU IRB #14-335). The collected information is subject to the confidentiality and privacy regulations

of the recruiting hospital. The hospital removes each participant's name and assigns a specific alpha-numeric code to the collected blood sample and corresponding paper work/information that is transferred to AU.

2.3.3 CBR

The CBR protocol was established to engage individuals all over the state of Alabama and to inspire underrepresented individuals to participate in the study through an educational and trust building process. Community partners (Supplementary Table 2.1) foster this effort by introducing the CBR team to potential participants at different events throughout the state. Recruitment efforts included scheduling education sessions to cancer support groups, attending Relay For Life events in different Alabama counties, and participating in community partner BC-specific events (i.e. walks and/or workshops). IRB-approved flyers/brochures were disseminated at all recruiting events. These strategies ultimately identified individuals interested in study participation; subsequently, a CBR team member scheduled enrollment appointments for those who met the criteria and expressed interest in the study. Enrollment appointments were scheduled at the convenience of the study participant. In order to address transportation and other barriers limiting research participation, the CBR team traveled to the study participants for their enrollment appointments (Figure 2.1). Since April 2017, the Gene Machine has been used for CBR travel, which is a refurbished bus that serves as study advertisement and a mobile recruitment and enrollment station (Supplementary Figure 2.1).

Upon study consent at a CBR enrollment session, similar to hospital recruitment, an individual shares demographic information, her/his personal and family history of cancer, and other cancer risk factors. From this information, a family pedigree is generated. Medical information about a participant's cancer diagnosis is also shared but, in this setting, medical reports

are provided through the participant. CBR study participants also consent to a blood draw for DNA extraction (AU IRB #14-335), which is carried out by a trained CBR team member. In circumstances when blood samples are not attainable/practical (i.e. individuals who had double mastectomies and lymph node removal), saliva samples can be provided. The CBR-collected samples are assigned an alpha-numeric code for laboratory use to protect participant confidentiality. Lastly, study participants agree to be contacted in the future for additional sampling and/or information pertaining to cancer risk, updates, and potential family member involvement. Upon enrollment, it is the job of the study proband to reach out to family members to inform them of the study, gauge interest, and inquire about study involvement. Once interest is expressed and permission granted, the CBR team can contact family members for an enrollment appointment.

2.3.4 DNA bank and database

The Merner DNA bank and database protocol (AU IRB #14-335) was established to organize the storage and use of collected information and samples. After samples are collected at enrollment sessions of protocols #14-232 or #15-111, they are transported to the AU laboratory for DNA extraction. Blood DNA is extracted following a protocol published by Miller *et al.*[95] The participant's DNA is then stored at 4°C in the DNA bank; the exact location of the participant's DNA is recorded in the database. The database only contains de-identified information including the alpha-numeric sample code along with demographic and medical information that corresponds to each sample/study participant. Furthermore, the database describes how each sample can be used in research.

2.4 Results

Upon 36-months of recruitment and enrollment, the AHCC has 242 individuals from 160 cancer-affected families (Supplementary Figure 2.2 and Figure 2.2). This includes 160 cancer probands and 82 cancer-affected and unaffected family members from 27 different counties in Alabama (Supplementary Figure 2.2 and Figure 2.1 and 2.2,); 81% of the study participants' counties of residence (22 of 27 counties) are completely medically underserved (Figure 2.1). Of all the cancer probands, 52% (n=84) were recruited through hospital recruitment (Supplementary Figure 2.2 and Figure 2.2A) and 48% (n=76) through CBR (Supplementary Figure 2.2 and Figure 2.2B and 2.2C). All family members were recruited through CBR (Supplementary Figure 2.2 and Figure 2.2C). Overall, 62% (n=99), 37% (n=59) and 1% (n=2) of the probands self-reported being of European, African, and Asian descent, respectively. Ninety percent (n=144) of the probands are BC cases (Supplementary Figure 2.2A and Table 2.1).

2.4.1 Recruitment and Enrollment

Although the recruitment efforts and enrollment sessions were carried out over a span of 36 months, all months did not receive equal efforts for recruiting and enrolling individuals into the study (Figure 2.2). Months in which resources (i.e. time and personnel) were allotted toward enrollment sessions are defined as Active Enrollment Months (AEMs), whereas no efforts were made in Inactive Enrollment Months. There were a total of 17 and 14 inactive enrollment months for hospital recruitment and CBR, respectively (Figure 2.2A and 2.2C).

2.4.1.1 Hospital recruitment

After the screening process, the research nurse contacted eligible individuals and approximately equal percentages accepted and declined participation. Accordingly, this 50% hospital participation rate resulted in the enrollment of 84 probands; 73% (n=61), 26% (n=22), and

1% (n=1) are European, African, and Asian American, respectively (Supplementary Figure 2.2A and Table 2.1). Of the 19 total AEMs, seven, four, and eight fell in 2015, 2016, and 2017, respectively, with an overall average enrollment of four individuals per AEM (Figure 2.2A). The majority of study participants enrolled in 2017 (n=44; Figure 2.2A) with an average enrollment of six per AEM. The least successful enrollment year was 2016 with only 13 new study participants, averaging three per AEM (Figure 2.2A). Ninety-four percent (n=79) of the probands are BC cases with 49.8 years being the average age of onset (Table 2.1). Of the 79 BC probands, 72% (n=57) are European American, 27% (n=21) are African American, and 1% (n=1) is Asian American (Supplementary Figure 2.2A). One of the BC probands is an African American male. He was diagnosed at 42 years of age with moderately differentiated, infiltrating ductal carcinoma and has a family history of the disease. OvC and PC cases represent 4% (n=3) and 2% (n=2) of the probands, respectively (Supplementary Figure 2.2A and Table 2.1).

2.4.1.2 CBR

Recruitment efforts involved presenting 12 education seminars to cancer support groups and attending 13 Relay for Life events as well as 15 other BC-specific events (Figure 2.2B). The latter of which typically occurred in October, BC awareness month, and primarily involved attending the same community partner-organized events each year. Most of the recruitment efforts occurred in 2015 and 2017 (Figure 2.2B). Overall, attending CBR recruiting events highly corresponded with AEMs (Figure 2.2B and 2.2C), except in October of 2016 when limited resources were allotted to study enrollment. In fact, overall, the least amount of recruitment and enrollment efforts were allotted for 2016 (Figure 2.2B and 2.2C). Of the 22 CBR AEMs, eight, four, and 10 fell in 2015, 2016, and 2017, respectively. With a total of 158 study participants who enrolled through CBR, the overall average enrollment rate was seven individuals per AEM

(Figure 2.2C). The majority of the CBR study participants enrolled in 2015 (n=86) averaging 11 per AEM. Both 2016 (n=21) and 2017 (n=51) averaged five new enrollees per AEM (Figure 2.2C).

A total of 76 CBR probands enrolled in the study of which 50% (n=38), 49% (n=37), and 1% (n=1) are European, African, and Asian American, respectively (Supplementary Figure 2.2A and Figure 2.2C). The majority of the probands enrolled in 2015 (n=36) and 2017 (n=34), averaging four and three new probands per AEM for each respective year. Only six probands enrolled in 2016 (Figure 2.2C). Overall, 20% were initially identified at an education session, 22% through a Relay for Life, and 22% at a BC event; the remaining 36% were informed of the study through word of mouth or general publicity (i.e. a newspaper article). However, this differed based on ethnicity. More African American probands enrolled in the study after attending an education session or meeting the CBR team at a Relay for Life event (30% and 27%, respectively) compared to European American probands (11% and 16%, respectively). Moreover, word of mouth/general publicity contributed to only 22% of the African American probands but 50% of the European American probands. Eighty-six percent (n=65) of the CBR probands are BC cases with 46.1 years being the average age of onset. Of the CBR BC probands, 49% (n=32) are European American, 49% (n=32) are African American, and 2% (n=1) are Asian American (Supplementary Figure 2.2A and Table 2.1). Through CBR, one European American male diagnosed with multifocal intraductal papilloma at 48 years of age enrolled in the study. In addition to BC, OvC, and PC cancer-affected individuals (Supplementary Figure 2.2A and Table 2.1), five unique cancer cases/probands were also enrolled through CBR when the individual had an apparent family history of BC, OvC, or PC (Supplementary Figure 2.2A). This includes three (one European American and two African American) females diagnosed with uterine cancer (UC) as well as one European American female diagnosed with colorectal cancer. The unique

cancer cases also included an European American male who was diagnosed with squamous cell skin cancer at 65, melanoma at 70, and pancreatic cancer at 72 years of age (Supplementary Figure 2.2A).

A total of 82 family members enrolled in the study through CBR (Supplementary Figure 2.2), of which 50, 15, and 17 family members enrolled in 2015, 2016, and 2017, respectively (Figure 2.2C); family member enrollment rates were six, four, and two per AEM for each respective year. Despite that some family members were recruited along with their proband at the same recruiting event, family member recruitment was highly dependent on study participants reaching out and informing additional family members about the study. Of the 82 family members, 27 were cancer-affected and 55 unaffected individuals (Supplementary Figure 2.2B). The majority (70%; n=57) were African American, of which 95% (n=54) were family members of BC probands. Overall, a total of 12 African American BC families with multiple cancer-affected study participants have enrolled in the study (Supplementary Figure 2.2B and Figure 2.3); the largest families are 1CAD and 1CAG - each with six and five cancer-affected study participants, respectively. Family 1CAD also has 10 cancer-unaffected study participants, making it the largest enrolled family (Figure 2.3). European Americans represent 28% (n=23) of the enrolled family members (Supplementary Figure 2.2B). The majority (91%; n=21) of the European American family members were of BC probands, composing a total of 15 families, four of which have multiple cancer-affected individuals (Supplementary Figure 2.2B).

2.5 Discussion

With the extreme need to include underrepresented populations in medical research,[71,73,75,76] the establishment of the AHCC is a timely and vital effort from which to gain insight. In order to focus on individuals with a predisposition to hereditary BC, recruitment criteria

were established to identify affected individuals with hallmark characteristics such as a family history of BC and associated cancers and early ages of onset.[2,30] A number of hereditary cancer syndromes exist for which BC is an associated cancer. HBOC Syndrome[13-17] is one such syndrome that is characterized by BC and/or OvC in multiple generations, as well as diagnoses under 45 years of age, women with multiple primary BCs or both BC and OvC, male BC, and/or a family history of certain other cancers, including PC, melanoma, and pancreatic cancer.[3] This recruitment effort targeted probands who were primarily diagnosed with BC, OvC, and PC, three associated cancers of HBOC syndrome. Furthermore, unique cancer cases with a family history of such cancers have been recruited. This strategy was developed in recognition that BC is typically not the only cancer noted on a hereditary BC pedigree. Thus, the recruitment criteria allow for the inclusion of individuals/families who may have a genetic predisposition but could have been excluded from a study solely enrolling BC probands. For instance, it allows alternate cancer probands to enroll into the study who are from families that have experienced BC mortalities or have family members diagnosed with BC but unwilling to participate. It also recognizes families with a higher proportion of males to females that are more likely to observe PC over BC. Ultimately, with the main goal of identifying BC genetic risk factors, BC cases represented the majority (90%) of the probands recruited into the study. Nonetheless, in order to expand OvC and PC proband recruitment, additional effort needs to be made, such as committing more time identifying individuals diagnosed with such cancer types in both the hospital and community settings. Regarding the latter, partnering with OvC and PC support groups will be key since, to date, our community partners are primarily BC support groups. Additionally, it is important to recognize other hereditary cancer syndromes, including Cowden, Li-Fraumeni, Hereditary Diffuse Gastric Cancer, and Peutz-Jeghers Syndromes that are all associated with different types of inherited cancers in addition to BC.[3] Overall, the particular gene/mutation involved in

pathogenesis dictates the predisposition to particular cancer types and the cancer patterns observed in a family. Ultimately, in order to encompass all possible inherited BC syndromes and inheritance patterns, it is important to keep the definition of family history broad, asking study participants to acknowledge all cancers that they are aware of in their family, and recognize and enroll unique cancer probands who have a family history of BC.

In order to offer BC genetic research participation to individuals in the medically underserved state of Alabama who would not normally be given the opportunity to participate in such a research study, both hospital recruitment and CBR mechanisms were established. Both recruitment mechanisms were instrumental in enrolling individuals into the study. Together, they led to the enrollment of 242 study participants who provided information and samples that have been incorporated into the Merner DNA bank and database. This includes 160 cancer probands (90% of which were BC probands) and 82 family members. Hospital recruitment is the most traditional mode of recruitment for a genetic research study.[96] Our seminal hospital recruitment efforts involved identifying and enrolling patients at EAMC, a regional hospital in Lee County that serves 6 medically underserved Alabama counties. Despite that intermittent enrollment periods due to unforeseen circumstances at EAMC resulted in inactive enrollment months, one research nurse devoted approximately 0.25 full-time equivalent (FTE) towards this project during AEMs. Thus, during times of active enrollment, the typical 10 hours of weekly effort towards the project was divided into approximately seven hours of eligibility screening and contacting patients, and approximately three hours of enrollment appointments and paper work/data entry. Overall, a much lower participation rate (~50%) was observed compared to reports from other hospitals that have enrolled for genetic studies (with claims as high as 100%); however, it is important to note that participation rates are known to vary between hospitals, and when Helgesson *et al*. compared factors that could influence such participation rate differences, the actual site of recruitment was

determined to be the most important factor.[97] Despite that the site-specific study coordinator's motivation, demeanor, knowledge, and ability to communicate and build trust can influence such rates,[97] it is important to recognize that the EAMC service area is medically underserved and individuals in the area have very rarely been offered to participate in a research study. Plus, many individuals have been negatively and justifiably influenced by historical events that cast doubt on even the most well-intended efforts,[98] which is likely another contributing factor. An investigation into the exact factors that influenced the initial EAMC participation rate and ways to improve is pertinent. Overall, upon being offered study participation, 84 probands enrolled into the study at EAMC during 19 AEMs. Therefore, an overall average of four EAMC study participants enrolled per AEM, which ranged from an average of three to six individuals per AEM for each year of the study. Stemming back to the potential impacts of a study coordinator, the different yearly averages ultimately correlated with the assigned research nurse. On another note, the percentages of European American (73%), African American (26%), and Asian American (1%) probands that enrolled into the study at EAMC closely represented the racial demographics of the cancer center's patient population, being 65% European American, 33% African American, and 2% other (averaged over the three years of the study). Interestingly, this is contrary to other clinic-based studies that typically have a difficult time enrolling ethnic minorities.[96,97]

CBR has been suggested to be an effective method to recruit medically underserved and underrepresented racial/ethnic minorities into research studies;[99] thus, we designed a CBR mechanism to overcome barriers known to hinder research participation.[85,86] In order to reach out to individuals all over the state, an educational and trust-building recruitment process that involves traveling to different Alabama counties/communities was established. Specifically, four unique modes of recruitment were developed: offering education seminars to cancer support groups, attending Relay for Life events, participating in BC-specific events, and word-of-mouth/general

publicity. Presenting education seminars and attending both Relay for Life and BC-specific events were all essential to the success of this project, since each mode yielded similar enrollment of CBR probands. Although word-of-mouth/general publicity led to the enrollment of the largest portion of the CBR probands (36%) compared to the other three modes of recruitment individually, this mode led to the enrollment of a smaller portion of African American probands (22%) compared to European American probands (50%). The CBR team, which is mainly composed of European Americans, likely influenced the discrepancy in ethnicities recruited through word-of-mouth/general publicity since the ethnicity of the recruitment team has been reported to greatly influence participation of underrepresented individuals in medical research studies [85]. Thus, diversifying the CBR team to adequately represent the targeted population will likely help. However, interestingly, CBR ultimately enrolled equal numbers of European and African American BC probands; thus, the trust building and educational components of the other modes of recruitment highly influenced African American enrollment. In fact, most African American probands were recruited after attending an education session or meeting a CBR team member at their local Relay for Life. This was accomplished by targeting African American BC support groups for education sessions and choosing to attend Relay for life events in predominantly African American communities; hence, why the proband ethnic proportions do not match the state racial demographic.[100]

The initial CBR efforts were carried out on an extremely small-scale. In 2015, the principal investigator (PI) and a graduate student carried out CBR. Our first community partner, SISTAs Can Survive Coalition (SCSC), fostered the invitations to our first education seminars. Furthermore, attending Relay for Life events not only identified some of our initial CBR study participants but also facilitated additional partnerships, which subsequently resulted in more invitations to education sessions and BC-specific events. In 2015, despite not measuring the exact

participation rate, the recruitment and subsequent enrollment of 86 study participants through eight AEMs (averaging 11 individuals per AEM), as well as all other related tasks (such as relationship building, modifying protocols/dissemination materials, traveling, DNA extractions, database management, etc.), consumed the majority of the PI's (70%) and graduate student's (30%) workload. However, the efforts put forth in that inaugural year were necessary to demonstrate proof of concept. In 2016, with the newly established cohort, the PI's focus changed to seeking research funds and initiating genetic analyses hence the observance of so many inactive enrollment months. In 2017, funds were obtained to hire a recruitment coordinator who worked 0.67 FTE and allotted approximately 20% effort each towards recruiting, enrolling, traveling (on the Gene Machine), extracting DNA, and managing the DNA bank and database. Upon training, the recruitment coordinator independently enrolled individuals into the study from May to December (averaging six study participants per AEM through that period). Ultimately, over a total of 22 CBR AEMs, 158 individuals enrolled in the study for an overall average enrollment of seven individuals per AEM. Moving forward, the goal is to expand the CBR team and designate duties to make CBR as efficient as possible.

Similar to the initial efforts of the Carolina BC Study,[89] our CBR efforts involve traveling to individuals for enrollment appointments. We travel all over the state to enroll eligible individuals in order to overcome logistic barriers to research participation; currently, we have enrolled individuals from 27 counties. In addition to proband enrollment, this component of CBR has also proven to be an excellent approach to enroll large families for genetic analyses. Similar to the approach used by the Family Information Service described by Dr. Henry Lynch in 2001,[101] the CBR team coordinates and attends large family gatherings to recruit and enroll a large number of family members in a single session. For example, the CBR team receives invitations to family reunions, which are phenomenal events to provide an education seminar and enroll individuals as

family members reunite. However, if a single session is not ideal for family members due to barriers in transportation and/or conflicting time-commitments, the CBR team also travels to different towns to recruit individuals from the same family. For instance, the team traveled to towns in three different counties to recruit members of family 1CAG in Figure 2.3. Overall, African Americans represented the majority (70%) of family members that enrolled in the study; thus, again, reiterating that CBR is a great mechanism to involve African Americans in genetic research and provides a collection of unique families/individuals for analyses. Furthermore, since most of Alabama is rural, travel includes visiting isolated communities that are likely enriched for ancestral genetic mutations. Thus, by traveling to these communities for recruitment, the CBR team can cater to underserved populations as well as harness their genetic potential and detect ancestral mutations in seemingly unrelated probands/families. After all, studying cohorts derived from isolated populations is currently an extremely palatable approach towards BC susceptibility gene discovery.[30]

The recruitment mechanisms and stemming biobank also allows the investigation of particular cancer disparities. Firstly, African American BC genetics is vastly understudied and less understood compared to European American BC genetics.[102] Studying African American hereditary BC is a priority of this study since African American women are more often diagnosed with an aggressive and less treatable BC sub-type and have a higher incidence rate of BC under the age of 40 compared to European Americans.[1] Similarly, African American males are more susceptible to PC compared to European Americans, and normally diagnosed at a younger age and with larger tumors.[103] Thus, considering that (i) an early age of onset is a hallmark of hereditary cancer, (ii) hereditary BC is associated with an increased risk of PC,[4] and (iii) the Black Women's Health Study has demonstrated there is a strong familial component of African American BC,[104] it is likely that genetic risk factors contribute towards the higher incidence rate of early onset and

aggressive BC and PC in African Americans and that the two disparities are genetically-linked. This potential link will be investigated, especially with the success of CBR regarding African American enrollment. Additionally, the families that have been recruited, such as our largest African American family, 1CAD, are excellent examples of both BC and PC segregating in the same family, highlighting our resources and the practicality to investigate the genetic overlap.

2.6 Conclusions

This initial report details the protocols that were established and carried out to enroll underrepresented individuals into a hereditary BC cancer genetic study and the subsequent development of a biobank from which samples can be used in future independent and collaborative cancer genetic studies. It specifically highlights the rudimentary accomplishments made during the first three years of the project and provides insight on how to continue and expand the efforts. A hospital recruitment protocol was established for its efficiency. It is the most standard mode of recruitment due to the ease of identifying study participants, obtaining complete medical records, and carrying out enrollment appointments. Therefore, in order to expand this efficient mode of recruitment, the protocol was strategically designed to add collaborating hospitals through IRB reliance agreements. However, it is important to note that site-specific enrollment rates will vary greatly depending on the percent FTE allocated to the project as well as each study coordinator's personality. Furthermore, due to Alabama being a significantly medically underserved state with double the national percentage of self-identifying African Americans, it was crucial to adapt and develop an alterative recruitment method. CBR focused on overcoming recruitment barriers, enabling our team to connect with even more underrepresented individuals in the state. We aspire to grow similarly to the Carolina BC Study,[89] which has continued to function for over 20 years and now has a large staff of interviewers, nurses, and technicians committed to the project.

Currently, we plan to continue to work closely with our partners and stay connected with the community as we traveling to events, education seminars, and enrollment appointments on the Gene Machine, which now has a strong presence on social media and has begun to unofficially brand our CBR efforts providing a new marketing component and mode of recruitment. Overall, we have learned that the effort required to include underrepresented individuals in research is immense and challenging. It is a vital effort that should no longer be underappreciated.

2.7 References

1 ACS. American Cancer Society: Breast Cancer Facts and Figures 2013-2014. (2014).

2 Apostolou, P. & Fostira, F. Hereditary breast cancer: the era of new susceptibility genes. *Biomed Res Int* **2013**, 747318, doi:10.1155/2013/747318 (2013).

3 Shiovitz, S. & Korde, L. A. Genetics of breast cancer: a topic in evolution. *Ann Oncol*, doi:10.1093/annonc/mdv022 (2015).

4 Berliner, J. L. & Fay, A. M. Risk assessment and genetic counseling for hereditary breast and ovarian cancer: recommendations of the National Society of Genetic Counselors. *J Genet Couns* **16**, 241-260, doi:10.1007/s10897-007-9090-7 (2007).

13 Lynch, H. T. & Krush, A. J. Carcinoma of the breast and ovary in three families. *Surg Gynecol Obstet* **133**, 644-648 (1971).

14 Lynch, H. T. *et al.* Tumor variation in families with breast cancer. *JAMA* **222**, 1631-1635 (1972).

15 Lynch, H. T. *et al.* Familial association of carcinoma of the breast and ovary. *Surg Gynecol Obstet* **138**, 717-724 (1974).

16 Hall, J. M. *et al.* Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* **250**, 1684-1689 (1990).

17 Narod, S. A. *et al.* Familial breast-ovarian cancer locus on chromosome 17q12-q23. *Lancet* **338**, 82-83 (1991).

30 Chandler, M. R., Bilgili, E. P. & **Merner, N. D.** A Review of Whole-Exome Sequencing Efforts Toward Hereditary Breast Cancer Susceptibility Gene Discovery. *Hum Mutat* **37**, 835-846, doi:10.1002/humu.23017 (2016).

71 Spratt, D. E. *et al.* Racial/Ethnic Disparities in Genomic Sequencing. *JAMA Oncol* **2**, 1070-1074, doi:10.1001/jamaoncol.2016.1854 (2016).

72 Wong, S. L. Medically underserved populations: disparities in quality and outcomes. *J Oncol Pract* **11**, 193-194, doi:10.1200/JOP.2015.004259 (2015).

73 Bustamante, C. D., Burchard, E. G. & De la Vega, F. M. Genomics for the world. *Nature* **475**, 163-165, doi:10.1038/475163a (2011).

74 Ashley, E. A. Towards precision medicine. *Nat Rev Genet* **17**, 507-522, doi:10.1038/nrg.2016.86 (2016).

75 Konkel, L. Racial and Ethnic Disparities in Research Studies: The Challenge of Creating More Diverse Cohorts. *Environ Health Perspect* **123**, A297-302, doi:10.1289/ehp.123-A297 (2015).

76 Streicher, S. A. *et al.* Reasons for participating and genetic information needs among racially and ethnically diverse biobank participants: a focus group study. *Journal of community genetics* **2**, 153-163, doi:10.1007/s12687-011-0052-2 (2011).

83 Frist, W. H. Overcoming disparities in U.S. health care. *Health Aff (Millwood)* **24**, 445-451, doi:10.1377/hlthaff.24.2.445 (2005).

84 Ubri, P. & Artiga, S. *Disparities in Health and Health Care: Five Key Questions and Answers*, <https://www.kff.org/disparities-policy/issue-brief/disparities-in-health-and-health-care-five-key-questions-and-answers/> (2016).

85 Farmer, D. F., Jackson, S. A., Camacho, F. & Hall, M. A. Attitudes of African American and low socioeconomic status white women toward medical research. *J Health Care Poor Underserved* **18**, 85-99, doi:10.1353/hpu.2007.0008 (2007).

86 Brown, D. J. Recruitment of Black Americans for research: A model of success. *Journal of Multicultural Nursing & Health* **10**, 19-23 (2004).

87 Tan, D. S., Mok, T. S. & Rebbeck, T. R. Cancer Genomics: Diversity and Disparity Across Ethnicity and Geography. *J Clin Oncol* **34**, 91-101, doi:10.1200/JCO.2015.62.0096 (2016).

88 Taylor, J. Y. Recruitment of three generations of African American women into genetics research. *Journal of transcultural nursing : official journal of the Transcultural Nursing Society* **20**, 219-226, doi:10.1177/1043659608330352 (2009).

89 Newman, B. *et al.* The Carolina Breast Cancer Study: integrating population-based epidemiology and molecular biology. *Breast Cancer Res Treat* **35**, 51-60 (1995).

90 Palmer, J. R., Ambrosone, C. B. & Olshan, A. F. A collaborative study of the etiology of breast cancer subtypes in African American women: the AMBER consortium. *Cancer Causes Control* **25**, 309-319, doi:10.1007/s10552-013-0332-8 (2014).

91 Susan G. Komen North Central Alabama Affiliate 2015 Community Profile. (2015).

92 *Alabama Department of Public Health, Rural Health, Shortage Area Designations*, <http://www.adph.org/ruralhealth/Default.asp?id=901> (2017).

93 Gyawu, R., Quansah, J. E., Fall, S., Gichuhi, P. N. & Bovell-Benjamin, A. C. Community food environment measures in the Alabama Black Belt: Implications for cancer risk reduction. *Prev Med Rep* **2**, 689-698, doi:10.1016/j.pmedr.2015.08.015 (2015).

94 Rastogi, S., Johnson, T. D., Hoeffel, E. M. & Drewery, M. P. United States Census Bureau. The Black Population: 2010. (2010).

95 Miller, S. A., Dykes, D. D. & Polesky, H. F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic acids research* **16**, 1215 (1988).

96 Salowe, R. *et al.* Cost and yield considerations when expanding recruitment for genetic studies: the primary open-angle African American glaucoma genetics study. *BMC Med Res Methodol* **17**, 101, doi:10.1186/s12874-017-0374-9 (2017).

97 Helgesson, G. Are enrollment sites the key to optimizing participation in genetic studies? *Genome Med* **3**, 41, doi:10.1186/gm257 (2011).

98 Brandt, A. M. Racism and research: the case of the Tuskegee Syphilis Study. *Hastings Cent Rep* **8**, 21-29 (1978).

99 Greiner, K. A. *et al.* Effective recruitment strategies and community-based participatory research: community networks program centers' recruitment in cancer prevention studies. *Cancer Epidemiol Biomarkers Prev* **23**, 416-423, doi:10.1158/1055-9965.EPI-13-0760 (2014).

100 *United States Census Bureau: Alabama Quick Facts*, <https://www.census.gov/quickfacts/fact/table/AL#viewtop> (2017).

101 Lynch, H. T. Family information service and hereditary cancer. *Cancer* **91**, 625-628 (2001).

102 Churpek, J. E. *et al.* Inherited predisposition to breast cancer among African American women. *Breast Cancer Res Treat* **149**, 31-39, doi:10.1007/s10549-014-3195-0 (2015).

103 Zenka, D. (2012).

104 Palmer, J. R., Boggs, D. A., Adams-Campbell, L. L. & Rosenberg, L. Family history of cancer and risk of breast cancer in the Black Women's Health Study. *Cancer Causes Control* **20**, 1733-1737, doi:10.1007/s10552-009-9425-9 (2009).

2.8 Tables

Table 2.1: Summary of proband characteristics.

AHCC		BC probands							OvC probands					PC probands				
		Total number of individuals			Individuals diagnosed ≤ 45 yoa	Individuals diagnosed > 45 yoa	Average age of BC onset*	Range of BC onset	Total	Individuals diagnosed ≤ 45 yoa	Individuals diagnosed > 45 yoa	Average age of OvC onset	Range of OvC onset	Total	Individuals diagnosed ≤ 45 yoa	Individuals diagnosed > 45 yoa	Average age of PC onset	Range of PC onset
		Female	Male	Total														
Hospital recruitment	African American	20	1	21	9	12	50.0	23-69	0	0	0	N/A	N/A	1	0	1	63	63
	European American	57	0	57	22	35	50.0	24-70	3	0	3	57	52-63	1	0	1	51	51
	Asian American	1	0	1	1	0	39.0	39	0	0	0	N/A	N/A	0	0	0	N/A	N/A
	Combined	78	1	79	32	47	49.8	23-70	3	0	3	57	52-63	2	0	2	57	51-63
CBR	African American	32	0	32	19	13	46.1	32-61	1	1	0	24	24	2	0	2	62.5	53-72
	European American	31	1	32	14	18	46.3	24-69	0	0	0	N/A	N/A	3	0	3	69.7	62-81
	Asian American	1	0	1	1	0	39	39	0	0	0	N/A	N/A	0	0	0	N/A	N/A
	Combined	64	1	65	34	31	46.1	24-70	1	1	0	24	24	5	0	5	66.8	53-81
Total	African American	52	1	53	28	25	47.6	23-69	1	1	0	24	24	3	0	3	62.7	53-72
	European American	88	1	89	36	53	48.7	24-70	3	0	3	57	52-63	4	0	4	65	51-81
	Asian American	2	0	2	2	0	39	39	0	0	0	N/A	N/A	0	0	0	N/A	N/A
	Combined	142	2	144	66	78	48.1	23-70	4	1	3	48.8	24-63	7	0	7	64	51-81

Legend: (yoa) years of age; (*) If individuals were diagnosed with multiple primary BC tumors, the age of BC used in this table was their age at the first BC diagnosis.

2.9 Figures

Figure 2.1: Map of Alabama divided into counties. Medically underserved areas and populations, and counties of residence of current study participants are highlighted; see figure legend. The original map was obtained from the Alabama Department of Public Health (ADPH) website (http://www.alabamapublichealth.gov/ruralhealth/assets/MUAPMap.pdf) with permission to use to demonstrate recruitment progress. According to the ADPH, medically underserved areas are "a measure of the number of health professionals and certain health outcomes that demonstrate a lack of access and impact on the health of the community. Medically underserved populations are very similar to medically underserved areas except that they are designating the low-income population rather than the geographical region" (http://www.alabamapublichealth.gov/ruralhealth/assets/MUAP_101.pdf).

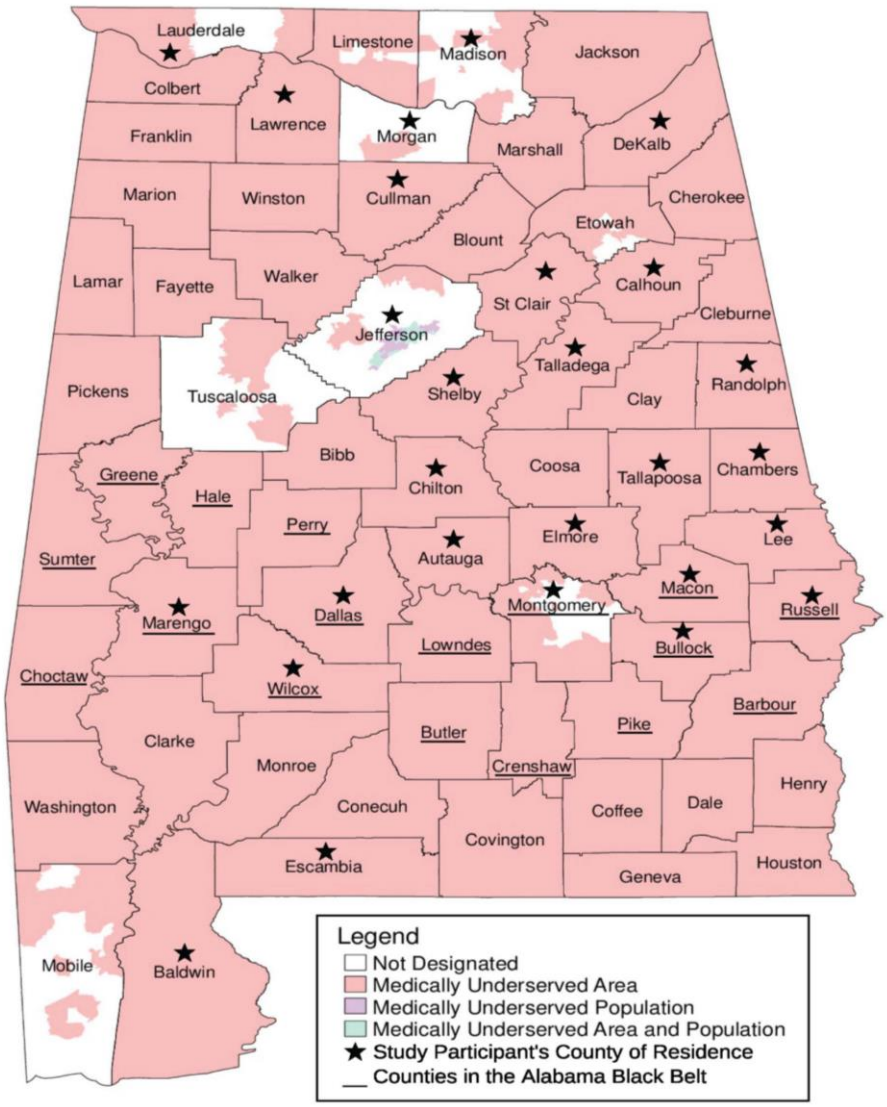
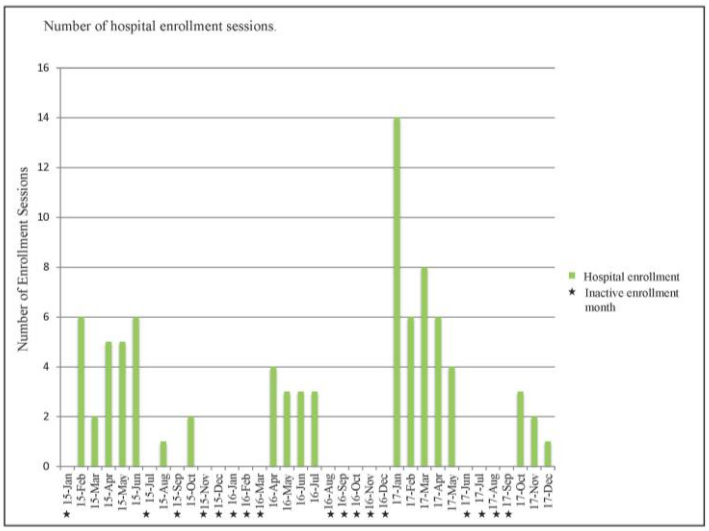
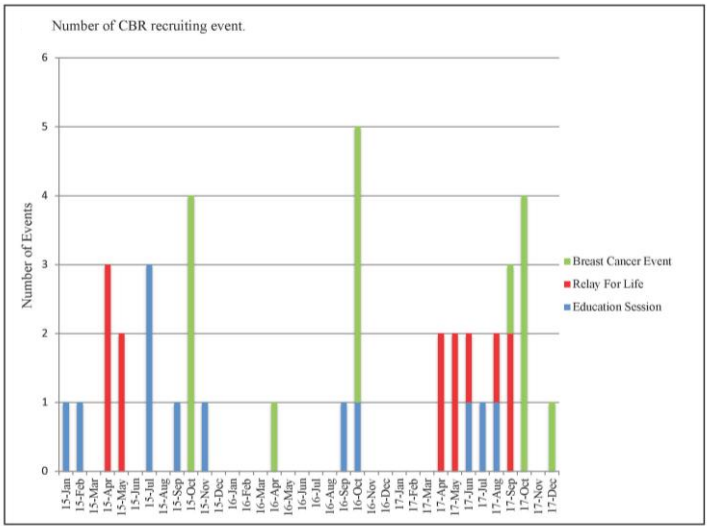


Figure 2.2: Recruiting events and enrollment sessions over the 36-month period.

A.



B.



C.

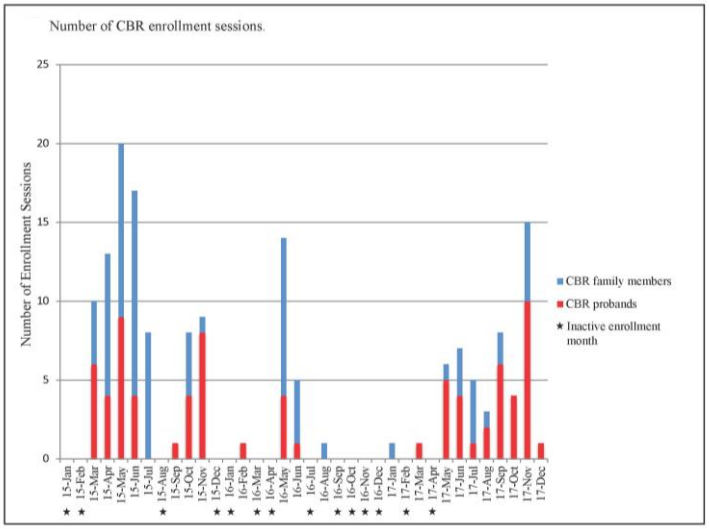
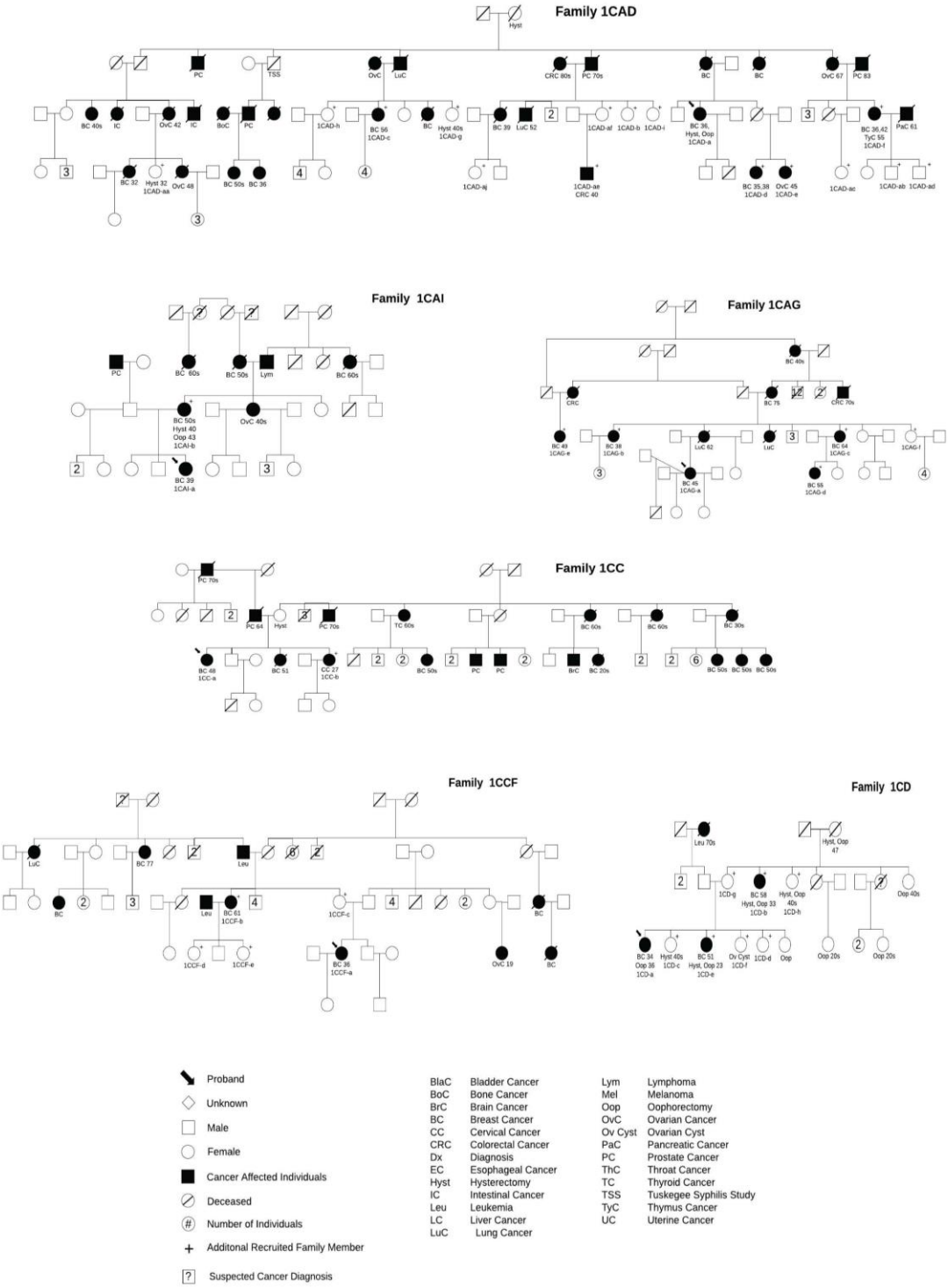


Figure 2.3: Selected African American pedigrees of families with multiple study participants.



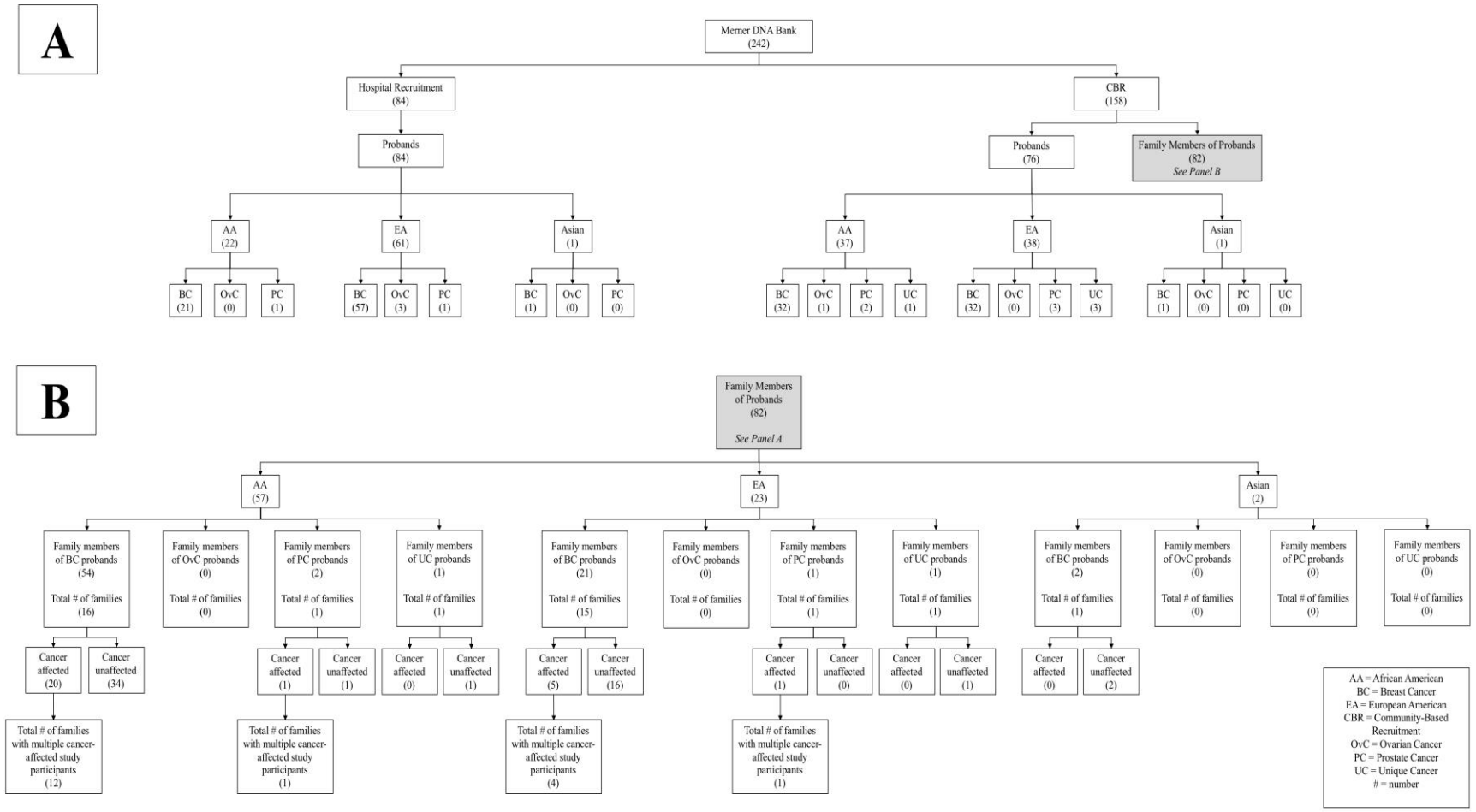
2.10 Supplementary Material

Supplementary Table 2.1: Descriptions of five IRB-approved community partners.

Name	Mission	Research Involvement
Alabama Breast and Cervical Cancer Early Detection Program (ABCCEDP)	An ADPH Initiative that provides free BC and cervical cancer screenings for underserved women who meet certain poverty eligibility guidelines.	Disseminates an IRB-approved information letter about the research study to women who were diagnosed with BC through the ABCCEDP.
Macon M.E.A.N.S for Cancer - Support Group	An African American BC support group in Tuskegee, Alabama that focuses on mentorship, education, advocacy, and nutrition support.	Disseminates brochures and flyers about the research study and invites the CBR team to support group meetings.
SCSC	An African American BC support group in Montgomery, Alabama that provides a culturally sensitive focus to combat and cope with cancer. Their mission is to increase cancer survivorship by improving quality of life of survivors amongst the medically underserved.	Disseminates brochures and flyers about the research study and invites the CBR team to support group meetings and speak at their annual BC walk. Founder, Carrie Nelson, is a patient advocate on this grant proposal and offers a patient perspective on the study design.
Working to Help Those In Pink (W.H.I.P.)	An African American BC support group in east Alabama that offers help to survivors, family members, and supporters of BC.	Disseminates brochures and flyers about the research study, and invites the CBR team to support group meetings and an education booth at their annual BC walk.
Young Breast Cancer Survivorship Network (YBCSN)	An initiative out of the University of Alabama at Birmingham School of Nursing that provides targeted online resources to young women facing BC, with a central focus to improve their quality of life through education, support, and networking.	Disseminates brochures and flyers about the research study, and has invited the CBR team to support group meetings and their annual BC workshop.

Supplementary Figure 2.1: The Gene Machine, a mobile recruitment and enrollment station.

Supplementary Figure 2.2: Flowchart of samples in the Merner DNA Bank. Panel A illustrates how the DNA samples in the bank are divided into recruitment mechanism, participation category (proband or family member), ethnicity, and cancer type. Panel B further explains characteristics of additional family member who have been recruited into the study.



Chapter 3: A research-based gene panel to investigate breast, ovarian and prostate cancer genetic risk.

This information has been submitted to *Cancer Genetics*: Bishop MR, Huskey ALW, Omeler SM, Hetzel J, Merner ND. A research-based gene panel to investigate breast, ovarian and prostate cancer genetic risk. *Cancer Genetics*; Submitted: July 1, 2018. Submission number: CG_2018_227.

3.1 Abstract

There is a concern that clinically accredited NGS gene panels that assess inherited cancer risk contain genes lacking clinical validity/utility. Therefore, a research-based NGS gene panel, B.O.P., was developed to evaluate the genetic risk of BC, OvC and/or PC. This manuscript serves as an introduction to B.O.P. and highlights the initial analytical validity through the assessment of 10 genes. Forty-three individuals from the AHCC were B.O.P. screened. Upon bioinformatics processing and variant filtering, 74 variants were carried through for validation. A total of 61 were TPs with an average sequencing depth of 659X and alternate allele frequency of 51%. The average FP sequencing depth was 34X and alternate allele frequency was 33%. Although low sequencing depth was not always indicative of a FP, high sequencing depths (>100X) signified a TP. Furthermore, sensitivity and specificity of *BRCA1/2* were calculated to be 100% and 92.3%, respectively. Overall, this screening enabled the establishment of criteria to alleviate future validation efforts and strongly supports the use of B.O.P. to further elucidate hereditary cancer susceptibility. Ultimately, continued B.O.P. screening aims to impact the clinical validity/utility of accredited NGS gene panels.

Keywords: Gene panel screening, next-generation sequencing, breast cancer, African American, and analytical validity

3.2 Introduction

Gene panels enable the simultaneous screening of a number of genes. Panels are typically customized for specific screening purposes; thus, the genes (and even specific gene regions) on such panels are unique to the screening goals. In recent years, with technological sequencing advances, panel-based screening has become extremely efficient and cost-effective. These advancements involve the targeted enrichment of selected genes followed by massively parallel sequencing, which is also known as NGS.[5,53]

Although NGS gene panels have been implemented into clinical practice to assess inherited risk of cancer, it has not occurred without debate.[5,53,105] Despite a proposed framework for the evaluation of clinical genetic tests by the ACCE (whose name was derived from the four suggested evaluation criteria, **a**nalytical validity, **c**linical validity, **c**linical utility, and **e**thic, legal, and social issues)[106], tests are currently regulated by CLIA, which primarily assesses analytical validity – the accuracy of mutation detection.[62] The American College of Medical Genetics and Genomics (ACMG) has also established clinical laboratory standards for NGS and highlights quality control (QC) challenges.[107] Thus, a concern is that clinically accredited NGS gene panels assessing inherited cancer risk contain genes lacking clinical validity and utility. Clinical validity considers the strength of the association between disease risk and particular gene mutations, while clinical utility refers to the impact of the genetic test on disease management (such as screening, surveillance, treatment, etc.).[5,53]

The NCCN aids in maximizing clinical utility by providing genetic risk assessment criteria and mutation-positive management strategies regarding risk genes that are clinically valid.[108] However, for some inherited cancers, clinically valid genes only account for a minority of the associated genes reported in the literature; for instance, over a hundred genes harboring variants of high, moderate and/or low risk have been reported to be associated with BC risk,[30] but the NCCN provides mutation-positive management strategies for only 11 BC susceptibility genes.[109] Furthermore, NCCN BC risk management strategies have primarily been developed for overtly pathogenic, truncation mutations since the clinical relevance of other variants in those 11 genes, such as missense variants, is commonly unknown, hence the term VUSs.[5]

In order to properly classify genes/variants as risk-associated, there is a need for further research efforts to provide evidence that either supports or refutes previous claims. Developing research-based NGS gene panels that target genes with previously suggested associations that lack sufficient evidence in addition to NCCN clinically actionable genes can help aid this effort. This manuscript serves as an introduction to one such panel, B.O.P., which is an acronym for **B**reast, **O**varian, and **P**rostate, highlights its initial analytical validity assessments, and discusses applications for future use that can impact the clinical validity and utility of accredited gene panels.

3.3 Materials and Methods

3.3.1 Ethical compliance and informed consent

All procedures performed in studies involving human participants were in accordance with the ethical standards of AU and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Specifically, this research was reviewed and approved by the AU IRB for the recruitment, enrollment and biobanking of the AHCC (IRB protocols 14-232, 14-335,

and 15-111).[69] Informed consent was obtained from all individual participants included in the study.

3.3.2 Panel design

B.O.P. targets ~500 kilobases (Kb) of DNA including genes that are suggested, predicted, or clinically proven to be associated with BC, OvC, and/or PC risk. Agilent Technologies Haloplex probes were designed using Agilent Technologies SureDesign software (https://earray.chem.agilent.com/suredesign/). The "Advanced HaloPlex" design allowed for the selection of the desired genes of which the targeted regions included both coding and non-coding exons as well as 10 intronic base pairs flanking the intron-exon boundaries. The probe set was designed for 100 base pair reads on an Illumina platform. Overall, probes were predicted to cover 98.93% of the targeted genes/regions (Table 3.1).

3.3.3 Capture and sequencing

The genomic DNA of 43 individuals (23 African American [AA] and 20 European American [EA]) from the AHCC[69] was selected to undergo the first B.O.P. screening; this included 41 BC-affected, 1 PC-affected and 1 UC-affected individuals (Figure 3.1 and Figure 3.2). Two study participants (1CAD-a and 1CAD-f) were knowingly related (first cousins). The HaloPlex HS Target Enrichment System For Illumina Sequencing Protocol (Version C0, December 2015) was followed for the targeted-capture, allowing each of the 43 samples to be uniquely barcoded/indexed, individually captured, and pooled in equimolar amounts for Illumina paired-end sequencing. One pooled sample with a final concentration of 24.13 nanomoles/liter (and DNA fragments ranging from 175 to 625 base pairs) was sent for sequencing on one lane of a flow cell on an Illumina HiSeq™ 2500 at the Genomic Services Laboratory (GSL) at HudsonAlpha Institute

for Biotechnology. The final DNA quality/quantity of the pooled sample was assessed using the High Sensitivity DNA kit using the ABI 2100 Bioanalyzer.

3.3.4 Bioinformatics analyses

The sequencing data generated for each indexed sample (43 forward and 43 reverse FASTQ files) were downloaded using the GSL's wget downloader (Figure 3.2). Trimmomatic (v.0.35) was used to trim the unique barcodes and Illumina adaptors. After generating trimmed FASTQ files, FastQC (v.0.10.1) was used to ensure that the repeated sequences had been trimmed from the sequences. The trimmed, paired forward and reverse FASTQ files were then aligned to the soft-masked human reference genome (GRCh38/hg38) using Burrows-wheeler Aligner (BWA v.0.7.12), generating SAM (Sequence Alignment/Mapping) files, which were then compressed into BAM (binary SAM) files and sorted using SAMtools (v.1.2). PicardTools (v.1.79) was used to add read groups and then index the sorted, compressed BAM files along with realigning insertions and deletions (indels). As previously suggested for Haloplex, duplicates were not marked or removed [110]. Variants in B.O.P targeted-regions were called from the sorted BAM files using the HaplotypeCaller tool in the Genome Analysis Tool Kit (GATK; v.3.4-46); the generated VCF (Variant Calling Format) files were then merged using Tabix (v.0.2.6) and VCFTools (v.0.1.12a). Variants in the merged VCF files were annotated using ANNOVAR (June2017). Overall, this pipeline was adapted from the GATK Best Practice Pipeline.[111] Lastly, Samtools flagstat (v.1.2) was used to gather metrics of the analyzed reads, and the DepthOfCoverage tool within the GATK (v.3.4-46) was used to calculate the depth of coverage for the targeted regions.

3.3.5 Analytical assessment

Ten BC susceptibility genes with clinical validity and utility based on NCCN guidelines [109] were selected for analytical validation (Table 3.1). Using Exome Variant Server (EVS) as a control repository, B.O.P. variants in those 10 genes were filtered for MAFs of less than or equal to 2% in both EAs and AAs.[112] Variants were further filtered; all coding variants as well as intronic variants that were located within 10 base pairs of an intron-exon boundary were carried through for validation using PCR and Sanger sequencing (Figure 3.1). Primer sequences and amplification conditions are available upon request. P values and ORs were calculated using Fisher exact test in R (v 3.5.1).

Upon consent and enrollment into the AHCC, study participants provided information about previous clinical genetic screening.[69] Thus, for genes with clinical screening results provided by the 43 participants involved in this initial B.O.P. screening, sensitivity and specificity were calculated. Sensitivity was defined as the total number of TPs divided by the sum of the total number of TPs and false negatives (FNs; TPs/ [TPs + FNs]). TPs were defined as (i) variants that had been previously identified through clinical gene screening, initially confirmed in the research laboratory by PCR and Sanger sequencing, and, subsequently, detected upon B.O.P. screening, or, (ii) in the case of no clinical screening results, as variants detected upon B.O.P. screening and then validated through PCR and Sanger sequencing (Figure 3.1 and 3.2). FNs were variants that had been previously identified through clinical gene screening and confirmed in the research laboratory by PCR and Sanger sequencing but not detected through B.O.P. screening. Specificity was defined as the ratio of the total number of true negatives (TNs) over the sum the total number of TNs and false positives (FPs; TNs / [TNs + FPs]). TNs were defined as individuals who had no pathogenic variants detected through clinical gene screening as well as B.O.P. screening. FPs were defined as variants detected through B.O.P. screening but not validated upon subsequent PCR and Sanger sequencing (Figure 3.1 and 3.2). Lastly, the false discovery rate (FP / [TP + FP]) was calculated

for the 10 BC susceptibility genes individually. FDRs were calculated in two ways, (i) including all B.O.P. called variants, and (ii) excluding B.O.P. variants with an alternate allele frequency less than or equal to 20%.

3.4 Results

The number of reads that passed QC assessment per individual averaged 11.3 million (M), with 98.6% of those reads mapping to the human genome. On average, 50.9% of the reads mapped to B.O.P. targeted regions (Supplementary Table 3.1). The average sequencing depth for all targeted base pairs was 809X; however, sequencing depth was not uniform with a large interquartile range (Table 3.1). A probe design report provided by Agilent Technologies predicted that 98.9% of the targeted base pairs would be covered at least 1X, which was similar to the actual coverage of 98.2%; thus, 1.8% of the targeted base pairs were not covered at all (Table 3.1). The ten clinically relevant BC susceptibility genes had average sequencing depths that ranged from 505X-1017X (Table 3.1). Furthermore, assessment of the 225 different regions that targeted those ten genes revealed that the majority had average sequencing depths between 800-899X but ranged from 68X-2053X (Table 3.2; Supplementary Table 3.2). Although rare, regions with average sequencing depths less than 100X missed on average 24.3% of the targeted base pairs and only covered 52.2%, 34.6%, and 21.3% of targeted base pairs at or greater than 20X, 50X, and 100X, respectively (Table 3.2; Supplementary Table 3.2). Regions with the highest average sequencing depth, 1500X or greater, had over 99% and 96.5% of the targeted base pairs covered at least 50X and 100X, respectively. However, 28.0% of the 225 regions-of-focus did not, on average, obtain 100% coverage at 1X, which included regions with average sequencing depths ranging from 68X-1354X (Table 3.2; Supplementary Table 3.2).

Upon variant annotation (Figure 3.2), a total of 24,915 variants (2,858 unique) were called. After filtering for variants in the ten genes (Table 3.1), a total of 1960 (287 unique) remained, 74 (56 unique) of which had MAFs less than 2% in both ethnicities (Table 3.3; Figure 3.2). A total of 61 of the 74 variants were validated and classified as TPs, averaging a sequencing depth of 659X and alternate allele frequency of 51%; this included 100% of the variants in seven out of the 10 genes (*ATM*, *BRCA2*, *CHEK2*, *NBN*, *PALB2*, *STK11*, and *TP53*), resulting in FDRs of 0 (Table 3.3). *BRCA1* and *CDH1* each had one FP, and *PTEN* had the highest FDR with 11 FPs, all in intron7/exon8 (Table 3.3). Despite that the average FP sequencing depth was 34X (ranging from 12X-63X), sequencing depths of the three regions harboring FPs revealed that all achieved an average greater than 427X (Table 3.3 and Supplementary Table 3.2). The average FP alternate allele frequency was 33%, ranging from 13%-68%.

Though not optimal, low sequencing depth was not always indicative of a FP. Of the 20 variants covered less than 100X, 13 were FPs, and 7 were TPs with average sequencing depth of 60X and alternate allele frequency of 48%, ranging from 33%-58% (Table 3.3). In contrast, higher coverage, such as sequencing depth greater than 100X, was an indicator of a TP; all 54 variants covered over 100X were determined to be TPs. This included two homozygous TPs, which were each covered over 1000X with the alternate allele being the only one detected (Table 3.3), and 52 heterozygous TPs that had an average sequencing depth of 724X and alternate allele frequency of 50%. To note the importance of alternate allele frequency, 95% of the TPs had an alternate allele frequency above 40% compared to only 23% of the FPs. Eighty five percent of the TPs had over 100X coverage and 40% alternate allele frequency (Table 3.3). No variants with an alternate allele frequency less than 20% were TPs, and additional filtering to exclude such variants improved FDRs (Table 3.3).

Prior to B.O.P. screening, positive and negative *BRCA1* and *BRCA2* mutation status was known for eight of the 43 study participants; thus, sensitivity and specificity could be calculated for those genes. Seven study participants had previously undergone clinical *BRCA1/2* screening; six reported negative results with no pathogenic variants identified. One individual, 1CB-a, received a positive report indicating a pathogenic *BRCA2* frame-shifting mutation (c.5611_5615delAGTAA [p.S1871fs] also known as c.5616_5620delAGTAA [p.K1872Nfs), which was confirmed using PCR and Sanger sequencing prior to B.O.P. screening (Figure 3.3 and Table 3.3). The eighth individual, 1CAD-a, had not personally obtained clinical gene screening; however, a deceased family member had undergone clinical *BRCA1/2* screening and received a positive report indicating a pathogenic missense mutation (*BRCA1* c.5387T>G [p.M1796R]). Thus, this individual was screened for the familial mutation using PCR and Sanger sequencing prior to B.O.P. screening and tested positive (Figure 3.3 and Table 3.3). Noteworthy, another BC-affected family member, 1CAD-f, tested negative for *BRCA1* p.M1796R in the research laboratory prior to B.O.P. screening but could not be considered a TN for the specificity calculation since full gene screening had not been carried out (Table 3.3). B.O.P. variant calling reported 12 and 14 variants in *BRCA1* and *BRCA2*, respectively (Table 3.3). Upon Sanger sequencing confirmation, this included 11 TPs, zero FNs, six TNs and one FP in *BRCA1*, and 14 TPs, six TNs, and zero FNs and FPs in *BRCA2* (Table 3.3), which corroborated the previously reported *BRCA1* and *BRCA2* mutation statuses. Therefore, B.O.P. screening of *BRCA1/2* resulted in 100% sensitivity and 92.3% specificity. However, specificity became 100% with the elimination of variants with alternate allele frequencies of 20% or less.

Of the 61 TPs, 45 were detected in AAs; this included 34 unique variants, eight of which were detected in multiple individuals (Table 3.3). According to ClinVar [113,114], the 34 variants were

categorized as pathogenic/risk factor (n=4), VUSs (n=11), or benign/likely benign (n=19). A total of five variants were predicted to be deleterious in Polyphen, two of which have been defined as pathogenic non-synonymous variants in ClinVar; the other three are currently classified as VUSs (Table 3.3). Of the eight variants detected in more than one individual, *BRCA2* c.5020A>G; p.S1674G, was identified in two first cousins. The remaining seven were in seemingly unrelated individuals. This includes *STK11* c.369G>A;p.Q123Q, a seemingly benign variant, which is reported to have a MAF of 1.5% in the general AA population but was detected in five of the 23 AAs in this study, indicating a MAF of 10.8% (*P value* 8.50 X 10^{-4}; *OR* 7.79 CI_{95}[2.32-20.70]). Furthermore, the 45 AA TPs were detected in 96% of the AAs screened, and multiple variants were detected in 70% of the cases. In contrast, 16 TPs were validated in 55% of the EAs, and only 20% had multiple variants. The difference in the number of individuals from each ethnicity with at least one TP was significant (*P value* 2.71 X 10^{-3}; *OR* 16.83 CI_{95}[1.93-819.72]) as well as the number of cases from each ethnicity with multiple TPs (*P value* 1.95 X 10^{-3}; *OR* 8.60 CI_{95}[1.89-49.30]). No overtly pathogenic variants were validated in EAs, but 50% of the EA TPs (8/16) were listed as a VUS, three of which were predicted to be deleterious in Polyphen.[115]

3.5 Discussion

Our group has developed B.O.P., a research-based NGS gene panel, which targets genes that have been suggested, predicted, or clinically proven to be associated with risk of BC, OvC, and/or PC. The overall purpose of this new panel is to gain additional insights toward the genetic risk of and overlap between those three cancers. This particular manuscript serves to highlight the initial assessment of 10 clinically valid BC susceptibility genes[109] and report seminal findings that hint towards the panel's discovery potential.

The targeted enrichment method for the B.O.P. panel is amplicon-based. It uses custom-designed Haloplex probes from Agilent Technologies to generate amplicons for sequencing on an Illumina platform. Haloplex is unique from other targeted enrichment methods in a number of ways. For instance, it uses restriction enzymes to fragment DNA prior to probe hybridization, allowing multiple samples to be fragmented simultaneously and shortening library preparation. Furthermore, Haloplex probes are uniquely designed to hybridize to the 5 and 3 prime ends of their targets, leading to circularized molecules prior to amplification.[110] These differences can affect sequencing outcomes; and, upon comparison to other methods, Samorodnitsky *et al*. concluded that Haloplex had the highest on-target read alignment and normalized sequencing depth but the least uniformity.[110] Noteworthy, despite reports of Haloplex resulting in a high percentage (>90%) of on-target read alignments,[110,116] only 50.9% of our QC passed reads mapped to the B.O.P. targeted regions. With other Haloplex gene panel studies not reporting such data,[117-119] it is difficult to make general conclusions about Haloplex on-target read alignment specificity. However, similar on-target read alignment percentages have been reported; Castera *et al.* used SureSelect baits in order to target hereditary BC and OvC susceptibility genes, and reported an average of 42% of reads on-target.[120] Ultimately, the percentage of off-target reads is likely dependent on a number of factors, including the specific genes/regions being targeted.[121] Of the reads that mapped on-target, B.O.P.'s overall sequencing depth averaged 809X, and each individually assessed gene obtained average sequencing depths from 505X-1017X. Nevertheless, large interquartile ranges indicated that depth was not uniform. This was expected since no current enrichment and sequencing approach provides complete uniformity primarily because of complex genomic regions that are very difficult to capture/sequence and result in low sequencing depths or even no coverage at all.[110,121]

By focusing on a select set of genes/regions, NGS gene panels target a smaller number of base pairs compared to more broad applications such as exome and WGS. The smaller target-capture provides the option to achieve a high average sequencing depth, which aids in variant identification.[110,122] Therefore, the overall goal is to obtain 100% coverage as well as the appropriate/desired sequencing depths at all targeted base pairs. Since this goal is not generally achieved, complementary assays can be used to fill in gaps, which is commonly implemented for clinical applications. In such cases, regions of low/no coverage are normally Sanger sequenced.[121,122] B.O.P. was able to cover, on average, 98.2% of its targeted base pairs at 1X. Being a research panel, no gap-filling assays were carried out; however, region-specific coverage analyses provided insight towards the feasibility of gap-filling. Gap-filling criteria has been described in a number of BC NGS gene panel publications, specifically, those that highlighted panel performance and analytical validity.[120,123-125] Being clinical panels, regions covered less than 20X[120] or 50X[123-125] were checked by conventional methods. Interestingly, only two B.O.P. regions had average sequencing depths less than 100X (68X and 82X), which would not have required complementary assays to fill in gaps according to the criteria set in the referenced studies.[120,123-125] This is despite that, on average, those two B.O.P. regions missed 24.3% of the targeted base pairs and only covered 52.2%, 34.6%, and 21.3% of targeted base pairs at or greater than 20X, 50X, and 100X, respectively. Furthermore, 63 of the 225 B.O.P. regions-of-focus did not, on average, obtain 100% coverage at 1X; these regions had average sequencing depths ranging from 68X-1354X. Thus, regions with high sequencing depths still had base pairs with no/low coverage, which happened to be where FPs were detected in this study. Therefore, only gap-filling regions with 'low' (20X or 50X) sequencing depths, will not guarantee 100% coverage. Establishing mapping criteria to ensure all base pairs are covered at a desired depth is ideal but would likely reveal gaps in too many regions, making gap-filling infeasible. Overall, gaps in B.O.P. as well as

other panels, even with gap-filling criteria, can provide less than definitive negative results;[122] however, in noting that, zero FNs were identified in *BRCA1* and *BRCA2*, resulting in 100% sensitivity.

In addition to gap-filling, conventional approaches are also used to validate called variants. A total of 1960 variants were detected in the 10 B.O.P. assessed genes and, to reduce the number of variants to validate for this analytical assessment, only variants with MAFs less than 2% in both ethnicities were Sanger sequenced. This included 74 variants, 61 of which were confirmed and defined as TPs revealing 13 FPs. The validation process ultimately provided insight regarding the likelihood of confirmation based on variant quality, such as sequencing depth since all 54 variants covered over 100X were TPs. These results corroborated the criteria established by Mu *et al*.,[126] which set high confidence calls as having a minimum sequencing depth of 100X and alternate allele frequency of 40%. Additionally, Mu *et al*. indicated that such calls did not require confirmation. Although, all B.O.P. variants covered at or above 100X were TPs despite alternate allele frequency, the criteria from Mu *et al.* will be implemented in the future in order to be thorough.[126] This will limit validation efforts to low confidence calls, reducing the cost and time of validation.

In this study, 22 variants had low confidence calls. This included nine TPs, seven of which were covered less than 100X and two that failed to meet the required alternate allele frequency. The remaining 13 were FPs with an average sequencing depth of 34X and alternate allele frequency of 33%, reiterating that low sequencing depths are susceptible to sequencing artifacts.[110,122] Interestingly, as mentioned in the previous paragraph, the regions harboring the FPs did not have low sequencing depths, stressing the potential lack of uniformity within a targeted region. On another note, 11 of the 13 FPs were in *PTEN*. Considering *PTEN* has a processed pseudogene, *PTENP1* on chromosome 9, their homology could have contributed to probe mis-

priming as well as read mis-alignments. Overall, encountering problematic regions, such as regions with high homology or GC rich content, is common and referred to in many studies.[117,121,122] Overall, for each assessed gene, FDRs ranged from 0 to 0.92, the latter being *PTEN*. Of course, FDRs improved as additional filtering was implemented. Initial B.O.P. specificity, which could only be calculated for *BRCA1*/*BRCA2*, was 92.3%. Upon filtering out variants with alternate alleles frequencies equal to or less than 20%, specificity was 100%. Ultimately, Sanger sequencing all low confidence calls will eliminate FPs and provide 100% specificity; therefore, it is common to complement NGS gene panels with Sanger sequencing validation in order to consider the test complete and optimize specificity.[120,123-127]

In addition to enabling B.O.P.'s initial analytical assessment, the first B.O.P. screening, which involved 43 individuals (23 AAs and 20 EAs) from the AHCC,[69] has provided insight regarding variant contributions and ethnic differences. Overall, compared to EAs, AAs had a significantly higher number of individuals with at least one TP (*P value* 2.71 X 10^{-3}) as well as individuals with multiple TPs (*P value* 1.95 X 10^{-3}). Of course, comparisons to ethnic-specific controls will determine if these differences contribute to an inherited cancer risk. Interestingly, according to ClinVar,[113,114] none of the variants identified in EAs were considered pathogenic/risk variants, whereas 17.4% (4/23) of the AAs had a variant with that classification. The majority of the detected variants were classified as VUSs or benign/likely benign; ultimately, elucidating how VUSs and, even, synonymous variants contribute towards risk is very important. Synonymous variants, though normally ignored and considered benign, can affect splicing, gene expression or translation dynamics, all of which can contribute to a disease phenotype.[128] To further stress their importance, they have been reported to act as driver mutations in human cancers [129] and, through this initial B.O.P. screening, *STK11* c.369G>A;p.Q123Q was detected in significantly more AA cases than controls (*P value* 8.50 X 10^{-4}). Additionally, despite recognizing that hereditary BC risk

is polygenic,[130] little effort has been put forth to thoroughly investigate all variants in clinically relevant BC susceptibility genes and determine if different variant combinations increase risk. Altogether, seemingly benign variant combinations could, in fact, be pathogenic, and paired with the striking difference between ethnicities regarding the number of cases with multiple variants, further investigation is warranted.

3.6 Conclusions

In summary, this effort assesses the analytical validity of the B.O.P. panel and demonstrates the panel's ability to accurately detect mutations in 10 NCCN clinically actionable genes [109]. Despite the potential biases of the B.O.P. capture and NGS, the high depth of coverage, low FDR, and great sensitivity and specificity strongly support the use of this research gene panel to further elucidate hereditary BC/OvC/PC genetics. Although the cohort for this initial assessment is small, B.O.P. has begun to determine the mutation contributions of clinically valid genes in different ethnicities as well as permit the investigation of VUSs and other variant types and their effect towards polygenic risk. Furthermore, continued B.O.P. screening will provide additional evidence to confirm or refute previously suggested susceptibility genes, lessening the number of genes that lack clinical validity on commercially available panels. Lastly, with the incorporation of candidate genes on B.O.P., it has the potential to identify novel genetic risk factors that are contributing towards BC, OvC, and PC. As described herein, the potential implications for the implementation of B.O.P. screening are immense and will, in the near future, lead to reductions in morbidity and mortality through risk management options.

3.7 References

5 Easton, D. F. *et al.* Gene-panel sequencing and the prediction of breast-cancer risk. *N Engl J Med* **372**, 2243-2257, doi:10.1056/NEJMsr1501341 (2015).

30 Chandler, M. R., Bilgili, E. P. & **Merner, N. D.** A Review of Whole-Exome Sequencing Efforts Toward Hereditary Breast Cancer Susceptibility Gene Discovery. *Hum Mutat* **37**, 835-846, doi:10.1002/humu.23017 (2016).

53 Shah, P. D. & Nathanson, K. L. Application of Panel-Based Tests for Inherited Risk of Cancer. *Annu Rev Genomics Hum Genet* **18**, 201-227, doi:10.1146/annurev-genom-091416-035305 (2017).

62 Centers for Disease Control and Prevention. *Clinical Laboratory Improvement Amendments (CLIA)*, <https://wwwn.cdc.gov/clia/> (2017).

69 Bishop, M. R. *et al.* Establishment of the Alabama Hereditary Cancer Cohort - strategies for the inclusion of underrepresented populations in cancer genetics research. *Mol Genet Genomic Med*, doi:10.1002/mgg3.443 (2018).

105 Spurdle, A. B., Bowman, M. A., Shamsani, J. & Kirk, J. Endometrial cancer gene panels: clinical diagnostic vs research germline DNA testing. *Mod Pathol* **30**, 1048-1068, doi:10.1038/modpathol.2017.20 (2017).

106 Haddow, J. E. & Palomaki, G. E. in *Human Genome Epidemiology: A Scientific Foundation for Using Genetic Information to Improve Health and Prevent Disease* (eds M. Khoury, J. Little, & W. Burke) 217-233 (Oxford University Press, 2003).

107 Rehm, H. L. *et al.* ACMG clinical laboratory standards for next-generation sequencing. *Genet Med* **15**, 733-747, doi:10.1038/gim.2013.92 (2013).

108 National Comprehensive Cancer Network. *NCCN Guidelines® & Clinical Resources*, <https://www.nccn.org/professionals/physician_gls/default.aspx#detection> (2018).

109 National Comprehensive Cancer Network. *Genetic/Familial High-Risk Assessment: Breast and Ovarian (Version I.2018)*, <https://www.nccn.org/store/login/login.aspx?ReturnURL=https://www.nccn.org/professionals/physician_gls/pdf/genetics_screening.pdf> (2018).

110 Samorodnitsky, E. *et al.* Comparison of custom capture for targeted next-generation DNA sequencing. *J Mol Diagn* **17**, 64-75, doi:10.1016/j.jmoldx.2014.09.009 (2015).

111 Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**, 11 10 11-33, doi:10.1002/0471250953.bi1110s43 (2013).

112 EVS. (Seattle, WA, 2017).

113 Landrum, M. J. *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* **42**, D980-985, doi:10.1093/nar/gkt1113 (2014).

114 Rehm, H. L., Harrison, S. M. & Martin, C. L. ClinVar Is a Critical Resource to Advance Variant Interpretation. *Oncologist* **22**, 1562, doi:10.1634/theoncologist.2017-0246 (2017).

115 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, doi:10.1002/0471142905.hg0720s76 (2013).

116 Lim, E. C. *et al.* Next-generation sequencing using a pre-designed gene panel for the molecular diagnosis of congenital disorders in pediatric patients. *Hum Genomics* **9**, 33, doi:10.1186/s40246-015-0055-x (2015).

117 Green, A. *et al.* Assessment of HaloPlex amplification for sequence capture and massively parallel sequencing of arrhythmogenic right ventricular cardiomyopathy-associated genes. *J Mol Diagn* **17**, 31-42, doi:10.1016/j.jmoldx.2014.09.006 (2015).

118 Erman, B. *et al.* Investigation of Genetic Defects in Severe Combined Immunodeficiency Patients from Turkey by Targeted Sequencing. *Scand J Immunol* **85**, 227-234, doi:10.1111/sji.12523 (2017).

119 Ortega-Moreno, L. *et al.* Molecular diagnosis of patients with epilepsy and developmental delay using a customized panel of epilepsy genes. *PLoS One* **12**, e0188978, doi:10.1371/journal.pone.0188978 (2017).

120 Castera, L. *et al.* Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture targeting multiple candidate genes. *Eur J Hum Genet* **22**, 1305-1313, doi:10.1038/ejhg.2014.16 (2014).

121 Xue, Y., Ankala, A., Wilcox, W. R. & Hegde, M. R. Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing. *Genet Med* **17**, 444-451, doi:10.1038/gim.2014.122 (2015).

122 Harismendy, O. *et al.* Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol* **10**, R32, doi:10.1186/gb-2009-10-3-r32 (2009).

123 Chong, H. K. *et al.* The validation and clinical implementation of BRCAplus: a comprehensive high-risk breast cancer diagnostic assay. *PLoS One* **9**, e97408, doi:10.1371/journal.pone.0097408 (2014).

124 Judkins, T. *et al.* Development and analytical validation of a 25-gene next generation sequencing panel that includes the BRCA1 and BRCA2 genes to assess hereditary cancer risk. *BMC Cancer* **15**, 215, doi:10.1186/s12885-015-1224-y (2015).

125 Lincoln, S. E. *et al.* A Systematic Comparison of Traditional and Multigene Panel Testing for Hereditary Breast and Ovarian Cancer Genes in More Than 1000 Patients. *J Mol Diagn* **17**, 533-544, doi:10.1016/j.jmoldx.2015.04.009 (2015).

126 Mu, W., Lu, H. M., Chen, J., Li, S. & Elliott, A. M. Sanger Confirmation Is Required to Achieve Optimal Sensitivity and Specificity in Next-Generation Sequencing Panel Testing. *J Mol Diagn* **18**, 923-932, doi:10.1016/j.jmoldx.2016.07.006 (2016).

127 Vysotskaia, V. S. *et al.* Development and validation of a 36-gene sequencing assay for hereditary cancer risk assessment. *PeerJ* **5**, e3046, doi:10.7717/peerj.3046 (2017).

128 Buske, O. J., Manickaraj, A., Mital, S., Ray, P. N. & Brudno, M. Identification of deleterious synonymous variants in human genomes. *Bioinformatics* **29**, 1843-1850, doi:10.1093/bioinformatics/btt308 (2013).

129 Supek, F., Minana, B., Valcarcel, J., Gabaldon, T. & Lehner, B. Synonymous mutations frequently act as driver mutations in human cancers. *Cell* **156**, 1324-1335, doi:10.1016/j.cell.2014.01.051 (2014).

130 Kuchenbaecker, K. B. *et al.* Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers. *Journal of the National Cancer Institute* **109**, doi:10.1093/jnci/djw302 (2017).

3.8 Tables

Table 3.1: Ten clinically relevant genes on the B.O.P. panel assessed for analytical validity.

Gene or targeted regions		Accession number	# of targeted regions	Size (bp)	Predicted target 1X coverage* (%)	Average sequencing depth (X)	Interquartile range (X)			% bases covered greater than or equal to:								
							First quartile	Median	Third quartile	1X	10X	20X	50X	100X	250X	500X	1000X	10000X
Genes investigated for analytical validation	*ATM*	NM_000051	65	15545	98.0	781	336	659	1068	97.5	95.7	93.6	87.4	78.3	60.0	41.0	20.8	0.8
	BRCA1	NM_007300	24	7750	98.5	1017	397	787	1271	98.1	97.1	95.6	90.5	81.9	65.2	47.4	27.2	1.1
	BRCA2	NM_000059	28	12078	99.1	960	445	803	1225	98.7	97.8	96.5	91.6	83.5	66.6	48.1	25.7	1.1
	CDH1	NM_004360	16	5269	98.9	934	418	834	1242	98.5	97.1	95.5	91.1	83.4	66.3	48.3	26.6	1.0
	CHEK2	NM_001005735	23	4605	96.5	726	279	588	1023	95.8	93.4	91.3	84.5	74.7	56.6	38.5	18.9	0.7
	NBN	NM_002485	22	6681	98.0	696	312	610	959	97.6	96.0	93.8	86.6	76.8	57.9	38.0	17.9	0.6
	PALB2	NM_024675	13	4318	100.0	1001	557	876	1261	99.9	99.0	98.0	94.2	86.8	70.9	52.3	28.3	1.1
	PTEN	NM_000314	10	10248	98.0	597	221	468	853	96.9	93.2	90.0	81.3	70.6	50.8	31.3	14.1	0.5
	STK11	NM_000455	10	3476	100.0	505	175	387	774	93.2	89.0	85.9	77.6	67.5	46.5	28.3	11.5	0.3
	TP53	NM_000546	14	4216	99.0	788	316	676	1039	97.4	94.6	92.6	87.3	77.7	59.3	41.7	21.4	0.8
All targeted B.O.P. regions			1417	499,521	98.9	809	359	687	1092	98.2	96.6	94.7	88.8	79.9	61.9	43.0	21.6	0.8

*Based on design report by Agilent Technologies

Table 3.2: Average sequencing depth analyses of the 225 regions targeting the ten B.O.P. genes being assessed.

Average region sequencing depth (X)	# of regions	Average % bases covered greater than or equal to:								
		1X	10X	20X	50X	100X	250X	500X	1000X	10000X
<100	2	75.7	63.2	52.2	34.6	21.3	7.3	1.8	0.9	0.0
100-199	4	84.4	75.9	69.1	54.8	39.5	15.6	5.0	2.3	0.0
200-299	7	91.5	88.8	82.6	70.1	54.9	26.1	8.1	2.7	0.0
300-399	12	97.0	93.1	88.2	77.9	64.5	38.5	16.5	5.0	0.0
400-499	22	97.8	95.6	92.8	83.2	71.8	47.9	24.8	7.0	0.1
500-599	21	98.4	96.9	94.8	87.5	76.0	56.7	32.9	11.1	0.2
600-699	27	99.3	98.2	96.2	89.6	79.2	59.3	37.9	14.9	0.5
700-799	28	99.1	97.8	96.6	91.1	81.6	64.5	42.7	19.1	0.7
800-899	29	99.6	98.9	98.0	93.8	85.2	68.8	49.4	24.1	0.9
900-999	13	100.0	99.7	99.2	96.3	88.6	73.2	54.0	27.8	1.2
1000-1099	11	99.8	99.4	98.9	96.1	89.4	74.3	57.6	31.4	1.3
1100-1199	12	99.5	99.1	98.1	95.1	89.0	72.4	57.2	34.6	1.4
1200-1299	14	99.8	99.7	99.4	97.2	91.4	77.2	61.8	37.6	1.6
1300-1399	3	99.8	99.7	99.6	98.7	93.6	80.6	68.3	41.3	2.0
1400-1499	7	100.0	100.0	99.8	98.8	94.7	81.6	68.2	44.2	1.7
>1500	13	100.0	100.0	99.9	99.2	96.5	84.8	72.4	51.3	2.2

Table 3.3. Summary of called variants after bioinformatics pipeline and variant filtering.

Gene Name	Chr	Start position	Ref. Allele	Alt. Allele	Function	Exon/ Intron	DNA Change	Amino Acid Change	Polyphen2 prediction ~	EA EVS#	AA EVS#	CLINVAR ^	Number of individuals with variant called	Individual/ Sample	Ethnicity	GT	GQ	Total Depth	% Alt. Allele	≥ 100X Depth AND 40% Alt Allele	Validation results (TP/FP)	Percent validated	FDR	
																							Including all called variants	Excluding variants with ≤ 20% Alt Allele
ATM (NM_000051)	chr11	108227849	C	G	NS	exon3	c.146C>G	p.S49C	P	0.0136	0.0027	Risk Factor	1	1CAIa	AA	Het	99	830	48%	Yes	TP	100%	0.00	0.00
	chr11	108229171	C	T	Intronic	intron3	c.186-7C>T	N/A	.	.	0.0142	Benign	1	4CAa	AA	Het	99	1685	43%	Yes	TP			
	chr11	108248860	C	T	NS	exon7	c.735C>T	p.V245V	.	0.0134	0.0023	Likely Benign	1	1EDa	EA	Het	99	907	50%	Yes	TP			
	chr11	108248927	T	G	Intronic	intron8	c.1066-6T>G	N/A	.	0.0026	0.0002	VUS	1	1CBEa	EA	Het	99	79	43%	No	TP			
	chr11	108249086	T	C	NS	exon9	c.1229T>C	p.V410A	B	0.0022	0.0009	VUS	1	1EEa	EA	Het	99	376	64%	Yes	TP			
	chr11	108251973	T	C	NS	exon11	c.1744T>C	p.F582L	B	0.0009	.	Likely Benign	1	1EADa	EA	Het	99	240	58%	Yes	TP			
	chr11	108253901	T	C	S	exon13	c.1986T>C	p.F662F	.	0.0006	.	Likely Benign	1	1EACa	EA	Het	99	331	54%	Yes	TP			
	chr11	108254034	T	C	NS	exon13	c.2119T>C	p.S707P	B	0.0109	0.0039	Likely Benign	1	1CBCa	AA	Het	99	581	41%	Yes	TP			
	chr11	108259051	C	A	NS	exon16	c.2442C>A	p.D814E	B	0.0001	0.0198	Likely Benign	1	1EBAa	AA	Het	99	195	37%	No	TP			
	chr11	108267276	T	C	NS	exon17	c.2572T>C	p.F858L	P	0.012	0.0032	Benign	1	1EAJa	EA	Het	99	227	47%	Yes	TP			
	chr11	108284478	G	T	Intronic	intron26	c.3993+5G>T	N/A	.	.	0.0098	Likely Benign	1	1CADf	AA	Het	99	368	59%	Yes	TP			
	chr11	108315883	G	A	NS	exon41	c.6067G>A	p.G2023R	D	0.0031	0.0007	VUS	1	1CGa	EA	Het	99	1725	54%	Yes	TP			
	chr11	108315904	A	G	NS	exon41	c.6088A>G	p.I2030V	B	.	0.0148	Likely Benign	1	1CAFa	AA	Het	99	745	43%	Yes	TP			
	chr11	108317409	G	A	NS	exon43	c.6235G>A	p.V2079I	B	0.0006	0.0166	Benign	1	1EAa	EA	Het	99	997	51%	Yes	TP			
	chr11	108327713	G	A	S	exon48	c.7044G>A	p.T2348T	.	.	0.0014	Likely Benign	1	1CACa	AA	Het	99	103	59%	Yes	TP			
BRCA1 (NM_007300)	chr17	43049113	A	G	Intronic	intron22	c.5469+8T>C	N/A	.	.	0.0145	Likely Benign	2	3CCa	AA	Het	99	1756	45%	Yes	TP	91.67%	0.08	0.00
														4CAa	AA	Het	99	1343	51%	Yes	TP			
	chr17	43051071	A	C	NS	exon21	c.5387T>G	pM1795R	D	.	.	Pathogenic	1	1CADa	AA	Het	99	114	60%	Yes	TP			
	chr17	43070958	C	T	NS	exon16	c.5019G>A	pM1673I	B	0.0152	0.002	VUS	1	1EAa	EA	Het	99	457	49%	Yes	TP			
	chr17	43091492	T	C	NS	exon10	c.4039A>G	pR1347G	B	0.0057	0.0011	VUS	1	1EACa	EA	Het	99	1579	49%	Yes	TP			
	chr17	43092352	T	C	NS	exon10	c.3169A>G	pS1057G	B	.	.	VUS	1	1CEa	AA	Het	99	131	51%	Yes	TP			
	chr17	43092509	T	C	NS	exon10	c.3022A>G	pM1008V	B	.	0.0023	VUS	1	1EAGa	AA	Het	99	187	42%	Yes	TP			
	chr17	43093035	T	A	S	exon10	c.2496A>T	p.P832P	.	.	.	Likely benign	1	1CDa	AA	Het	99	1466	48%	Yes	TP			
	chr17	43093626	A	G	S	exon10	c.1905T>C	pN635N	.	.	0.0005	Likely benign	1	1CFa	AA	Het	99	264	52%	Yes	TP			
	chr17	43094408	G	T	NS	exon10	c.1123C>A	pL375I	P	.	.	.	1	1CGa	EA	Het	99	63	13%	No	FP			
	chr17	43097280	G	T	NS	exon7	c.557C>A	p.S186Y	D	.	0.0068	VUS	2	3CCa	AA	Het	99	814	50%	Yes	TP			
														4CAa	AA	Het	99	407	50%	Yes	TP			
BRCA2 (NM_000059)	chr13	32332629	C	T	NS	exon10	c.1151C>T	p.S384F	D	0.0015	0.0002	VUS	1	1CDa	AA	Het	99	1229	52%	Yes	TP	100%	0.00	0.00
	chr13	32332753	A	G	S	exon10	c.1275A>G	p.E425E	.	.	0.0098	Benign	1	1CAIa	AA	Homo	99	1046	100%	Yes	TP			
	chr13	32333266	T	C	S	exon10	c.1788T>C	p.D596D	.	0.0003	0.0163	Likely Benign	2	1CCBa	AA	Het	99	1115	51%	Yes	TP			
														1EBa	AA	Het	99	347	50%	Yes	TP			
	chr13	32333276	T	C	NS	exon10	c.1798T>C	p.Y600H	B	.	0.0055	VUS	1	1CAGa	AA	Het	99	72	42%	No	TP			
	chr13	32333395	TG	-	Intronic	intron10	c.1919+8delTG	N/A	.	0.0005	0.0059	VUS	1	1CAGa	AA	Het	99	19	58%	No	TP			
	chr13	32339875	A	G	NS	exon11	c.5020A>G	pS1674G	B	.	.	VUS	2	1CADa	AA	Het	99	344	61%	Yes	TP			
														1CADf	AA	Het	99	1294	47%	Yes	TP			

	chr13	32339554	C	T	S	exon11	c.5199C>T	pS1733S	.	0.0054	0.0009	Likely Benign	1	1CBEa	EA	Het	99	1398	47%	Yes	TP			
	chr13	32339966_32339970	AGTAA	-	FSD	exon11	c.5611_5615delAGTAA[s]	pS1871fs[s]	.	.	.	Pathogenic	1	1CBa	AA	Het	99	279	57%	Yes	TP			
	chr13	32340678	G	A	NS	exon11	c.6323G>A	pR2108H	B	0.0015	0.0068	VUS	1	1CAHa	AA	Hom	99	1,090	100%	Yes	TP			
	chr13	32357750	G	A	S	exon16	c.7626G>A	p.T2542T	.	.	0.0061	Likely Benign	1	1EAHa	AA	Het	99	36	33%	No	TP			
	chr13	32363385	T	C	NS	exon18	c.8183T>C	pV2728A	P	.	.	VUS	1	1CADf	AA	Het	99	1001	52%	Yes	TP			
	chr13	32371035	A	C	NS	exon20	c.8567A>C	pE2856A	D	0.002	0.0005	VUS	1	1CABa	EA	Het	99	47	53%	No	TP			
CDH1 (NM_004360)	chr16	68801830	A	G	S	exon3	c.324A>G	pR108R	.	0.0003	0.0061	Likely Benign	2	1CAFa	AA	Het	99	1,210	49%	Yes	TP	75%	0.25	0.00
														1CBHa	AA	Het	99	120	48%	Yes	TP			
	chr16	68813324	G	T	NS	exon9	c.1149G>T	pQ383H	B	.	.	.	1	1EEa	EA	Het	99	20	20%	No	FP			
	chr16	68819894	G	C	S	exon11	c.1680G>C	p.T560T	.	0.0029	0.0007	Likely Benign	1	1CAIa	AA	Het	99	3,053	53%	Yes	TP			
CHEK2 (NM_001005735)	chr22	28695232	A	G	NS	exon13	c.1399T>C	pY467H	D	0.0003	.	VUS	1	1EAa	EA	Het	99	567	47%	Yes	TP	100%	0.00	0.00
NBN (NM_002485)	chr8	89955458	T	C	NS	exon10	c.1222A>G	pK408E	P	.	0.0082	Likely Benign	1	1CGa	AA	Het	99	1341	53%	Yes	TP	100%	0.00	0.00
	chr8	89980833	A	G	S	exon4	c.381T>C	pA127A	.	0.0045	0.0018	Likely Benign	1	1EAJa	EA	Het	99	164	45%	Yes	TP			
	chr8	89984520	C	T	Intronic	intron1	c.37+5G>A	N/A	.	0.001	0.0182	Likely Benign	2	1CAHa	AA	Het	99	177	54%	Yes	TP			
														1CGa	AA	Het	99	109	53%	Yes	TP			
PALB2 (NM_024675)	chr16	23629888	T	C	S	exon5	c.2256A>G	pG752G	.	.	0.0055	Likely Benign	1	1CCBa	AA	Het	99	1152	51%	Yes	TP	100%	0.00	0.00
	chr16	23635127	T	G	S	exon4	c.1419A>C	pP473P	.	.	0.0084	Likely Benign	1	1CAIa	AA	Het	99	955	49%	Yes	TP			
	chr16	23635536	A	G	NS	exon4	c.1010T>C	pL337S	B	0.0197	0.0036	VUS	1	1RGa	EA	Het	99	281	52%	Yes	TP			
	chr16	23638125	T	C	NS	exon2	c.53A>G	pK18R	D	.	0.0155	VUS	2	1CGa	AA	Het	99	276	49%	Yes	TP			
														1EAGa	AA	Het	99	111	42%	Yes	TP			
PTEN (NM_000314)	chr10	87931070	C	T	S	exon4	c.234C>T	p.T78T	.	.	0.0002	Likely Benign	1	1CAa	EA	Het	99	339	53%	Yes	TP	83.33%	0.92	0.92
	chr10	87960892	A	T	Intronic	intron7	c.802-2A>T	N/A	.	.	.	Pathogenic	4	1CAFa	AA	Het	91	27	30%	No	FP			
														1CDa	AA	Het	99	45	29%	No	FP			
														1CFa	AA	Het	96	41	46%	No	FP			
														1EAJa	EA	Het	86	12	42%	No	FP			
	chr10	87960895	C	A	NS	exon8	c.804C>A	pD268E	B	.	.	Not provided	3	1CGa	AA	Het	99	32	34%	No	FP			
														1EEa	EA	Het	83	34	24%	No	FP			
														3CGa	AA	Het	99	63	27%	No	FP			
	chr10	87960902	G	T	NS	exon8	c.810G>T	pM270I	P	.	.	.	3	1CABa	EA	Het	90	19	32%	No	FP			
														1CAFa	AA	Het	90	25	32%	No	FP			
														1CGa	AA	Het	41	34	68%	No	FP			
	chr10	87960906	C	T	NS	exon8	c.814C>T	pH272Y	D	.	.	.	1	1CGa	AA	Het	89	31	29%	No	FP			
STK11 (NM_000455)	chr19	1218495	G	A	S	exon2	c.369G>A	pQ123Q	.	0.0002	0.0154	Likely Benign	5	1CAGa	AA	Het	99	107	46%	Yes	TP	100%	0.00	0.00
														1CBHa	AA	Het	99	81	57%	No	TP			
														1CEa	AA	Het	99	88	49%	No	TP			
														1EBAa	AA	Het	99	849	51%	Yes	TP			
														1EBa	AA	Het	99	904	49%	Yes	TP			
TP53 (NM_000546)	chr17	7670613	A	C	NS	exon10	c.1096T>G	pS366A	B	.	.	VUS	1	1EBAa	AA	Het	99	863	49%	Yes	TP	100%	0.00	0.00
	chr17	7673776	G	A	NS	exon8	c.844C>T	pR282W	D	0.0002	.	Pathogenic	1	1RCa	AA	Het	99	240	28%	No	TP			

Table 3.3 Key: (Chr) Chromosome; (Ref.) Reference; (Alt.) Alternate; (NS) Nonsynonymous; (S) Synonymous; (FSD) Frame-shift deletion; (EA) European American; (AA) African American; (~) Polyphen2 HDIV prediction; (B) Benign; (P) Possibly damaging; (D) Damaging; (#) esp6500siv2; (^) most severe clinical significance classification; (GT) Genotype; (GQ) Genotype quality; (Het) Heterozygous; (Homo) Homozygous; ($) The deletion was named using ANNOVAR (v.); however, it is within a short tandem repeat and commonly referred to as *BRCA2* c.5616_5620del5 (p.K1872Nfs) since Human Genome Variant Society (HGVS) nomenclature rules state to arbitrarily assign the deletion to the most 3' nucleotide.

3.9 Figures

Figure 3.1: Screening process for individuals in the AHCC. *In situations where blood is unattainable, another set of three distinct biological samples (i.e. saliva) is collected.

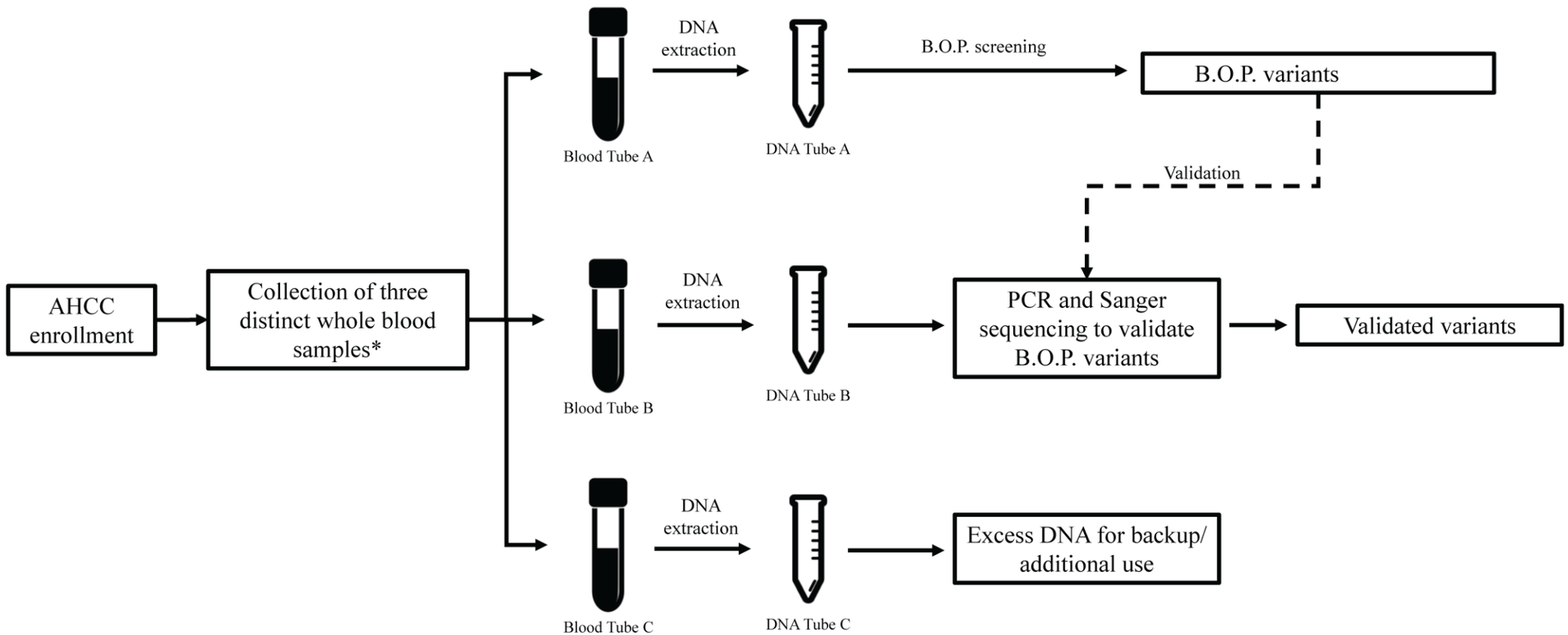


Figure 3.2: Pipeline for B.O.P. panel screening: targeted capture, NGS, bioinformatics pipeline, and variant filtering.

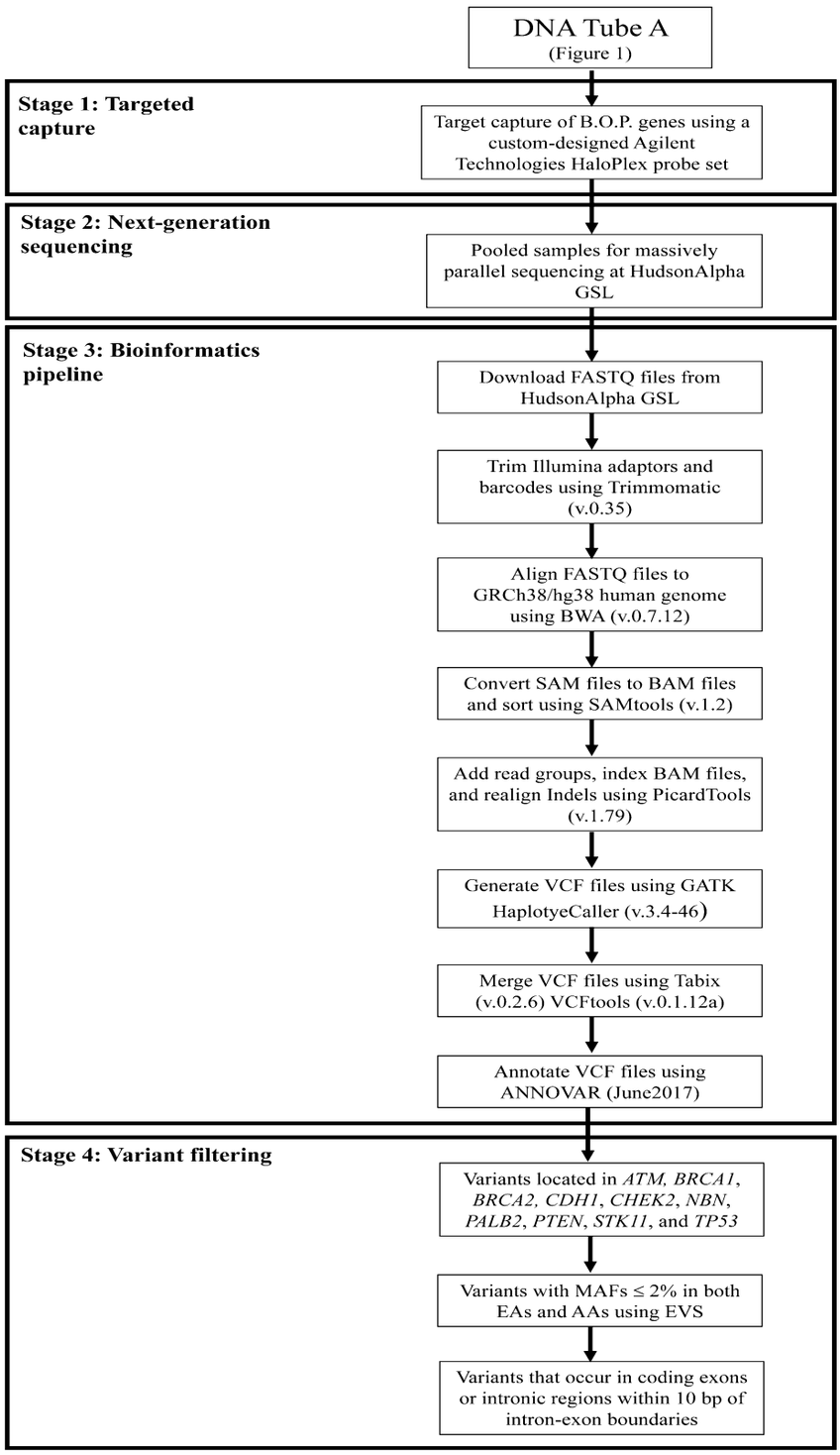
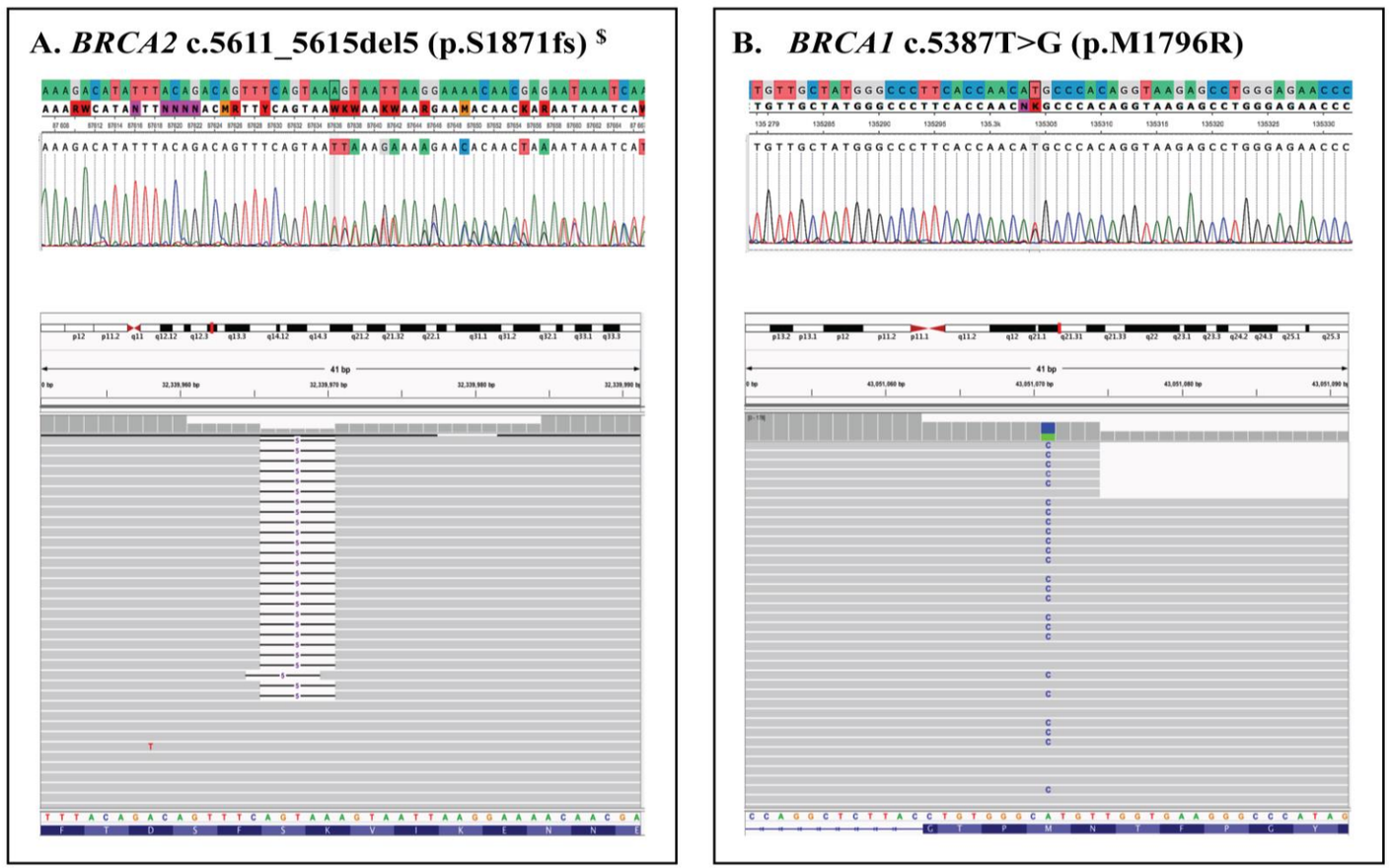


Figure 3.3: B.O.P. positive controls: mutations that were previously reported through clinical gene screening. The Sanger sequence electropherogram (above) and the Integrative Genomics Viewer (v.2.4.5) image (below) are depicted for two B.O.P. screened positive controls: 1CB-a (Panel A) and 1CAD-a (Panel B). ($) This deletion is within a short tandem repeat and is commonly referred to as BRCA2 c.5616_5620del5 (p.K1872Nfs) since HGVS nomenclature rules arbitrarily assign.



3.10 Supplementary Material

Supplementary Table 3.1:

Sample	Genome			B.O.P. target regions			Calculations		
	QC passed reads	%Mapped Reads	%Properly Paired Reads	QC passed reads	%Mapped Reads	%Properly Paired Reads	Total number of reads mapped to genome	Total number of reads mapped to target region	% of total mapped reads on target
1CAA-a	3466931	99.45%	98.92%	2244003	100.00%	99.64%	3466931	2244003	64.73%
1CA-a	7358715	97.60%	96.49%	2950926	100.00%	99.70%	7358715	2950926	40.10%
1CAB-a	4693516	99.02%	98.33%	2562564	100.00%	99.63%	4693516	2562564	54.60%
1CAC-a	1076138	99.25%	98.70%	602854	100.00%	99.72%	1076138	602854	56.02%
1CAD-a	4999317	98.31%	97.57%	2487768	100.00%	99.57%	4999317	2487768	49.76%
1CAD-f	17380472	98.66%	97.98%	10043275	100.00%	99.45%	17380472	10043275	57.78%
1CAE-a	2104120	98.74%	98.02%	950271	100.00%	99.63%	2104120	950271	45.16%
1CAF-a	7481433	99.06%	98.37%	3899157	100.00%	99.60%	7481433	3899157	52.12%
1CAG-a	926011	97.14%	96.35%	463059	100.00%	99.44%	926011	463059	50.01%
1CAH-a	13972990	99.15%	98.49%	6790535	100.00%	99.68%	13972990	6790535	48.60%
1CAI-a	13941582	99.02%	98.44%	8162274	100.00%	99.65%	13941582	8162274	58.55%
1CB-a	8651447	98.85%	98.17%	3590191	100.00%	99.68%	8651447	3590191	41.50%
1CBC-a	10756010	99.04%	98.36%	5430742	100.00%	99.66%	10756010	5430742	50.49%
1CBE-a	11413750	98.56%	97.99%	7262357	100.00%	99.61%	11413750	7262357	63.63%
1CBH-a	1053204	97.66%	97.04%	498489	100.00%	99.50%	1053204	498489	47.33%
1CC-a	6904994	98.96%	98.42%	3982834	100.00%	99.65%	6904994	3982834	57.68%
1CCB-a	12192260	98.31%	97.61%	6187971	100.00%	99.56%	12192260	6187971	50.75%
1CD-a	12964959	99.25%	98.67%	7526111	100.00%	99.64%	12964959	7526111	58.05%
1CE-a	849962	99.13%	98.28%	401321	100.00%	99.65%	849962	401321	47.22%
1CF-a	11840537	98.67%	97.95%	5659114	100.00%	99.65%	11840537	5659114	47.79%

1CG-a	15140542	99.03%	98.45%	8680038	100.00%	99.64%	15140542	8680038	57.33%
1CH-a	10686868	98.68%	98.07%	6218516	100.00%	99.58%	10686868	6218516	58.19%
1EAJ-a	3615706	99.06%	98.43%	1738139	100.00%	99.66%	3615706	1738139	48.07%
1EAA-a	15854996	99.00%	98.25%	8132822	100.00%	99.53%	15854996	8132822	51.30%
1EA-a	10171087	99.03%	98.39%	4533599	100.00%	99.70%	10171087	4533599	44.57%
1EAC-a	13389945	98.75%	98.06%	7270723	100.00%	99.67%	13389945	7270723	54.30%
1EAD-a	1765462	97.87%	97.18%	1016065	100.00%	99.57%	1765462	1016065	57.55%
1EAE-a	5042685	98.95%	98.15%	2516319	100.00%	99.56%	5042685	2516319	49.90%
1EAF-a	5964662	93.39%	92.58%	2886393	100.00%	99.23%	5964662	2886393	48.39%
1EAG-a	2329107	98.84%	98.07%	1178518	100.00%	99.53%	2329107	1178518	50.60%
1EAH-a	1133042	98.45%	97.85%	570220	100.00%	99.63%	1133042	570220	50.33%
1EAI-a	15427715	98.32%	97.64%	8208697	100.00%	99.60%	15427715	8208697	53.21%
1EBA-a	9564055	98.84%	98.24%	4798432	100.00%	99.65%	9564055	4798432	50.17%
1EB-a	8172055	99.23%	98.63%	3894157	100.00%	99.69%	8172055	3894157	47.65%
1EC-a	7612387	98.83%	97.95%	2810482	100.00%	99.65%	7612387	2810482	36.92%
1ED-a	8215227	97.25%	96.55%	3961643	100.00%	99.65%	8215227	3961643	48.22%
1EE-a	10931675	99.07%	98.38%	6153136	100.00%	99.69%	10931675	6153136	56.29%
1EF-a	124980500	99.10%	98.19%	60356533	100.00%	99.69%	124980500	60356533	48.29%
1EG-a	5630466	98.69%	98.07%	2945296	100.00%	99.62%	5630466	2945296	52.31%
1EH-a	8344391	98.34%	97.37%	2949704	100.00%	99.70%	8344391	2949704	35.35%
1EI-a	1852922	98.92%	98.15%	732455	100.00%	99.62%	1852922	732455	39.53%
3CC-a	31874003	98.32%	97.27%	16836337	100.00%	99.54%	31874003	16836337	52.82%
4CA-a	14014903	98.58%	97.84%	7824383	100.00%	99.45%	14014903	7824383	55.83%
Average	11296343	98.57%	97.86%	5765312	100.00%	99.61%	11296343	5765312	50.91%

Supplementary Table 3.1 key: (%) Percent.

Supplementary Table 3.2:

Target Region	Total Average of 43 Individuals Screened	Avg % at 1X	Avg % at 10X	Avg % at 20X	Avg % at 50X	Avg % at 100X	Avg % at 250X	Avg % at 500X	Avg % at 1000X	Avg % at 10000X
chr10:87863103-87864558	300	95.86	87.19	81.30	68.64	56.03	33.77	15.32	3.83	0.04
chr10:87866663-87867059	859	98.63	97.29	96.83	93.48	84.68	69.96	48.97	23.94	0.60
chr10:87894015-87894336	329	100.00	97.02	91.97	80.77	67.62	38.26	14.40	3.85	0.03
chr10:87925503-87925567	429	100.00	99.78	96.31	86.44	77.81	48.83	25.29	3.54	0.00
chr10:87931036-87931099	419	100.00	98.47	97.42	86.34	69.88	48.70	25.88	3.60	0.00
chr10:87933003-87933497	773	100.00	99.41	98.01	91.70	82.32	63.17	40.58	18.40	0.52
chr10:87952108-87952269	1080	100.00	100.00	99.94	97.10	90.11	78.72	59.79	36.98	1.88
chr10:87957843-87958029	911	100.00	99.70	99.05	95.69	87.58	73.41	54.17	25.83	1.55
chr10:87960884-87961319*	427	91.98	82.57	78.20	64.71	51.20	30.55	22.66	11.28	0.33
chr10:87965277-87971940	640	96.77	93.86	90.94	82.71	72.52	53.06	33.27	15.45	0.56
chr11:108222822-108223196	678	100.00	99.80	98.88	93.27	82.95	64.16	41.62	13.39	0.76
chr11:108224774-108224881	716	100.00	100.00	98.77	94.19	83.08	65.68	41.48	18.68	0.30
chr11:108227585-108227706	1499	100.00	99.93	99.78	98.08	91.99	80.69	64.79	41.01	1.60
chr11:108227766-108227898	564	81.95	80.70	78.92	72.50	64.74	46.09	31.42	15.97	0.40
chr11:108229168-108229333	541	100.00	100.00	99.09	89.49	80.47	57.39	31.67	10.81	0.20
chr11:108235660-108235844	1142	100.00	98.21	91.98	85.80	76.43	53.83	43.70	31.06	1.17
chr11:108243943-108244128	915	100.00	100.00	99.72	97.58	89.84	72.65	52.35	28.32	0.90
chr11:108244778-108245036	1162	100.00	99.42	98.74	95.86	89.90	73.09	58.31	35.17	1.13

chr11:108246954-108247137	691	100.00	100.00	98.73	92.48	82.90	64.36	38.96	17.54	0.93
chr11:108248923-108249112	314	99.66	97.53	91.64	81.13	70.43	43.45	13.36	2.26	0.00
chr11:108250691-108251082	952	100.00	100.00	99.04	94.70	87.67	69.90	51.05	26.95	1.39
chr11:108251827-108252041	1096	99.68	99.16	98.93	97.23	90.25	74.66	59.85	35.70	1.66
chr11:108252807-108252922	361	100.00	97.31	92.90	85.13	67.76	35.93	17.03	6.45	0.00
chr11:108253804-108254049	908	100.00	99.81	99.25	96.66	86.50	73.30	52.53	24.72	0.86
chr11:108256205-108256350	1491	100.00	100.00	99.78	97.40	92.96	80.04	64.47	43.23	1.56
chr11:108257471-108257616	580	100.00	98.79	97.34	90.94	79.43	55.15	34.34	13.58	0.56
chr11:108258976-108259085	771	100.00	99.30	98.33	92.35	84.20	65.05	45.31	18.55	0.38
chr11:108267161-108267352	863	100.00	100.00	99.49	96.15	88.69	72.69	51.12	22.87	1.08
chr11:108268400-108268619	1354	99.28	99.16	98.96	97.21	93.25	77.28	64.29	39.45	1.96
chr11:108271054-108271156	1182	100.00	99.05	96.72	93.13	87.82	68.11	52.79	31.92	1.47
chr11:108271241-108271416	1219	100.00	100.00	99.50	96.37	90.51	76.05	60.10	34.68	1.43
chr11:108272522-108272617	1049	97.85	95.92	95.75	94.60	85.32	73.15	57.48	27.41	1.43
chr11:108272712-108272862	442	100.00	99.48	96.92	87.43	77.33	54.87	23.31	5.88	0.00
chr11:108279481-108279618	775	100.00	100.00	100.00	96.95	86.33	72.32	46.90	19.27	0.84

chr11:108280985-108281178	1028	100.00	100.00	98.63	94.53	88.05	71.08	53.33	27.23	1.20
chr11:108282700-108282889	528	100.00	98.60	95.34	84.16	72.23	50.58	32.57	12.88	0.00
chr11:108284217-108284483	400	100.00	97.95	95.91	83.84	68.89	48.39	22.90	4.90	0.00
chr11:108287588-108287725	857	100.00	100.00	99.82	97.33	88.62	72.78	49.20	25.36	0.79
chr11:108288967-108289113	849	96.00	95.09	93.43	85.87	77.83	59.23	44.34	25.38	0.82
chr11:108289592-108289811	238	99.88	95.21	87.10	74.37	57.35	23.90	8.98	3.38	0.00
chr11:108292609-108292803	223	99.41	96.96	90.23	73.76	56.86	21.00	5.21	2.42	0.00
chr11:108293303-108293487	411	100.00	100.00	99.29	88.32	75.83	48.70	22.56	4.33	0.00
chr11:108294917-108296002	794	85.41	83.87	83.09	78.71	73.23	58.94	40.51	21.09	0.72
chr11:108297277-108297392	521	100.00	98.36	96.46	88.72	71.96	53.12	30.73	7.86	0.00
chr11:108299704-108299895	410	99.43	97.10	91.84	80.94	68.67	44.83	25.17	4.54	0.00
chr11:108301638-108301799	876	99.62	98.74	98.02	93.16	84.29	66.68	47.72	25.82	1.31
chr11:108302843-108303039	739	100.00	100.00	99.85	95.95	85.87	70.91	43.64	20.12	0.24
chr11:108304665-108304862	401	100.00	99.04	94.70	82.22	68.19	41.88	22.71	8.37	0.13
chr11:108307887-108307994	796	100.00	98.80	98.15	95.26	85.85	70.25	42.36	24.40	0.17
chr11:108310150-108310325	1175	100.00	99.46	98.37	94.30	88.54	71.23	54.00	32.11	1.35

chr11:108312401-108312508	1251	100.00	100.00	100.00	99.50	94.06	79.16	64.99	39.78	2.33
chr11:108315813-108315921	1189	100.00	100.00	100.00	99.04	95.84	79.67	67.87	38.47	1.45
chr11:108316001-108316123	463	100.00	100.00	97.15	87.24	78.43	55.52	25.19	4.99	0.00
chr11:108317363-108317531	1151	100.00	99.46	98.93	97.03	90.68	72.58	58.37	35.31	1.23
chr11:108319944-108320068	655	100.00	100.00	99.24	93.13	84.65	68.32	40.80	14.68	0.32
chr11:108321291-108321430	1416	100.00	100.00	99.75	99.12	95.05	81.66	66.26	44.17	1.38
chr11:108325300-108325554	426	98.37	94.87	89.72	78.17	62.55	39.16	23.69	10.17	0.00
chr11:108326048-108326235	507	100.00	99.29	97.65	89.74	77.27	56.72	29.62	7.07	0.07
chr11:108327635-108327768	820	100.00	100.00	99.48	96.17	86.97	72.04	49.31	26.57	0.49
chr11:108329011-108329248	1206	100.00	100.00	99.99	96.63	90.10	76.53	59.26	35.64	1.40
chr11:108330204-108330431	767	100.00	99.27	97.90	93.07	84.65	65.43	45.52	19.75	1.03
chr11:108331434-108331567	1651	100.00	100.00	100.00	99.38	97.74	83.39	71.35	50.26	1.93
chr11:108331869-108332047	1298	100.00	100.00	99.92	98.44	92.95	80.73	61.79	37.90	1.25
chr11:108332752-108332910	821	99.93	99.71	99.08	96.52	85.90	69.86	49.15	21.41	1.07
chr11:108333876-108333978	214	89.88	86.99	81.88	71.31	54.84	22.03	5.40	2.12	0.00
chr11:108334959-108335119	884	100.00	99.39	97.17	93.99	83.03	64.32	46.63	29.22	0.81

chr11:108335835-108335971	2053	100.00	100.00	99.44	98.30	95.97	85.79	71.19	56.14	3.16
chr11:108343212-108343381	486	100.00	96.00	92.30	81.94	66.77	40.43	19.47	9.58	0.55
chr11:108345733-108345918	791	100.00	99.46	97.69	94.01	84.94	67.49	44.30	21.51	0.93
chr11:108346511-108346602	233	100.00	96.97	88.35	73.84	59.26	29.96	7.72	2.10	0.00
chr11:108347269-108347375	68	55.87	46.03	35.64	25.95	19.05	10.43	1.52	0.76	0.00
chr11:108353756-108353890	1141	100.00	100.00	100.00	99.85	91.04	79.93	62.59	40.56	1.88
chr11:108354801-108354884	1138	100.00	100.00	99.58	97.06	91.55	73.82	58.51	35.70	1.47
chr11:108365072-108365228	849	100.00	100.00	99.33	95.95	87.66	72.04	50.09	26.27	0.92
chr11:108365315-108369112	666	96.52	92.11	88.90	81.01	71.90	54.46	36.83	17.45	0.64
chr13:32315464-32315677	699	100.00	99.84	99.17	93.93	85.12	62.99	39.51	17.52	0.43
chr13:32316412-32316537	1228	100.00	100.00	99.76	98.84	94.59	80.30	66.87	42.09	1.79
chr13:32319067-32319335	1611	100.00	100.00	100.00	99.33	96.04	84.02	71.54	47.53	2.20
chr13:32325066-32325194	382	99.78	98.17	92.02	85.28	69.21	42.09	15.73	6.34	0.18
chr13:32326091-32326160	609	100.00	100.00	100.00	94.85	83.72	66.94	36.41	8.84	0.00
chr13:32326232-32326292	1327	100.00	100.00	100.00	100.00	93.02	83.72	74.95	42.89	2.33
chr13:32326489-32326623	824	100.00	99.83	98.04	93.35	86.08	64.48	45.99	24.35	0.67
chr13:32329433-32329502	517	100.00	99.37	98.60	89.07	77.74	59.30	33.79	5.55	0.00
chr13:32330909-32331040	1219	100.00	100.00	100.00	98.34	94.53	78.95	66.99	37.60	1.80
chr13:32332262-32333397	809	99.96	99.76	99.09	93.91	85.12	67.95	46.65	21.36	1.00
chr13:32336255-32341206	1137	97.22	96.74	95.96	92.58	85.63	70.64	53.49	30.92	1.37
chr13:32344548-32344663	786	100.00	100.00	100.00	94.83	89.03	75.62	46.22	20.12	0.08
chr13:32346817-32346906	966	100.00	99.90	99.74	96.12	89.58	76.43	51.68	30.84	1.50

chr13:32354851-32355298	734	99.85	97.17	94.66	87.13	77.01	60.72	43.58	19.65	0.61
chr13:32356418-32356619	1264	100.00	100.00	99.53	94.86	89.07	72.85	52.61	30.33	1.07
chr13:32357732-32357939	1635	100.00	100.00	100.00	98.22	94.92	82.01	67.13	44.61	1.73
chr13:32362513-32362703	621	100.00	97.61	93.12	86.25	73.69	47.35	34.85	16.31	0.84
chr13:32363169-32363543	1244	100.00	99.88	99.82	98.06	91.94	78.01	63.97	38.43	1.79
chr13:32370392-32370567	600	100.00	100.00	99.37	93.66	83.88	63.77	36.65	12.28	0.00
chr13:32370946-32371110	867	100.00	99.79	99.37	95.71	86.86	72.35	52.97	24.71	0.93
chr13:32376660-32376801	941	100.00	100.00	100.00	97.60	90.72	74.77	55.08	23.82	1.90
chr13:32379307-32379525	958	100.00	99.24	97.62	92.95	84.73	66.40	49.04	28.88	1.32
chr13:32379740-32379923	1281	100.00	99.39	98.58	94.23	87.73	71.44	57.91	35.93	1.49
chr13:32379997-32380155	164	98.13	85.72	76.05	59.34	42.42	18.80	4.08	1.27	0.00
chr13:32394679-32394943	530	96.70	95.84	93.31	85.95	78.30	56.84	31.88	7.18	0.00
chr13:32395981-32396126	363	99.86	96.97	94.21	80.88	68.71	44.42	20.88	4.20	0.00
chr13:32396888-32397054	543	100.00	98.00	95.65	86.93	76.55	55.41	32.65	9.93	0.31
chr13:32398152-32399682	650	99.23	97.29	94.70	86.83	75.67	54.79	35.32	16.09	0.47
chr16:23603152-23603679	733	100.00	99.19	97.75	91.38	82.83	64.63	41.52	19.47	0.85
chr16:23607854-23608022	653	98.16	96.86	95.27	90.36	79.81	60.41	38.67	16.23	0.69
chr16:23613994-23614101	1573	100.00	100.00	100.00	100.00	97.55	84.84	73.86	49.37	2.00
chr16:23621352-23621488	1097	100.00	99.83	99.83	98.00	91.40	77.81	61.84	35.36	1.39
chr16:23622959-23623140	922	100.00	99.62	98.67	96.64	90.60	73.04	54.04	28.84	0.92
chr16:23623999-23624104	465	100.00	98.93	97.81	88.37	78.43	53.67	30.79	6.80	0.00
chr16:23626226-23626407	540	99.30	97.19	92.98	86.43	72.26	51.83	34.05	15.42	0.00
chr16:23629194-23629285	723	100.00	100.00	100.00	95.55	86.30	68.63	47.20	15.29	0.66
chr16:23629630-23630479	1372	100.00	99.99	99.89	98.92	94.50	80.65	65.80	41.56	1.56
chr16:23634852-23636344	1075	99.86	98.34	97.02	92.37	85.58	70.41	52.86	30.10	1.15
chr16:23637840-23637962	818	100.00	100.00	99.62	97.22	85.97	70.11	48.53	21.65	1.13
chr16:23638060-23638139	753	100.00	100.00	100.00	94.80	86.72	71.40	46.15	18.59	0.81
chr16:23641100-23641367	707	100.00	99.73	99.47	96.60	85.10	71.47	47.79	17.08	0.36

chr16:68737215-68737473	589	100.00	99.87	98.55	92.85	82.54	63.83	37.74	10.01	0.09
chr16:68738287-68738421	610	100.00	97.71	93.14	85.54	73.67	42.44	29.39	17.25	0.77
chr16:68801660-68801903	1137	100.00	100.00	99.85	98.06	92.48	79.33	64.35	36.31	1.48
chr16:68808414-68808577	435	99.79	98.09	95.54	87.36	73.23	52.22	25.17	6.39	0.00
chr16:68808683-68808858	1206	100.00	100.00	99.80	98.79	92.09	79.02	66.62	39.49	2.05
chr16:68810187-68810351	1449	100.00	100.00	100.00	99.00	95.45	82.01	69.62	46.23	2.02
chr16:68811674-68811869	1956	100.00	100.00	99.98	99.93	98.59	87.32	77.09	61.38	2.04
chr16:68812125-68812273	722	100.00	98.27	94.33	86.25	80.35	50.70	34.10	20.85	0.87
chr16:68813303-68813572*	642	100.00	97.28	94.17	84.13	74.24	55.43	38.14	16.41	0.68
chr16:68815505-68815769	775	100.00	98.47	97.68	91.33	81.46	66.12	44.71	18.29	0.91
chr16:68819270-68819435	1729	100.00	99.71	99.53	98.30	96.95	86.19	71.58	51.76	1.51
chr16:68821991-68822235	555	91.54	91.40	90.68	87.39	77.53	60.54	33.79	13.30	0.00
chr16:68823389-68823636	1572	100.00	100.00	99.96	99.60	96.55	84.29	70.80	49.74	1.93
chr16:68828164-68828314	854	98.47	94.23	92.89	87.93	80.98	62.18	41.56	21.89	0.48
chr16:68829644-68829807	613	100.00	99.89	99.78	92.69	83.36	63.18	38.58	10.82	0.28
chr16:68833280-68835551	868	97.51	95.63	93.38	88.87	80.76	63.71	46.32	25.07	1.00
chr17:43044285-43045812	629	99.14	96.93	93.48	83.33	70.19	47.41	29.40	15.14	0.49
chr17:43047633-43047713	690	100.00	100.00	100.00	95.92	84.35	67.47	43.26	15.07	0.77
chr17:43049111-43049204	577	100.00	99.97	98.14	92.13	79.80	64.79	31.57	8.56	0.25
chr17:43051053-43051127	310	100.00	98.60	94.57	83.31	68.49	37.89	14.11	2.33	0.00
chr17:43057042-43057145	1466	100.00	100.00	100.00	100.00	97.96	82.94	76.25	46.84	2.33
chr17:43063323-43063383	557	100.00	96.00	93.13	89.33	68.01	53.33	33.17	12.51	0.00
chr17:43063864-43063961	1539	100.00	100.00	100.00	100.00	95.85	81.92	73.85	43.62	2.33
chr17:43067598-43067705	869	100.00	100.00	100.00	96.28	87.92	75.34	58.51	24.38	1.47
chr17:43070918-43071248	1150	97.08	96.57	96.57	91.09	85.46	69.31	48.99	31.20	1.20
chr17:43074321-43074531	1409	100.00	99.76	99.55	99.28	95.19	81.85	70.17	44.02	1.67
chr17:43076478-43076621	716	100.00	99.73	98.61	91.64	82.69	66.19	42.60	16.79	0.68
chr17:43079324-43079409	988	100.00	100.00	100.00	98.62	90.10	76.88	55.64	31.26	1.03

chr17:43082394-43082585	264	78.28	77.56	76.41	64.88	58.07	36.03	11.60	2.39	0.00
chr17:43090934-43091042	943	100.00	100.00	100.00	95.99	86.93	74.70	58.08	25.46	1.30
chr17:43091425-43094870*	1265	97.79	96.93	95.94	92.62	85.62	71.49	55.05	34.69	1.48
chr17:43095836-43095932	1553	100.00	100.00	100.00	100.00	97.33	82.88	74.70	50.73	2.33
chr17:43097234-43097299	553	100.00	100.00	98.41	91.05	82.06	61.31	33.23	10.04	0.00
chr17:43099765-43099890	2053	100.00	100.00	100.00	99.63	97.47	88.70	77.61	60.18	2.33
chr17:43104112-43104271	370	97.25	93.15	87.41	77.58	64.96	41.43	17.89	6.08	0.17
chr17:43104858-43104966	655	100.00	100.00	98.95	92.21	84.60	63.81	41.41	11.56	0.28
chr17:43106446-43106543	827	100.00	99.67	98.67	93.59	83.31	66.51	51.99	22.50	0.85
chr17:43115716-43115789	499	100.00	100.00	96.98	89.63	79.67	57.79	30.10	5.81	0.00
chr17:43124007-43124125	645	100.00	99.20	96.81	91.85	79.67	63.00	40.73	12.53	0.00
chr17:43125261-43125493	892	100.00	100.00	99.79	98.00	88.96	74.39	54.83	24.48	0.92
chr17:7661769-7662024	499	84.77	81.53	80.08	74.30	66.55	47.76	26.20	13.42	0.05
chr17:7666076-7666254	333	100.00	97.82	93.42	82.96	63.59	38.35	17.28	6.77	0.00
chr17:7668392-7669700	842	100.00	99.26	97.39	92.15	83.43	64.16	45.89	22.90	0.84
chr17:7670599-7670725	1186	100.00	100.00	100.00	97.93	92.09	76.88	63.75	36.84	1.80
chr17:7673197-7673349	767	100.00	97.74	95.49	88.04	76.83	53.72	43.20	24.31	1.14
chr17:7673525-7673618	736	100.00	99.26	99.26	95.96	84.86	68.00	44.25	17.81	1.06
chr17:7673691-7673847	493	100.00	100.00	98.62	91.85	82.27	56.99	29.37	6.58	0.00
chr17:7674171-7674300	1703	100.00	100.00	99.95	99.50	97.91	86.71	73.29	53.57	2.43
chr17:7674516-7674981	602	99.96	96.83	94.41	85.12	73.60	55.13	36.25	13.78	0.73
chr17:7675043-7675503	1054	100.00	99.95	99.70	96.38	88.98	73.65	51.69	27.85	0.82
chr17:7675984-7676282	776	92.84	77.30	72.24	67.26	55.71	40.10	33.61	22.63	0.86
chr17:7676372-7676632	407	81.15	69.20	64.32	59.20	48.62	29.78	20.50	11.03	0.39
chr17:7677315-7677444	1261	100.00	100.00	100.00	100.00	95.69	83.51	69.59	45.08	2.31
chr17:7687367-7687560	654	100.00	100.00	98.60	94.75	82.60	63.26	39.75	14.86	0.50
chr19:1205789-1207213	303	84.79	76.79	72.13	60.78	49.99	28.17	13.52	6.07	0.10
chr19:1218407-1218510	999	100.00	100.00	100.00	100.00	93.22	80.48	61.67	32.33	0.65

chr19:1219314-1219423	802	99.49	93.18	89.61	82.02	75.47	62.91	50.64	24.60	0.78
chr19:1220363-1220515	882	100.00	99.62	99.57	97.86	89.89	77.16	58.02	23.97	1.02
chr19:1220571-1220727	729	100.00	100.00	99.58	91.97	83.47	65.68	38.77	13.98	0.83
chr19:1221203-1221350	490	100.00	99.42	97.85	91.00	78.77	58.45	28.45	4.30	0.00
chr19:1221939-1222016	596	100.00	100.00	100.00	92.73	83.51	71.23	40.13	8.08	0.00
chr19:1222975-1223182	1065	100.00	100.00	100.00	99.03	92.56	79.03	63.84	30.08	1.58
chr19:1226444-1226673	549	100.00	99.92	96.94	90.19	81.56	56.41	32.13	12.26	0.05
chr19:1227583-1228445	480	97.62	95.21	91.86	83.56	72.19	46.80	28.09	10.42	0.20
chr22:28687733-28687996	395	95.20	90.10	86.37	75.80	65.08	43.16	23.66	6.27	0.00
chr22:28689125-28689420	1226	100.00	100.00	99.38	94.56	88.46	71.02	51.10	32.41	1.31
chr22:28694022-28694127	605	99.10	96.62	94.28	85.63	78.06	60.76	40.11	13.32	0.13
chr22:28695117-28695252	2041	100.00	99.64	99.41	97.94	91.03	84.05	67.80	48.36	2.31
chr22:28695700-28695883	676	100.00	99.37	98.30	92.44	82.04	66.26	43.99	16.86	0.00
chr22:28696888-28696997	814	100.00	100.00	100.00	96.64	87.23	71.98	49.52	20.48	0.95
chr22:28699754-28699947	692	100.00	100.00	99.24	91.83	82.31	66.60	43.17	15.70	0.80
chr22:28703495-28703576	576	100.00	99.15	94.27	87.60	77.14	55.34	26.16	12.00	0.40
chr22:28709996-28710069	440	100.00	97.49	93.81	85.36	78.82	48.28	18.42	5.56	0.00
chr22:28711899-28712267	997	100.00	100.00	99.89	97.21	89.88	75.00	58.01	28.90	1.36
chr22:28719385-28719495	831	100.00	97.66	96.19	89.80	81.01	65.57	45.81	24.79	1.34
chr22:28721508-28721641	111	98.68	90.84	82.07	58.93	34.83	5.98	2.33	1.01	0.00
chr22:28724555-28724933	225	73.36	69.33	64.14	55.84	42.25	23.95	10.61	4.52	0.00
chr22:28724955-28725134	803	100.00	99.81	97.65	91.07	83.39	61.83	43.15	20.24	1.05
chr22:28725233-28725377	888	100.00	100.00	100.00	95.94	88.33	73.60	54.60	26.98	1.91
chr22:28726235-28726371	641	100.00	99.05	97.15	92.33	83.21	59.63	38.77	16.35	0.00
chr22:28727970-28728132	849	100.00	100.00	99.57	97.43	89.33	74.15	52.65	26.03	0.21
chr22:28729272-28729386	161	40.96	28.53	26.56	23.47	21.28	15.81	9.85	4.90	0.00
chr22:28730410-28730558	528	98.10	93.76	91.21	83.14	70.30	56.33	34.66	12.20	0.00
chr22:28734393-28734737	902	99.42	98.13	96.32	92.06	84.01	65.16	48.54	25.44	1.09

chr22:28737250-28737333	701	100.00	95.10	92.17	80.13	67.67	49.42	37.83	17.90	0.77
chr22:28737875-28738048	1048	100.00	99.83	98.45	96.23	90.62	71.77	56.31	33.76	1.03
chr22:28741759-28742432	592	99.61	95.10	92.71	82.36	70.28	48.82	30.52	14.71	0.47
chr8:89933326-89935622	626	94.62	92.79	90.22	82.27	73.06	54.36	33.95	16.03	0.48
chr8:89936077-89936133	196	100.00	98.57	91.60	77.28	59.64	21.87	3.63	1.88	0.00
chr8:89937016-89937357	724	100.00	99.53	98.58	92.25	82.21	65.08	40.34	17.75	0.62
chr8:89943243-89943376	1497	100.00	100.00	100.00	99.05	94.19	82.00	65.76	43.63	1.54
chr8:89946130-89946635	614	98.07	97.78	96.12	89.48	79.62	59.66	36.18	12.53	0.24
chr8:89947814-89947902	891	100.00	100.00	100.00	96.84	87.57	74.87	55.97	24.23	0.39
chr8:89953234-89953701	1058	100.00	99.83	99.17	95.38	89.27	70.20	53.36	31.07	1.05
chr8:89955273-89955565	842	100.00	100.00	99.04	94.20	86.49	68.69	45.41	20.91	0.77
chr8:89957965-89958032	1059	100.00	100.00	100.00	96.62	91.25	76.30	62.97	30.27	1.20
chr8:89958715-89958864	203	100.00	98.37	90.08	76.80	55.49	25.65	7.38	1.89	0.00
chr8:89964400-89964517	770	100.00	99.96	99.47	96.73	84.03	74.40	49.88	19.61	0.37
chr8:89970354-89970567	711	97.93	96.53	96.23	89.05	80.26	61.80	40.23	16.54	0.49
chr8:89971163-89971300	427	99.46	97.47	95.39	82.62	74.31	46.37	26.50	6.69	0.00
chr8:89978210-89978333	718	100.00	99.98	99.12	94.11	85.20	67.04	42.51	16.69	0.67
chr8:89980724-89980903	1289	100.00	99.91	99.91	99.10	92.95	81.56	67.69	42.06	1.43
chr8:89981365-89981665	883	99.88	99.53	98.37	93.02	84.74	63.10	46.80	25.20	1.18
chr8:89981943-89982012	332	92.06	86.87	81.02	72.91	62.50	35.56	14.89	4.98	0.00
chr8:89982394-89982865	709	100.00	99.40	98.73	92.32	82.41	65.10	41.47	18.68	0.59
chr8:89984198-89984414	82	95.50	80.30	68.84	43.29	23.50	4.11	1.98	1.00	0.00
chr8:89984515-89984743	573	98.34	93.05	90.47	85.51	72.77	55.79	36.13	13.03	0.44
chr8:89984907-89984983	458	100.00	100.00	98.37	88.79	80.03	53.19	24.13	5.92	0.00
chr8:90003102-90003238	662	100.00	99.98	95.52	89.92	75.69	52.97	38.02	19.09	0.36

Supplementary Table 3.2 key: (Chr) Chromosome; (*) Regions of exons with false positives; (Avg) Average; (%) Percent.

Chapter 4: Gene panel screening for insight towards breast cancer susceptibility in different ethnicities.

This information will be submitted to *Genes, Chromosomes, and Cancer*: Bishop MR, Omeler SM, Huskey ALW, and Merner ND. Gene panel screening for insight towards breast cancer susceptibility in different ethnicities. *Genes, Chromosomes, and Cancer*.

4.1 Abstract:

Despite the advantages of NGS gene panel screening, few efforts have been published on gene-panel testing in minority populations. Therefore, African American BC genetics remains vastly understudied and is less understood compared to European American BC susceptibility. Thus, 97 BC-affected individuals of African and European descent from the AHCC were screened using the B.O.P. research-based gene-panel. Upon sequencing and bioinformatic processing, rare coding variants in 14 cancer susceptibility genes were assessed and compared between ethnicities. Of the 107 unique rare variants, 27 and 80 were classified as damaging and seemingly benign, respectively. Although there were no significant differences between the ethnicities regarding damaging variants, African American BC cases were more likely to have at least one seemingly benign variant and more likely to have multiple seemingly benign variants compared to European American BC cases (*P value* 2.05 X 10^{-3} and *P value* 3.02 X 10^{-3}, respectively). Interestingly, three variants (*BRCA2* p.S976I [*P value* 3.67 X 10^{-6}], *ATM* p.F763L [*P value* 6.40 X 10^{-3}], and *RAD51D* p.L84H [*P value* 0.019]) were associated with African American BC when compared to ethnic-specific controls and are currently not reported as pathogenic in ClinVar. Ultimately, B.O.P. screening provides essential insight towards the

variant contributions in clinically relevant cancer susceptibility genes and differences between ethnicities, stressing the need for future research efforts to include all variant types in all ethnicities to elucidate BC genetics that may explain current disparities.

4.2 Introduction:

The introduction of NGS and implementation of gene-panel screening has greatly reduced sequencing cost and has enabled the analysis of multiple genes harboring disease risk variants in parallel.[30,53] Despite these advantages, few efforts have been published on gene-panel testing in minority populations. Specifically regarding the assessment of hereditary BC risk, gene-panel screenings have focused on cohorts of mostly white, non-Hispanic individuals[53]; only a small number of studies sought to exclusively examine populations of Asian[131-133] or African[102] descent. However, a few screening reports of multi-ethnic cohorts have been published, which have begun to contribute towards a better understanding of genetic risk in different ethnicities.[134-136] Overall, similar to past publications, African Americans have been reported to have a higher rate of VUSs, which directly reflects the fact that the majority of BC genetics research has failed to focus on underrepresented individuals.[134-137] Even with these efforts, African American BC genetics remains vastly understudied and is less understood compared to European American BC susceptibility.[102]

In addition to ethnicity, it is important to acknowledge that study inclusion criteria, screening technologies, panel size, variant calling algorithms, and variability in the interpretation of variant classes all contribute towards the mutations/variants being reported.[53,102,134] Thus, there is a critical need for additional research to further elucidate the complete spectrum of BC risk variants to improve clinical risk assessment and management strategies. The AHCC

provides an opportunity to study BC genetics in underrepresented individuals.[69] Thus, 97 BC-affected individuals of either African or European descent from the AHCC were screened using the research-based gene-panel, B.O.P. (Chapter 3). Rare variants in 14 cancer susceptibility genes were assessed and compared between ethnicities.

4.3 Materials and Methods:

Ninety-seven (35 African American and 62 European American) seemingly unrelated BC cases were selected from the AHCC[69] for this study, and the average age of BC onset was 45.7 and 47.4 years of age, respectively. Genomic DNA from each individual was screened using a custom-designed gene panel, B.O.P., which targets genes that are suggested, predicted, or clinically proven to be associated with risk of BC, OvC and/or PC (Chapter 3). DNA libraries were prepared following the HaloPlex HS Target Enrichment System For Illumina Sequencing Protocol (Version C0, December 2015) and subsequently sequenced on an Illumina HiSeq™ 2500 at the GSL at HudsonAlpha Institute for Biotechnology. Following capture and sequencing, variants were called using an in-house bioinformatics pipeline (Chapter 3).

Fourteen genes that were targeted on the B.O.P. panel and NCCN clinical management guidelines regarding the genetic risk of BC and/or OvC[54,138] were selected for variant analysis (*ATM* [NM_000051], *BARD1* [NM_000465], *BRCA*1 [NM_007300], *BRCA2* [NM_000059], *CDH1* [NM_004360], *BRIP1* [NM_032043], *CHEK2* [NM_001005735], *NBN* [NM_002485], *PALB2* [NM_024675], *PTEN* [NM_000314], *RAD51C* [NM_058216], *RAD51D* [NM_001142571], *STK11* [NM_000455], *TP53* [NM_000546]). The depth of coverage of each assessed gene was calculated using DepthOfCoverage tool within the GATK (v.3.4-46) and ranged from 408X-970X (Supplementary Table 4.1). Only variants within coding regions of the 14 genes were further investigated. Next, variants were filtered using ethnic-specific MAF of

≤1% from controls in the National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project EVS.[139] Additionally, known sequencing artifacts from previous screening and validation were removed (Chapter 3).

After filtering, TPs were identified according to criteria established through B.O.P.'s initial analytical assessment (Chapter 3). Therefore, TPs included variants with high confidence calls (depth of coverage ≥100X and alternate allele frequency ≥40%), as well as variants with low confidence calls (depth of coverage <100X and alternate allele frequency <40%) that were subsequently validated through PCR and Sanger sequencing. All TP variants were organized into categories based on classifications in ClinVar[140] as well as predicted pathogenicity (Figure 4.1). A variant not reported as pathogenic in ClinVar was predicted to be pathogenic if it was a truncation variant, or a missense variant predicted to be damaging/pathogenic in two of three prediction software tools (PolyPhen, SIFT [Sorting Intolerant From Tolerant], and MutationTaster).[115,140-142] All other variant types (i.e. synonymous variants) were considered seemingly benign (Figure 4.1). Using the Fisher's exact test in R (v 3.5.1), the number of individuals with variants in each category (Figure 4.1) were compared between ethnicities, and variants identified in more than one BC-affected individual were compared to ethnic-specific controls from EVS.[139]

4.4 Results:

A total of 129 (107 unique) rare variants were classified as TPs (Figure 4.1). Overall, the African American BC cases were more likely to harbor at least one rare variant in one of the 14 assessed susceptibility genes when compared to the European American BC cases (*P value* 0.016; Table 4.1). Additionally, 54% of African American BC cases and 32% of European American BC

cases harbored multiple rare variants amidst the 14 genes, resulting in a *P value* trending toward significance (*P value* 0.052; Table 4.1).

Of the 107 unique rare variants, 27 (25.2%) were classified as damaging, but only eight (three truncating and five missense) were pathogenic according to ClinVar[140] (Figure 4.1 and Table 4.2); thus, a clinically relevant variant was detected in 8% (8/97) of the BC cases screened (explaining 11% of the African Americans and 6% of European Americans; *P value* 0.454; Table 4.1 and 4.2). Three of the 27 damaging variants (*BARD1* c.1972C>T [p.R658C], *BRCA2* c.2926_2927delinsAT [p.S976I], and *BRCA2* c.9976A>T [p.K3326X]) were identified in more than one BC-affected individual, all of which specific to an ethnicity (Table 4.2). Interestingly, *BRCA2* p.S976I, which was detected in three African American BC cases, is currently classified as a VUS in ClinVar[140] but was determined to be associated with African American BC risk when compared to ethnic-specific controls (*P value* 3.67 X 10^{-6}; Table 4.2). Due to the overlapping variants, B.O.P. screening identified a total of 32 damaging variants, of which 10 and 22 were identified in African Americans and European Americans, respectively. Most (70%) of the damaging variants in the African American BC cases were in *BRCA1* (n=5) and *BRCA2* (n=2), and the remaining three were found in *CHEK2* (n=1), *NBN* (n=1), and *TP53* (n=1) (Figure 4.2 and Table 4.2). Regarding the European American BC cases, damaging variants were identified in *ATM* (n=7), *BARD1* (n=3), *BRCA2* (n=5), *BRIP1* (n=1), *CDH1* (n=1), *CHEK2* (n=3), *NBN* (n=1), and *PALB2* (n=1) (Figure 4.2 and Table 4.2). Noteworthy, three European Americans had more than one damaging variant (Table 4.1 and 4.2); each had a predicted to be pathogenic *BRCA2* variant along with either a damaging variant in *BARD1* or *CDH1* (Table 4.2). Overall, there were no statistically significant differences between ethnicities regarding damaging variants (Table 4.1).

In contrast, the majority (74.8%) of the 107 unique variants were classified as seemingly benign (Figure 4.1); these 80 rare seemingly benign variants consisted of 41 missense variants, two non-frameshifting deletions, and 37 synonymous variants. Twelve of the 80 unique variants were detected in multiple BC-affected individuals (Table 4.3). Of note, this included two missense variants that appear to be associated with African American BC: *ATM* c.2289T>A [p.F763L] (*P value* 6.40 X 10^{-3}) and *RAD51D* c.251T>A [p.L84H] (*P value* 0.019; Table 4.3). Interestingly, 83% of the African American BC cases compared to 50% of the European American BC cases had at least one seemingly benign variant in one of the 14 genes (*P value* 2.05 X 10^{-3}; Table 4.1). Specifically, African Americans were more likely to have at least one seemingly benign missense (*P value* 5.79 X 10^{-3}) as well as at least one synonymous variant (*P value* 0.041; Table 4.1). Furthermore, African American BC cases were more likely to have multiple seemingly benign variants throughout the 14 genes compared to European American BC cases (*P value* 3.02 X 10^{-3}; Table 4.1).

4.5 Discussion:

Involving underrepresented individuals in cancer genetics research is crucial in order to better understand risk in different ethnicities. Herein, 97 BC-affected individuals from the AHCC[69] were screened using the B.O.P. gene panel (Chapter 3) to identify variants in 14 cancer susceptibility genes and compare the spectrum of variants between African and European Americans. The 14 assessed genes were selected because pathogenic variants in those genes can influence clinical management; in fact, the NCCN has established BC and/or OvC risk management guidelines regarding genetic testing results for each of the genes.[54]

Even though guidelines have been developed for the clinical interpretation of sequencing variants,[63] in reality, classification still varies amongst different clinical laboratories, and variant reclassification is an issue.[134,143] Taking this into consideration for the classification of variants identified in this study, we defined damaging variants as reported to be pathogenic in ClinVar[140] or predicted to be pathogenic. Variants had predicted pathogenicity if they truncated the protein or had at least two prediction software tools classify them as deleterious/damaging.[115,141,142] All other variant types were considered seemingly benign. We recognize the limitations of such categories. For instance, some truncation mutations, such as *BRCA2* c.9976A>T [p.K3326X], are classified as benign in ClinVar.[140] However, it is important to acknowledge that BC risk variants are on a spectrum of associated risk[30] and *BRCA2* c.9976A>T [p.K3326X] has been reported as a low risk variant.[48,52] Furthermore, our classification system was strongly influenced by the fact that protein-truncating mutations in known BC susceptibly genes have indisputable evidence of pathogenicity.[5] On another note, the prediction software tools that were used in this study (PolyPhen, SIFT and MutationTaster), which are the most commonly used in clinical laboratories for missense classification, are known to overlook 20-35% of known pathogenic missense variants.[63,115,141,142,144] Furthermore, with algorithms that mainly consider amino acid evolutionary conservation and/or the effect of the amino acid substitution,[63,144] coding variants (both missense and synonymous) that affect binding sites or have epigenetic modifications are completely ignored. Seemingly benign variants might play a bigger role in susceptibility than once thought, considering that we identified two seemingly benign missense variants that appear to be associated with African American BC: *ATM* c.2289T>A [p.F763L] (*P value* 6.40 X 10^{-3}) and *RAD51D* c.251T>A [p.L84H] (*P value* 0.019). Also, it is interesting that there is a statistically significant difference in the percentage of African American BC cases that had at

least one seemingly benign variant in one of the 14 genes compared to European American BC cases (83% versus 50%, respectively; *P value* 2.05 X 10^{-3}). Further investigation (such as comparing large cohorts of ethnic-specific cases to controls) may reveal that seemingly benign variants explain a portion of the missing BC heritability, which, based on our results, would be monumental in understanding African American BC genetics and how such variants contribute to disparities.[69,82,102] Overall, this stresses that variant classification is fluid and reclassification/reanalysis can profoundly impact clinical management.[63,143]

Previously, the initial B.O.P. screening, which assessed coding variants with MAFs ≤2% in both ethnicities in cancer-affected cases from the AHCC, identified a significant difference in the number of individuals from each ethnicity with at least one variant (*P value* 2.71 X 10^{-3}) (Chapter 3). Interestingly, this significant difference was consistent when the cohort was increased to include 97 BC cases; significantly more African American BC cases had at least one rare variant (MAF ≤1%) in the 14 assessed genes compared to European American BC cases (89% versus 65%, respectively; *P value* 0.016). This was primarily a result of the aforementioned difference in seemingly benign variants, specifically with seemingly benign missense variants having the largest impact (*P value* 5.79 X 10^{-3}), corroborating the fact that African American BC cases are typically reported to have more VUSs.[134-137] On the contrary, there were no significant differences between the ethnicities regarding damaging variants, including clinically significant variants as well as those predicted to be pathogenic. Whether predicted to be pathogenic or benign, the true interpretation of a variant can only be revealed through an integrated analysis that considers multiple factors.[145] Therefore, in addition to prediction software that were implemented in this study to assess amino acid conservation and R-

group changes, functional and co-SA are essential components of the complex analysis for interpretation.[5,145]

NGS, including gene panel screening, detects the full spectrum of variants in the targeted region(s) for each individual screened, which provides an opportunity to explore how combinations of variants contribute towards polygenic risk.[146] Despite the fact that recent efforts examining polygenic risk of BC have focused on common variants,[48] rare variants that modify risk in *BRCA1* and *BRCA2* mutation carriers have been identified.[147,148] Considering this, assessing combinations of rare variants is likely a vital missing component for polygenic BC risk assessment. Forty percent of the BC cases B.O.P. screened (19 African Americans and 20 European Americans), had multiple rare variants in the 14 clinically relevant cancer susceptibility genes. Three European Americans had more than one damaging variant, which was not statistically significant compared to African Americans. Interestingly, multiple, seemingly-benign variants were detected in significantly more African American BC cases (*P value* 3.02 X 10^{-3}). Although, individually, these seemingly benign variants may only slightly elevate risk, specific combinations of these variants may multiplicatively influence risk of developing BC. Therefore, comparing such rare variant combinations between cases and ethnic-specific controls using NGS approaches will provide essential insight towards polygenic BC risk, particularly in African Americans.[5] This effort requires having individual sequencing files from each assessed control, which were not available for this study.

Furthermore, with the launch of NGS, several WES investigations have been carried out to identify novel BC risk variants; however, the majority of these were relatively unsuccessful due to the heterogeneity of BC genetics.[30] Noteworthy, the successful WES studies focused on relatively homogeneous populations;[30,149] thus, suggesting that investigating homogeneous

cohorts is a palatable approach to enhance our understanding of BC genetics. By screening cancer cases from the AHCC, which was established through strategic recruitment mechanisms that involved traveling to isolated and rural communities in Alabama, the detection of ancestral mutations in seemingly unrelated individuals was anticipated.[69] Overall, this B.O.P. screening suggests that the AHCC is relatively homogeneous since a total of 15 rare variants in the 14 cancer susceptibility genes were detected in multiple seemingly unrelated BC cases. Interestingly, three of those variants (*BRCA2* p.S976I [*P value* 3.67 X 10^{-6}], *ATM* p.F763L [*P value* 6.40 X 10^{-3}], and *RAD51D* p.L84H [*P value* 0.019]) were associated with African American BC when compared to ethnic-specific controls, all of which are currently not reported to be pathogenic in ClinVar.[140] Additionally, while this study focused on variants with ethnic-specific MAF≤1%, a previous B.O.P. analysis identified a common, synonymous variant (*STK11* c.369G>A [p.Q123Q]) associated with African American BC (*P value* 8.50 X 10^{-4}) when compared to ethnic-specific controls (MAF of 1.5%) (Chapter 3). Though screening other cohorts is required to validate these findings, the detection of significant *P values* demonstrates the power of this approach.

4.6 Conclusions

Only a small percentage (8%) of the B.O.P. screened BC cases were explained by a variant reported to be pathogenic in ClinVar[140]; more specifically, at least one clinically relevant variant was detected in 11% of African American and 6% of European American BC cases. However, a larger portion of BC cases (28.6% of African Americans and 30.6% of European Americans) had at least one damaging variant in the 14 assessed genes, which was consistent with previous findings.[30,102,150] Of further note, despite the fact that *BRCA1* and *BRCA2* are the

most common BC susceptibility genes,[150] no damaging *BRCA1* variants were detected in the European American BC cases. These results further emphasize that the majority of African American and European American individuals with familial/hereditary BC remain genetically unsolved.[5,102] In the past, the majority of association studies focused on overtly damaging variants with a presumed loss of function, consistently ignoring seemingly benign variants.[5,30] However, this gene panel screening ultimately provides essential insight towards the variant contributions in clinically relevant cancer susceptibility genes and the observed differences between European and African Americans. Therefore, future research must broaden the search for potential genetic risk factors to include all variant types in all ethnicities to elucidate BC genetics that may explain current disparities.

4.7 References:

5 Easton, D. F. *et al.* Gene-panel sequencing and the prediction of breast-cancer risk. *N Engl J Med* **372**, 2243-2257, doi:10.1056/NEJMsr1501341 (2015).

30 Chandler, M. R., Bilgili, E. P. & **Merner, N. D.** A Review of Whole-Exome Sequencing Efforts Toward Hereditary Breast Cancer Susceptibility Gene Discovery. *Hum Mutat* **37**, 835-846, doi:10.1002/humu.23017 (2016).

48 Mavaddat, N. *et al.* Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst* **107**, doi:10.1093/jnci/djv036 (2015).

52 Meeks, H. D. *et al.* BRCA2 Polymorphic Stop Codon K3326X and the Risk of Breast, Prostate, and Ovarian Cancers. *J Natl Cancer Inst* **108**, doi:10.1093/jnci/djv315 (2016).

53 Shah, P. D. & Nathanson, K. L. Application of Panel-Based Tests for Inherited Risk of Cancer. *Annu Rev Genomics Hum Genet* **18**, 201-227, doi:10.1146/annurev-genom-091416-035305 (2017).

54 National Comprehensive Cancer Network. *Genetic/Familial High-Risk Assessment: Breast and Ovarian (Version I.2019)*, <https://www.nccn.org/professionals/physician_gls/pdf/genetics_screening.pdf> (2019).

63 Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405-424, doi:10.1038/gim.2015.30 (2015).

69 Bishop, M. R. *et al.* Establishment of the Alabama Hereditary Cancer Cohort - strategies for the inclusion of underrepresented populations in cancer genetics research. *Mol Genet Genomic Med*, doi:10.1002/mgg3.443 (2018).

82 Daly, B. & Olopade, O. I. A perfect storm: How tumor biology, genomics, and health care delivery patterns collide to create a racial survival disparity in breast cancer and proposed interventions for change. *CA Cancer J Clin* **65**, 221-238, doi:10.3322/caac.21271 (2015).

102 Churpek, J. E. *et al.* Inherited predisposition to breast cancer among African American women. *Breast Cancer Res Treat* **149**, 31-39, doi:10.1007/s10549-014-3195-0 (2015).

115 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, doi:10.1002/0471142905.hg0720s76 (2013).

131 Lin, P. H. *et al.* Multiple gene sequencing for risk assessment in patients with early-onset or familial breast cancer. *Oncotarget* **7**, 8310-8320, doi:10.18632/oncotarget.7027 (2016).

132 Mannan, A. U. *et al.* Detection of high frequency of mutations in a breast and/or ovarian cancer cohort: implications of embracing a multi-gene panel in molecular diagnosis in India. *J Hum Genet* **61**, 515-522, doi:10.1038/jhg.2016.4 (2016).

133 Konrad, K. *et al.* Current practice of diabetes education in children and adolescents with type 1 diabetes in Germany and Austria: analysis based on the German/Austrian DPV database. *Pediatric diabetes* **17**, 483-491, doi:10.1111/pedi.12330 (2016).

134 Yorczyk, A., Robinson, L. S. & Ross, T. S. Use of panel tests in place of single gene tests in the cancer genetics clinic. *Clin Genet* **88**, 278-282, doi:10.1111/cge.12488 (2015).

135 Ricker, C. *et al.* Increased yield of actionable mutations using multi-gene panels to assess hereditary cancer susceptibility in an ethnically diverse clinical cohort. *Cancer Genet* **209**, 130-137, doi:10.1016/j.cancergen.2015.12.013 (2016).

136 Susswein, L. R. *et al.* Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients referred for next-generation cancer panel testing. *Genet Med* **18**, 823-832, doi:10.1038/gim.2015.166 (2016).

137 Nanda, R. *et al.* Genetic testing in an ethnically diverse cohort of high-risk women: a comparative analysis of BRCA1 and BRCA2 mutations in American families of European and African ancestry. *Jama* **294**, 1925-1933, doi:10.1001/jama.294.15.1925 (2005).

138 Network, N. C. C. Genetic/Familial High-Risk Assessment: Breast and Ovarian (Version I.2018.

139 National Heart, L. a. B. I. Exome Sequencing Project (ESP).

140 Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic acids research* **44**, D862-868, doi:10.1093/nar/gkv1222 (2016).

141 Sim, N. L. *et al.* SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res* **40**, W452-457, doi:10.1093/nar/gks539 (2012).

142 Schwarz, J. M., Rodelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* **7**, 575-576, doi:10.1038/nmeth0810-575 (2010).

143 Turner, S. A., Rao, S. K., Morgan, R. H., Vnencak-Jones, C. L. & Wiesner, G. L. The impact of variant classification on the clinical management of hereditary cancer syndromes. *Genet Med*, doi:10.1038/s41436-018-0063-z (2018).

144 Thusberg, J., Olatubosun, A. & Vihinen, M. Performance of mutation pathogenicity prediction methods on missense variants. *Hum Mutat* **32**, 358-368, doi:10.1002/humu.21445 (2011).

145 Calo, V. *et al.* The Clinical Significance of Unknown Sequence Variants in BRCA Genes. *Cancers (Basel)* **2**, 1644-1660, doi:10.3390/cancers2031644 (2010).

146 Chatterjee, N., Shi, J. & Garcia-Closas, M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet* **17**, 392-406, doi:10.1038/nrg.2016.27 (2016).

147 Sokolenko, A. P. *et al.* High prevalence of GPRC5A germline mutations in BRCA1-mutant breast cancer patients. *Int J Cancer* **134**, 2352-2358, doi:10.1002/ijc.28569 (2014).

148 Thiffault, I. *et al.* Germline truncating mutations in both MSH2 and BRCA2 in a single kindred. *British journal of cancer* **90**, 483-491, doi:10.1038/sj.bjc.6601424 (2004).

149 Cybulski, C. *et al.* Germline RECQL mutations are associated with breast cancer susceptibility. *Nat Genet* **47**, 643-646, doi:10.1038/ng.3284 (2015).

150 Kurian, A. W. *et al.* Clinical evaluation of a multiple-gene sequencing panel for hereditary cancer risk assessment. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **32**, 2001-2009, doi:10.1200/JCO.2013.53.6607 (2014).

4.8 Tables

Table 4.1: Ethnic comparison between different variant categories.

Assessed categories from Figure 4.1			African American		European American		Comparison of African and European Americans
			Average age of BC onset: 45.7		Average age of BC onset: 47.4		
			All BC cases (n=35)		All BC cases (n=62)		All BC cases
			Number of individuals	Percentage	Number of individuals	Percentage	*P value*
Number of individuals with:	at least one variant		31	89%	40	65%	0.016
	multiple variants		19	54%	20	32%	0.052
	at least one damaging variant		10	29%	19	31%	1.000
	at least one damaging variant that has been reported as pathogenic in ClinVar	Total	4	11%	4	6%	0.454
		Truncation	1	3%	2	3%	1.000
		Missense	3	9%	2	3%	0.348
	at least one damaging variant that has not been reported as pathogenic in ClinVar	Total	6	17%	16	26%	0.450
		Truncation	0	0%	3	5%	0.551
		Missense	6	17%	15	24%	0.456
	multiple damaging variants		0	0%	3	5%	0.551
	at least one seemingly benign variant	Total	29	83%	31	50%	2.05E-03
		Missense	21	60%	19	31%	5.79E-03
		Non-frameshifting deletion	4	11%	3	5%	0.249
		Synonymous	15	43%	14	23%	0.041
	multiple seemingly benign		15	43%	9	15%	3.02E-03

Table 42: Damaging variants detected after BCP screening 97 BC affected cases.

Gene Name	Chr	Position	Ref. Allele	Alt. Allele	Exon	DNA Change	Amino Acid Change	Protein Function	Reported to be pathogenic in ClinVar	CLINVAR^	Predicted to be pathogenic*	EA EVS[#]	AA EVS[#]	Number of individuals with variant called	Individual/ Sample	Genotype	Ethnicity	Age of BC onset
ATM (NM_000051)	chr11	108249096	T	C	exon 9	c.1229T>C	p.V410A	Missense	No	VUS	Yes	0.0022	0.0009	1	1EE-1	Het	EA	35
		108304736	A	T	exon 37	c.5558A>T	p.D1853V	Missense	No	Likely benign	Yes	0.0069	0.0009	1	1ECC-a	Het	EA	43
		108315883	G	A	exon 41	c.6067G>A	p.G2023R	Missense	No	VUS	Yes	0.0031	0.0007	1	1CG-a	Het	EA	46
		108316018	A	G	exon 42	c.6103A>G	p.T2035A	Missense	No	VUS	Yes	0	0	1	1EEA-a	Het	EA	53
		108330381	T	G	exon 50	c.7475T>G	p.L2492R	Missense	No	VUS	Yes	0.0001	0	1	1EDI-a	Het	EA	60
		108345882	C	T	exon 58	c.8558C>T	p.T2853M	Missense	No	VUS	Yes	0.0001	0.0002	1	1CBF-a	Het	EA	56
		108347290	C	T	exon 59	c.8596C>T	p.L2866F	Missense	No	VUS	Yes	0	0	1	1CBB-a	Het	EA	42
BARD1 (NM_000465)	chr2	214728861_214728862	TG	-	exon 11	c.2148_ 2149del	p.T716fs	Truncation	Yes	Pathogenic	N/A	0	0	1	1EEJ-a	Het	EA	70
		214730440	G	A	exon 10	c.1972C>T	p.R658C	Missense	No	Likely benign	Yes	0.0078	0.0023	2	1ECI-a	Het	EA	40
															1EFA-a	Het	EA	61
BRCA1 (NM_007300)	chr17	43051071	A	C	exon 21	c.5387T>G	p.M1796R	Missense	Yes	Pathogenic	N/A	0	0	1	1CAD-a	Het	AA	36
		43071232	G	A	exon 16	c.4745C>T	p.T1582I	Missense	No	VUS	Yes	0	0.0054	1	1EDC-a	Het	AA	56
BRCA2 (NM_000059)	chr13	32319232	G	C	exon 3	c.223G>C	p.A75P	Missense	No	VUS	Yes	0.0005	0	1	1EEJ-a	Het	EA	70
		32337281_ 32337282	TC	AT	exon 11	c.2926_2927 delinsAT	p.S976I	Missense	No	VUS	Yes	0	0	3	1ECJ-a	Het	AA	45
															1ECG-a	Het	AA	49
															1EEG-a	Het	AA	45
		32337453	A	T	exon 11	c.3098A>T	p.D1033V	Missense	No	VUS	Yes	0	0.0005	1	1CCC-a	Het	AA	40
		32339966_ 32339970	AGTAA	-	exon 11	c.5611_ 5615del[$]	p.S1871fs	Truncation	Yes	Pathogenic	N/A	0	0	1	1CB-a	Het	AA	36
		32379413	G	A	exon 22	c.8851G>A	p.A2951T	Missense	No	Likely benign	Yes	0.0057	0.0018	1	1CBJ-a	Het	EA	48
		32398489	A	T	exon 27	c.9976A>T	p.K3326X	Truncation	No	VUS	Yes	0.0084	0.0027	3	1CCA-a	Het	EA	43
															1ECI-a	Het	EA	40
															1EFJ-a	Het	EA	68

BRIP1 (NM_032043)	chr17	61847211	G	A	exon 6	c.517C>T	p.R173C	Missense	No	VUS	Yes	0.0049	0.0011	1	1ED-a	Het	EA	24
CDH1 (NM_004360)	chr16	68822063	G	A	exon 12	c.1774G>A	p.A592T	Missense	No	VUS	Yes	0.0063	0.0011	1	1EFJ-a	Het	EA	68
CHEK2 (NM_001005735)	chr22	28695232	A	G	exon 13	c.1399T>C	p.Y467H	Missense	No	VUS	Yes	0.0003	0	1	1EA-a	Het	EA	28
		28725099	A	G	exon 5	c.599T>C	p.I200T	Missense	Yes	Pathogenic	N/A	0.0024	0	1	1CCI-a	Het	EA	50
		28725338	T	C	exon 4	c.478A>G	p.R160G	Missense	Yes	Pathogenic	N/A	0.0001	0.0002	1	1EAB-a	Het	EA	36
		28734468	G	A	exon 2	c.254C>T	p.P85L	Missense	Yes	Pathogenic	N/A	0	0.007	1	1EDB-a	Het	AA	57
NBN (NM_002485)	chr8	89970539	C	T	exon 7	c.721G>A	p.A241T	Missense	No	VUS	Yes	0	0	1	1CCG-a	Het	AA	39
		89982770	G	-	exon 2	c.123delC	p.I41fs	Truncation	Yes	Pathogenic	N/A	0	0	1	1EBI-a	Het	EA	36
PALB2 (NM_024675)	chr16	23638125	T	C	exon 2	c.53A>G	p.K18R	Missense	No	VUS	Yes	0	0.0155	1	1EDH-a	Het	EA	64
TP53 (NM_000546)	chr17	7673776	G	A	exon 8	c.844C>T	p.R282W	Missense	Yes	Pathogenic	N/A	0.0002	0	1	1EC-a	Het	AA	23

Table 2 Key: (Chr) Chromosome; (Ref.) Reference; (EA) European American; (AA) African American; predicted to be damaging/pathogenic in two of three prediction software tools (PolyPhen, SIFT and MutationTaster); (#) esp6500siv2; (^) most severe clinical significance classification; (Het) Heterozygous; ($) The deletion was named using ANNOVAR (v.); however, it is within a short tandem repeat and commonly referred to as *BRCA2* c.5616_5620del5 (p.K1872Nfs) since Human Genome Variant Society (HGVS) nomenclature rules state to arbitrarily assign the deletion to the most 3' nucleotide.

Table 4.3: Seemingly benign variants detected after BCOP screening 97 BC affected cases.

Gene Name	Chr	Start position (hg38)	Ref. Allele	Alt. Allele	Exon	DNA Change	Amino Acid Change	Protein Function	Classification of variant type (Figure 4.1)	Reported to be pathogenic in ClinVar	CLINVAR ^	Predicted to be pathogenic *	EA EVS#	AA EVS#	Number of individuals with variant called	Individual / Sample	Genotype	Ethnicity	Age of BC onset
ATM (NM_000051)	chr11	108257519	T	A	exon 15	c.2289T>A	p.F763L	Missense	Seemingly benign	No	VUS	No	.	0.0014	2	1EEE-a	Het	AA	67
																1EEG-a	Het	AA	45
BARD1 (NM_000465)	chr2	214780779_214780799	TGGTGAAGAACA TTCAGGCAA	-	exon 4	c.1075_1095del	p.359_365del	Non-frameshifting deletion	Seemingly benign	No	Benign	No	.	.	6	1CAH-a	Het	AA	43
																1EC-a	Het	AA	23
																1ECE-a	Het	AA	42
																1ED-a	Het	EA	24
																1EEB-a	Het	EA	65
																1EEC-a	Het	EA	58
BRCA1 (NM_007300)	chr17	43091492	T	C	exon 10	c.4039A>G	p.R1347G	Missense	Seemingly benign	No	VUS	No	0.0067	0.0011	2	1EAC-a	Het	EA	50
																1EDF-a	Het	EA	45
BRCA2 (NM_000059)	chr13	32338613	G	T	exon 11	c.4258G>T	p.D1420Y	Missense	Seemingly benign	No	VUS	No	0.0054	0.0012	2	1EEA-a	Het	EA	53
																1EFA-a	Het	EA	61
	chr13	32340702	A	G	exon 11	c.6347A>G	p.H2116R	Missense	Seemingly benign	No	VUS	No	0.0005	0.0095	2	1EDH-a	Het	EA	64
																1CBA-a	Het	AA	51
CDH1 (NM_004360)	chr16	68801830	A	G	exon 3	c.324A>G	p.R108R	Synonymous	Seemingly benign	No	Likely benign	No	0.0003	0.0061	2	1CAF-a	Het	AA	50
																1CBH-a	Het	AA	32
	chr16	68822185	C	T	exon 12	c.1896C>T	p.H632H	Synonymous	Seemingly benign	No	Benign	No	0.0094	0.0541	2	1EBF-a	Het	EA	51
																1EI-a	Het	EA	37
	chr16	68833484	C	T	exon 16	c.2634C>T	p.G878G	Synonymous	Seemingly benign	No	Likely benign	No	0.0088	0.0389	2	1EBF-a	Het	EA	51
																1EI-a	Het	EA	37
NBN (NM_002485)	chr8	89980833	A	G	exon 4	c.381T>C	p.A127A	Synonymous	Seemingly benign	No	Likely benign	No	0.0045	0.0018	2	1EAJ-a	Het	EA	35
																1ECF-a	Het	EA	40
PALB2 (NM_024675)	chr16	23629898	T	C	exon 5	c.2256A>G	p.G752G	Synonymous	Seemingly benign	No	VUS	No	.	0.0055	2	1CBA-a	Het	AA	51
																1CCB-a	Het	AA	43
		58695161	G	A		c.376G>A	p.A126T	Missense		No	VUS	No			3	1CAB-a	Het	EA	60

RAD51C (NM_058216)	chr17				exon 2				Seemingly benign				0.0064	0.0011		1EAB-a	Het	EA	36
																1EAH-a	Het	AA	40
RAD51D (NM_001142571)	chr17	35116931	A	T	exon 3	c.251T>A	p.L84H	Missense	Seemingly benign	No	Likely benign	No	.	0.0026	2	1CAH-a	Het	AA	43
																1ECJ-a	Het	AA	45

Table 43 Key: (Chr) Chromosome; (Ref.) Reference; (EA) European American; (AA) African American; predicted to be damaging/pathogenic in two of three prediction software tools (PolyPhen, SIFT and MutationTaster); (#) esp6500siv2; (^) most severe clinical significance classification; (Het) Heterozygous.

4.9 Figures

Figure 4.1: Description and classifications of called variants detected after B.O.P. gene panel screening, NGS, bioinformatics processing, and filtering. **(^)** most severe clinical significance classification in ClinVar**; (*) Variants were considered to have predicted pathogenicity if two of the three prediction software tools (PolyPhen, SIFT and MutationTaster) predicted the variant to be damaging/pathogenic or if the variant prematurely truncated the protein; (AAs) African Americans; (EAs) European Americans.**

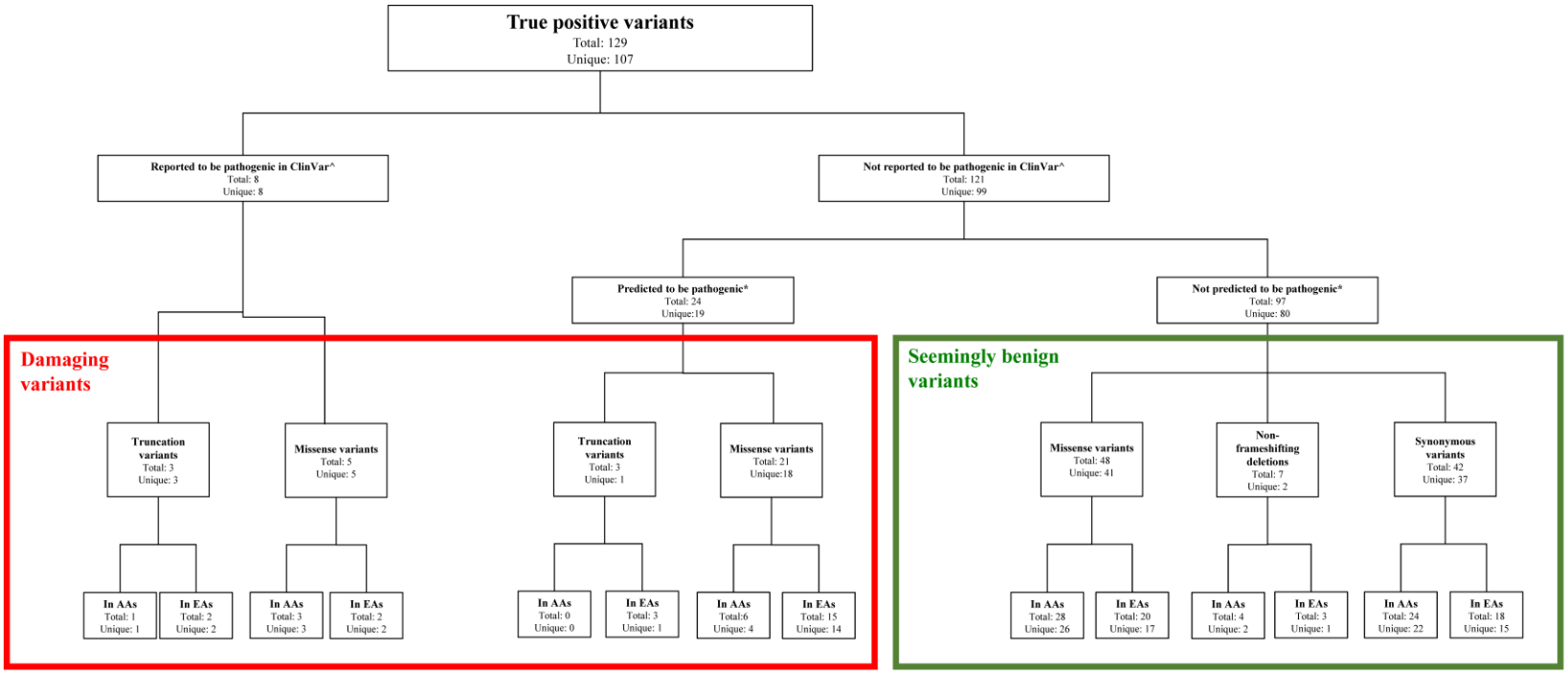
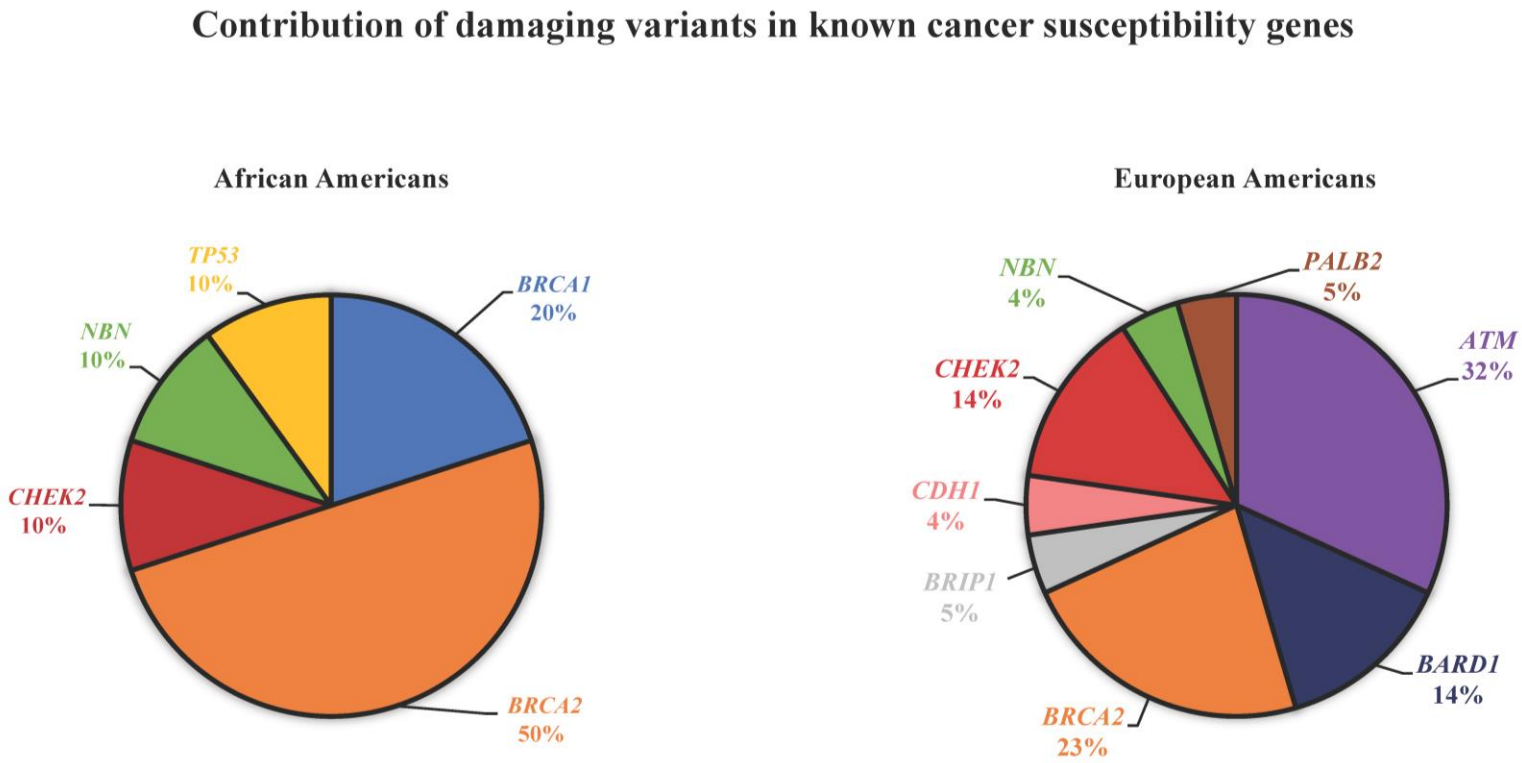


Figure 4.2: Damaging variant contributions amongst the 14 cancer susceptibility genes for both ethnicities.



4.10 Supplementary Material

Supplementary Table 4.1: Summary of coverage for the fourteen assessed genes from the B.O.P. panel.

Gene or targeted regions		Accession number	# of targeted regions	Size (bp)	Average sequencing depth (X)	Interquartile range (X)			% bases covered greater than or equal to:									
						First quartile	Median	Third quartile	1X	10X	20X	50X	100X	250X	500X	1000X	5000X	10000X
Genes investigated for analytical validation	*ATM*	NM_000051	65	15545	735	327	620	1008	97.7	96.5	95.2	91.8	86.1	70.2	48.8	22.6	0.7	0.3
	BARD1	NM_000465	13	6329	847	428	757	1131	99.2	98.2	97.2	93.8	88.3	74.6	56.8	28.4	0.8	0.5
	BRCA1	NM_007300	24	7750	970	382	734	1242	98.1	97.6	96.9	94.0	88.8	74.2	55.1	31.0	1.2	0.5
	BRCA2	NM_000059	28	12078	915	434	780	1203	98.7	98.2	97.6	95.1	90.4	77.1	57.6	30.4	1.0	0.5
	BRIP1	NM_032043	20	8566	692	305	624	984	95.3	93.8	92.7	89.2	83.5	68.1	48.0	21.6	0.7	0.3
	CDH1	NM_004360	16	5269	864	374	748	1180	98.5	97.6	96.7	94.0	89.1	74.1	54.5	28.8	0.8	0.4
	CHEK2	NM_001005735	23	4605	665	258	541	933	96.1	94.4	93.0	88.4	81.2	64.7	43.9	19.3	0.9	0.3
	NBN	NM_002485	22	6681	685	315	594	950	97.7	96.5	95.0	90.6	84.3	68.6	46.9	20.8	0.7	0.2
	PALB2	NM_024675	13	4318	932	508	804	1219	99.9	99.4	98.9	96.7	92.3	80.6	60.9	31.2	0.9	0.4
	PTEN	NM_000314	10	10248	554	201	435	795	97.1	94.2	92.0	85.8	77.2	58.5	36.6	14.8	0.6	0.2
	RAD51C	NM_058216	11	3173	751	360	653	1016	99.6	98.5	97.4	93.4	87.4	72.0	51.3	23.3	0.7	0.3
	RAD51D	NM_001142571	11	2803	787	398	729	1072	99.9	99.3	98.3	95.3	88.9	74.7	54.0	25.5	0.8	0.4
	STK11	NM_000455	10	3476	408	127	304	615	93.3	89.6	87.1	79.7	69.5	47.1	25.7	8.1	0.5	0.1
	TP53	NM_000546	14	4216	709	281	588	962	97.6	95.3	94.1	91.0	83.7	66.7	46.4	21.2	0.7	0.4
All targeted B.O.P. regions			1417	499521	745	327	628	1021	98.2	97.0	95.9	92.3	86.2	70.2	49.2	23.0	0.7	0.4

Supplementary Table 4.1 key: (#) Number; (%) Percent.

Chapter 5: Conclusions for the Alabama Hereditary Cancer Cohort and genetic analysis.

5.1 Recruitment and enrollment for establishment of the AHCC

Overall, this dissertation outlines the protocols that were established and implemented to enroll underrepresented individuals into a hereditary BC cancer genetic study and the subsequent development of a biobank, the AHCC. The primary accomplishments made during the first three years of this project provides essential insight on how to continue and expand the efforts to create a unique cohort for independent and collaborative genetic research studies. First, the hospital recruitment protocol, which is the most standard mode of recruitment, was established for its efficiency in identifying study participants and collecting complete medical records (i.e. pathology reports and clinical gene screening reports). In order to expand this efficient mode of recruitment, collaborations with additional hospitals will be established through IRB reliance agreements. Second, due to Alabama being a significantly medically underserved state with double the national percentage of self-identifying African Americans, an alternative method, CBR, was developed to overcome recruitment barriers and enabled our team to connect with even more underrepresented individuals. Overall, we have learned that the effort required to include underrepresented individuals in research is immense and challenging. It is an important effort that should no longer be underappreciated and adaptive mechanisms to recruit and enroll such individuals must be developed based on the communities needs to overcome common recruitment barriers.[81]

5.2 Conclusions for the analytical validity of the B.O.P. gene panel

Regarding the analytical validity assessment of the B.O.P. panel, a summary of the panel's ability to accurately detect mutations in 10 NCCN clinically actionable genes was reported.[109] Despite the potential biases of the B.O.P. capture and NGS, the high depth of coverage, low FDR, and

great sensitivity and specificity for *BRCA1* and *BRCA2* strongly support the use of this research gene panel to further explore hereditary BC, OvC, and PC genetics. Upon B.O.P. screening 43 individuals from the AHCC followed by bioinformatics processing and variant filtering, 74 variants identified in 10 clinically-relevant genes were carried through for validation. After identifying the 61 TPs (average sequencing depth of 659X and alternate allele frequency of 51%) and 13 false positives (average sequencing depth was 34X and alternate allele frequency was 33%), it was concluded that high sequencing depths (>100X) signified a TP, but low sequencing depth was not always indicative of a false positive. Furthermore, after assessing individuals with previous clinical gene screening, the sensitivity and specificity of *BRCA1/2* were calculated to be 100% and 92.3%, respectively. In the end, this screening aided in establishing criteria to alleviate future validation efforts and strongly confirms the continued use of the B.O.P. gene panel to further investigate hereditary cancer susceptibility.

5.3 Conclusions for the genetic analysis of known cancer susceptibility genes in different ethnicities using the B.O.P. gene panel

After B.O.P. screening 97 BC-affected individuals of African and European descent from the AHCC and bioinformatic processing, rare coding variants in 14 cancer susceptibility genes were assessed and compared between ethnicities. The identified TPs (107 unique rare variants) were divided into two categories: damaging and seemingly benign. Interestingly, only a small portion (8%) of the B.O.P. screened BC-affected cases were explained by a variant reported to be pathogenic in ClinVar;[140] however, a larger portion of BC cases (28.6% of African Americans and 30.6% of European Americans) had at least one damaging variant in the 14 assessed genes, which was consistent with previous findings.[30,102,150] This analysis further emphasizes that the

majority of African American and European American individuals with familial/hereditary BC remain genetically unsolved.[5,102] Although there were no significant differences between the ethnicities regarding damaging variants, African American BC cases were more likely to have at least one seemingly benign variant and more likely to have multiple seemingly benign variants compared to European American BC cases. Moreover, a total of 15 rare variants were identified in more than one seemingly unrelated BC-affected cases, strongly suggesting that the AHCC is a relatively homogeneous cohort. Of further interest, three variants not reported to be pathogenic in ClinVar (*BRCA2* p.S976I [*P value* 3.67 X 10^{-6}], *ATM* p.F763L [*P value* 6.40 X 10^{-3}], and *RAD51D* p.L84H [*P value* 0.019]) were associated with African American BC when compared to ethnic-specific controls from EVS. Consequently, future BC genetics studies should assess all variant types in all ethnicities to elucidate BC genetics that may explain current disparities.

5.4 Overall conclusions

In summary, this dissertation describes the establishment of the AHCC and stemming genetic analyses using the custom designed B.O.P. capture panel, massively parallel sequencing, and an in-house bioinformatics pipeline. Overall, involving underrepresented individuals in cancer genetics research is crucial in order to better understand risk in different ethnicities, and these efforts ultimately provide the continuous opportunity to explore hereditary cancer genetics research in underrepresented individuals from Alabama to improve risk assessment and increase insight to the disease mechanism. Although the cohort for this initial assessment is small, B.O.P. has begun to determine the mutation contributions of clinically valid genes in different ethnicities as well as permit the investigation of VUSs and other variant types and their effect towards polygenic risk. Furthermore, continued B.O.P. screening will provide additional

evidence to confirm or refute previously suggested susceptibility genes, lessening the number of genes that lack clinical validity on commercially available panels. Lastly, with the incorporation of candidate genes on B.O.P., it has the potential to identify novel genetic risk factors that are contributing towards BC, OvC, and PC. Not only are the potential implications for the implementation of B.O.P. screening immense, in the near future, these findings will lead to reductions in BC morbidity and mortality in underrepresented individuals through more accurate risk assessment and management options.

5.5 References

5 Easton, D. F. *et al.* Gene-panel sequencing and the prediction of breast-cancer risk. *N Engl J Med* **372**, 2243-2257, doi:10.1056/NEJMsr1501341 (2015).

30 Chandler, M. R., Bilgili, E. P. & **Merner, N. D.** A Review of Whole-Exome Sequencing Efforts Toward Hereditary Breast Cancer Susceptibility Gene Discovery. *Hum Mutat* **37**, 835-846, doi:10.1002/humu.23017 (2016).

81 Bishop, M. R. *et al.* Establishment of the Alabama Hereditary Cancer Cohort - strategies for the inclusion of underrepresented populations in cancer genetics research. *Molecular Genetics & Genomic Medicine (Accepted with minor revisions. Revised manuscript (MGG3-2018-03-0049.R1) submitted on 06/06/18)*, doi:DOI:10.1002/mgg3.443 (2018).

102 Churpek, J. E. *et al.* Inherited predisposition to breast cancer among African American women. *Breast Cancer Res Treat* **149**, 31-39, doi:10.1007/s10549-014-3195-0 (2015).

109 National Comprehensive Cancer Network. *Genetic/Familial High-Risk Assessment: Breast and Ovarian (Version I.2018)*, <https://www.nccn.org/store/login/login.aspx?ReturnURL=https://www.nccn.org/professionals/physician_gls/pdf/genetics_screening.pdf> (2018).

140 Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic acids research* **44**, D862-868, doi:10.1093/nar/gkv1222 (2016).

150 Kurian, A. W. *et al.* Clinical evaluation of a multiple-gene sequencing panel for hereditary cancer risk assessment. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **32**, 2001-2009, doi:10.1200/JCO.2013.53.6607 (2014).

6: Complete Reference List

1 ACS. American Cancer Society: Breast Cancer Facts and Figures 2013-2014. (2014).

2 Apostolou, P. & Fostira, F. Hereditary breast cancer: the era of new susceptibility genes. *Biomed Res Int* **2013**, 747318, doi:10.1155/2013/747318 (2013).

3 Shiovitz, S. & Korde, L. A. Genetics of breast cancer: a topic in evolution. *Ann Oncol*, doi:10.1093/annonc/mdv022 (2015).

4 Berliner, J. L. & Fay, A. M. Risk assessment and genetic counseling for hereditary breast and ovarian cancer: recommendations of the National Society of Genetic Counselors. *J Genet Couns* **16**, 241-260, doi:10.1007/s10897-007-9090-7 (2007).

5 Easton, D. F. *et al.* Gene-panel sequencing and the prediction of breast-cancer risk. *N Engl J Med* **372**, 2243-2257, doi:10.1056/NEJMsr1501341 (2015).

6 OMIM®. Online Mendelian Inheritance in Man. (2015).

7 Miki, Y. *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**, 66-71 (1994).

8 Wooster, R. *et al.* Identification of the breast cancer susceptibility gene BRCA2. *Nature* **378**, 789-792, doi:10.1038/378789a0 (1995).

9 Lalloo, F. & Evans, D. G. Familial breast cancer. *Clin Genet* **82**, 105-114, doi:10.1111/j.1399-0004.2012.01859.x (2012).

10 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, doi:10.1086/375033 (2003).

11 King, M. C., Marks, J. H. & Mandell, J. B. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. *Science* **302**, 643-646, doi:10.1126/science.1088759 (2003).

12 Fackenthal, J. D. & Olopade, O. I. Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations. *Nat Rev Cancer* **7**, 937-948, doi:10.1038/nrc2054 (2007).

13 Lynch, H. T. & Krush, A. J. Carcinoma of the breast and ovary in three families. *Surg Gynecol Obstet* **133**, 644-648 (1971).

14 Lynch, H. T. *et al.* Tumor variation in families with breast cancer. *JAMA* **222**, 1631-1635 (1972).

15 Lynch, H. T. *et al.* Familial association of carcinoma of the breast and ovary. *Surg Gynecol Obstet* **138**, 717-724 (1974).

16 Hall, J. M. *et al.* Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* **250**, 1684-1689 (1990).

17 Narod, S. A. *et al.* Familial breast-ovarian cancer locus on chromosome 17q12-q23. *Lancet* **338**, 82-83 (1991).

18 Morton, N. E. Sequential tests for the detection of linkage. *Am J Hum Genet* **7**, 277-318 (1955).

19 Wooster, R. *et al.* Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science* **265**, 2088-2090 (1994).

20 Gonzalez-Neira, A. *et al.* Genomewide high-density SNP linkage analysis of non-BRCA1/2 breast cancer families identifies various candidate regions and has greater power than microsatellite studies. *BMC Genomics* **8**, 299, doi:10.1186/1471-2164-8-299 (2007).

21 Hilbers, F. S. *et al.* Exome sequencing of germline DNA from non-BRCA1/2 familial breast cancer cases selected on the basis of aCGH tumor profiling. *PLoS One* **8**, e55734, doi:10.1371/journal.pone.0055734 (2013).

22 Ng, S. B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272-276, doi:10.1038/nature08250 (2009).

23 Ng, S. B. *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* **42**, 30-35, doi:10.1038/ng.499 (2010).

24 Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* **18**, 1851-1858, doi:10.1101/gr.078212.108 (2008).

25 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

26 Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat Biotechnol* **26**, 1135-1145, doi:10.1038/nbt1486 (2008).

27 Ku, C. S., Naidoo, N. & Pawitan, Y. Revisiting Mendelian disorders through exome sequencing. *Hum Genet* **129**, 351-370, doi:10.1007/s00439-011-0964-2 (2011).

28 Bamshad, M. J. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* **12**, 745-755, doi:10.1038/nrg3031 (2011).

29 Kiezun, A. *et al.* Exome sequencing and the genetic basis of complex traits. *Nat Genet* **44**, 623-630, doi:10.1038/ng.2303 (2012).

30 Chandler, M. R., Bilgili, E. P. & **Merner, N. D.** A Review of Whole-Exome Sequencing Efforts Toward Hereditary Breast Cancer Susceptibility Gene Discovery. *Hum Mutat* **37**, 835-846, doi:10.1002/humu.23017 (2016).

31 Ponder, B. A., Antoniou, A., Dunning, A., Easton, D. F. & Pharoah, P. D. Polygenic inherited predisposition to breast cancer. *Cold Spring Harbor symposia on quantitative biology* **70**, 35-41, doi:10.1101/sqb.2005.70.029 (2005).

32 Michailidou, K. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* **45**, 353-361, 361e351-352, doi:10.1038/ng.2563 (2013).

33 Easton, D. F. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087-1093, doi:10.1038/nature05887 (2007).

34 Hunter, D. J. *et al.* A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* **39**, 870-874, doi:10.1038/ng2075 (2007).

35 Stacey, S. N. *et al.* Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* **39**, 865-869, doi:10.1038/ng2064 (2007).

36 Stacey, S. N. *et al.* Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* **40**, 703-706, doi:10.1038/ng.131 (2008).

37 Ahmed, S. *et al.* Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet* **41**, 585-590, doi:10.1038/ng.354 (2009).

38 Thomas, G. *et al.* A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nat Genet* **41**, 579-584, doi:10.1038/ng.353 (2009).

39 Zheng, W. *et al.* Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet* **41**, 324-328, doi:10.1038/ng.318 (2009).

40 Antoniou, A. C. *et al.* A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population. *Nat Genet* **42**, 885-892, doi:10.1038/ng.669 (2010).

41 Turnbull, C. *et al.* Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet* **42**, 504-507, doi:10.1038/ng.586 (2010).

42 Fletcher, O. *et al.* Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. *J Natl Cancer Inst* **103**, 425-435, doi:10.1093/jnci/djq563 (2011).

43 Haiman, C. A. *et al.* A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. *Nat Genet* **43**, 1210-1214, doi:10.1038/ng.985 (2011).

44 Ghoussaini, M. *et al.* Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nat Genet* **44**, 312-318, doi:10.1038/ng.1049 (2012).

45 Siddiq, A. *et al.* A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. *Hum Mol Genet* **21**, 5373-5384, doi:10.1093/hmg/dds381 (2012).

46 Garcia-Closas, M. *et al.* Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nat Genet* **45**, 392-398, 398e391-392, doi:10.1038/ng.2561 (2013).

47 Michailidou, K. *et al.* Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet* **47**, 373-380, doi:10.1038/ng.3242 (2015).

48 Mavaddat, N. *et al.* Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst* **107**, doi:10.1093/jnci/djv036 (2015).

49 Narod, S. A. Breast cancer prevention in the era of precision medicine. *J Natl Cancer Inst* **107**, doi:10.1093/jnci/djv078 (2015).

50 Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Res* **46**, D754-D761, doi:10.1093/nar/gkx1098 (2018).

51 Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**, 308-311 (2001).

52 Meeks, H. D. *et al.* BRCA2 Polymorphic Stop Codon K3326X and the Risk of Breast, Prostate, and Ovarian Cancers. *J Natl Cancer Inst* **108**, doi:10.1093/jnci/djv315 (2016).

53 Shah, P. D. & Nathanson, K. L. Application of Panel-Based Tests for Inherited Risk of Cancer. *Annu Rev Genomics Hum Genet* **18**, 201-227, doi:10.1146/annurev-genom-091416-035305 (2017).

54 National Comprehensive Cancer Network. *Genetic/Familial High-Risk Assessment: Breast and Ovarian (Version I.2019)*, <https://www.nccn.org/professionals/physician_gls/pdf/genetics_screening.pdf> (2019).

55 Sanger, F. & Coulson, A. R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* **94**, 441-448 (1975).

56 Huang, X. C., Quesada, M. A. & Mathies, R. A. DNA sequencing using capillary array electrophoresis. *Anal Chem* **64**, 2149-2154 (1992).

57 Fackenthal, D. L., Chen, P. X. & Das, S. Denaturing high-performance liquid chromatography for mutation detection and genotyping. *Methods Mol Biol* **311**, 73-96, doi:10.1385/1-59259-957-5:073 (2005).

58 Reed, G. H., Kent, J. O. & Wittwer, C. T. High-resolution DNA melting analysis for simple and efficient molecular diagnostics. *Pharmacogenomics* **8**, 597-608, doi:10.2217/14622416.8.6.597 (2007).

59 Shendure, J., Mitra, R. D., Varma, C. & Church, G. M. Advanced sequencing technologies: methods and goals. *Nat Rev Genet* **5**, 335-344, doi:10.1038/nrg1325 (2004).

60 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**, 12629-12633, doi:10.1073/pnas.1007983107 (2010).

61 Walsh, T. *et al.* Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. *Proc Natl Acad Sci U S A* **108**, 18032-18037, doi:10.1073/pnas.1115052108 (2011).

62 Centers for Disease Control and Prevention. *Clinical Laboratory Improvement Amendments (CLIA)*, <https://wwwn.cdc.gov/clia/> (2017).

63 Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405-424, doi:10.1038/gim.2015.30 (2015).

64 Thompson, D., Easton, D. & Breast Cancer Linkage, C. Variation in BRCA1 cancer risks by mutation position. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **11**, 329-336 (2002).

65 Thompson, D., Easton, D. & Breast Cancer Linkage, C. Variation in cancer risks, by mutation position, in BRCA2 mutation carriers. *American journal of human genetics* **68**, 410-419, doi:10.1086/318181 (2001).

66 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. *American journal of human genetics* **81**, 873-883, doi:10.1086/521032 (2007).

67 Goldgar, D. E. *et al.* Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. *American journal of human genetics* **75**, 535-544, doi:10.1086/424388 (2004).

68 Plon, S. E. *et al.* Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Human mutation* **29**, 1282-1291, doi:10.1002/humu.20880 (2008).

69 Bishop, M. R. *et al.* Establishment of the Alabama Hereditary Cancer Cohort - strategies for the inclusion of underrepresented populations in cancer genetics research. *Mol Genet Genomic Med*, doi:10.1002/mgg3.443 (2018).

70 NIH. Inclusion of Women and Minorities as Participants in Research Involving Human Subjects- Policy Implementation Page. (2015).

71 Spratt, D. E. *et al.* Racial/Ethnic Disparities in Genomic Sequencing. *JAMA Oncol* **2**, 1070-1074, doi:10.1001/jamaoncol.2016.1854 (2016).

72 Wong, S. L. Medically underserved populations: disparities in quality and outcomes. *J Oncol Pract* **11**, 193-194, doi:10.1200/JOP.2015.004259 (2015).

73 Bustamante, C. D., Burchard, E. G. & De la Vega, F. M. Genomics for the world. *Nature* **475**, 163-165, doi:10.1038/475163a (2011).

74 Ashley, E. A. Towards precision medicine. *Nat Rev Genet* **17**, 507-522, doi:10.1038/nrg.2016.86 (2016).

75 Konkel, L. Racial and Ethnic Disparities in Research Studies: The Challenge of Creating More Diverse Cohorts. *Environ Health Perspect* **123**, A297-302, doi:10.1289/ehp.123-A297 (2015).

76 Streicher, S. A. *et al.* Reasons for participating and genetic information needs among racially and ethnically diverse biobank participants: a focus group study. *Journal of community genetics* **2**, 153-163, doi:10.1007/s12687-011-0052-2 (2011).

77 Daly, B. & Olopade, O. I. Race, ethnicity, and the diagnosis of breast cancer. *Jama* **313**, 141-142, doi:10.1001/jama.2014.17323 (2015).

78 Breast Cancer Facts & Figures. *American Cancer Society* (2015).

79 Stark, A. *et al.* African ancestry and higher prevalence of triple-negative breast cancer: findings from an international study. *Cancer* **116**, 4926-4932, doi:10.1002/cncr.25276 (2010).

80 Dietze, E. C., Sistrunk, C., Miranda-Carboni, G., O'Regan, R. & Seewaldt, V. L. Triple-negative breast cancer in African-American women: disparities versus biology. *Nature reviews. Cancer* **15**, 248-254, doi:10.1038/nrc3896 (2015).

81 Bishop, M. R. *et al.* Establishment of the Alabama Hereditary Cancer Cohort - strategies for the inclusion of underrepresented populations in cancer genetics research. *Molecular Genetics & Genomic Medicine (Accepted with minor revisions. Revised manuscript (MGG3-2018-03-0049.R1) submitted on 06/06/18)*, doi:DOI:10.1002/mgg3.443 (2018).

82 Daly, B. & Olopade, O. I. A perfect storm: How tumor biology, genomics, and health care delivery patterns collide to create a racial survival disparity in breast cancer and proposed interventions for change. *CA Cancer J Clin* **65**, 221-238, doi:10.3322/caac.21271 (2015).

83 Frist, W. H. Overcoming disparities in U.S. health care. *Health Aff (Millwood)* **24**, 445-451, doi:10.1377/hlthaff.24.2.445 (2005).

84 Ubri, P. & Artiga, S. *Disparities in Health and Health Care: Five Key Questions and Answers*, <https://www.kff.org/disparities-policy/issue-brief/disparities-in-health-and-health-care-five-key-questions-and-answers/> (2016).

85 Farmer, D. F., Jackson, S. A., Camacho, F. & Hall, M. A. Attitudes of African American and low socioeconomic status white women toward medical research. *J Health Care Poor Underserved* **18**, 85-99, doi:10.1353/hpu.2007.0008 (2007).

86 Brown, D. J. Recruitment of Black Americans for research: A model of success. *Journal of Multicultural Nursing & Health* **10**, 19-23 (2004).

87 Tan, D. S., Mok, T. S. & Rebbeck, T. R. Cancer Genomics: Diversity and Disparity Across Ethnicity and Geography. *J Clin Oncol* **34**, 91-101, doi:10.1200/JCO.2015.62.0096 (2016).

88 Taylor, J. Y. Recruitment of three generations of African American women into genetics research. *Journal of transcultural nursing : official journal of the Transcultural Nursing Society* **20**, 219-226, doi:10.1177/1043659608330352 (2009).

89 Newman, B. *et al.* The Carolina Breast Cancer Study: integrating population-based epidemiology and molecular biology. *Breast Cancer Res Treat* **35**, 51-60 (1995).

90 Palmer, J. R., Ambrosone, C. B. & Olshan, A. F. A collaborative study of the etiology of breast cancer subtypes in African American women: the AMBER consortium. *Cancer Causes Control* **25**, 309-319, doi:10.1007/s10552-013-0332-8 (2014).

91 Susan G. Komen North Central Alabama Affiliate 2015 Community Profile. (2015).

92 *Alabama Department of Public Health, Rural Health, Shortage Area Designations,* <http://www.adph.org/ruralhealth/Default.asp?id=901> (2017).

93 Gyawu, R., Quansah, J. E., Fall, S., Gichuhi, P. N. & Bovell-Benjamin, A. C. Community food environment measures in the Alabama Black Belt: Implications for cancer risk reduction. *Prev Med Rep* **2**, 689-698, doi:10.1016/j.pmedr.2015.08.015 (2015).

94 Rastogi, S., Johnson, T. D., Hoeffel, E. M. & Drewery, M. P. United States Census Bureau. The Black Population: 2010. (2010).

95 Miller, S. A., Dykes, D. D. & Polesky, H. F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic acids research* **16**, 1215 (1988).

96 Salowe, R. *et al.* Cost and yield considerations when expanding recruitment for genetic studies: the primary open-angle African American glaucoma genetics study. *BMC Med Res Methodol* **17**, 101, doi:10.1186/s12874-017-0374-9 (2017).

97 Helgesson, G. Are enrollment sites the key to optimizing participation in genetic studies? *Genome Med* **3**, 41, doi:10.1186/gm257 (2011).

98 Brandt, A. M. Racism and research: the case of the Tuskegee Syphilis Study. *Hastings Cent Rep* **8**, 21-29 (1978).

99 Greiner, K. A. *et al.* Effective recruitment strategies and community-based participatory research: community networks program centers' recruitment in cancer prevention studies. *Cancer Epidemiol Biomarkers Prev* **23**, 416-423, doi:10.1158/1055-9965.EPI-13-0760 (2014).

100 *United States Census Bureau: Alabama Quick Facts,* <https://www.census.gov/quickfacts/fact/table/AL#viewtop> (2017).

101 Lynch, H. T. Family information service and hereditary cancer. *Cancer* **91**, 625-628 (2001).

102 Churpek, J. E. *et al.* Inherited predisposition to breast cancer among African American women. *Breast Cancer Res Treat* **149**, 31-39, doi:10.1007/s10549-014-3195-0 (2015).

103 Zenka, D. (2012).

104 Palmer, J. R., Boggs, D. A., Adams-Campbell, L. L. & Rosenberg, L. Family history of cancer and risk of breast cancer in the Black Women's Health Study. *Cancer Causes Control* **20**, 1733-1737, doi:10.1007/s10552-009-9425-9 (2009).

105 Spurdle, A. B., Bowman, M. A., Shamsani, J. & Kirk, J. Endometrial cancer gene panels: clinical diagnostic vs research germline DNA testing. *Mod Pathol* **30**, 1048-1068, doi:10.1038/modpathol.2017.20 (2017).

106 Haddow, J. E. & Palomaki, G. E. in *Human Genome Epidemiology: A Scientific Foundation for Using Genetic Information to Improve Health and Prevent Disease* (eds M. Khoury, J. Little, & W. Burke) 217-233 (Oxford University Press, 2003).

107 Rehm, H. L. *et al.* ACMG clinical laboratory standards for next-generation sequencing. *Genet Med* **15**, 733-747, doi:10.1038/gim.2013.92 (2013).

108 National Comprehensive Cancer Network. *NCCN Guidelines® & Clinical Resources*, <https://www.nccn.org/professionals/physician_gls/default.aspx#detection> (2018).

109 National Comprehensive Cancer Network. *Genetic/Familial High-Risk Assessment: Breast and Ovarian (Version I.2018)*, <https://www.nccn.org/store/login/login.aspx?ReturnURL=https://www.nccn.org/professionals/physician_gls/pdf/genetics_screening.pdf> (2018).

110 Samorodnitsky, E. *et al.* Comparison of custom capture for targeted next-generation DNA sequencing. *J Mol Diagn* **17**, 64-75, doi:10.1016/j.jmoldx.2014.09.009 (2015).

111 Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**, 11 10 11-33, doi:10.1002/0471250953.bi1110s43 (2013).

112 EVS. (Seattle, WA, 2017).

113 Landrum, M. J. *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* **42**, D980-985, doi:10.1093/nar/gkt1113 (2014).

114 Rehm, H. L., Harrison, S. M. & Martin, C. L. ClinVar Is a Critical Resource to Advance Variant Interpretation. *Oncologist* **22**, 1562, doi:10.1634/theoncologist.2017-0246 (2017).

115 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, doi:10.1002/0471142905.hg0720s76 (2013).

116 Lim, E. C. *et al.* Next-generation sequencing using a pre-designed gene panel for the molecular diagnosis of congenital disorders in pediatric patients. *Hum Genomics* **9**, 33, doi:10.1186/s40246-015-0055-x (2015).

117 Green, A. *et al.* Assessment of HaloPlex amplification for sequence capture and massively parallel sequencing of arrhythmogenic right ventricular cardiomyopathy-associated genes. *J Mol Diagn* **17**, 31-42, doi:10.1016/j.jmoldx.2014.09.006 (2015).

118 Erman, B. *et al.* Investigation of Genetic Defects in Severe Combined Immunodeficiency Patients from Turkey by Targeted Sequencing. *Scand J Immunol* **85**, 227-234, doi:10.1111/sji.12523 (2017).

119 Ortega-Moreno, L. *et al.* Molecular diagnosis of patients with epilepsy and developmental delay using a customized panel of epilepsy genes. *PLoS One* **12**, e0188978, doi:10.1371/journal.pone.0188978 (2017).

120 Castera, L. *et al.* Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture targeting multiple candidate genes. *Eur J Hum Genet* **22**, 1305-1313, doi:10.1038/ejhg.2014.16 (2014).

121 Xue, Y., Ankala, A., Wilcox, W. R. & Hegde, M. R. Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing. *Genet Med* **17**, 444-451, doi:10.1038/gim.2014.122 (2015).

122 Harismendy, O. *et al.* Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol* **10**, R32, doi:10.1186/gb-2009-10-3-r32 (2009).

123 Chong, H. K. *et al.* The validation and clinical implementation of BRCAplus: a comprehensive high-risk breast cancer diagnostic assay. *PLoS One* **9**, e97408, doi:10.1371/journal.pone.0097408 (2014).

124 Judkins, T. *et al.* Development and analytical validation of a 25-gene next generation sequencing panel that includes the BRCA1 and BRCA2 genes to assess hereditary cancer risk. *BMC Cancer* **15**, 215, doi:10.1186/s12885-015-1224-y (2015).

125 Lincoln, S. E. *et al.* A Systematic Comparison of Traditional and Multigene Panel Testing for Hereditary Breast and Ovarian Cancer Genes in More Than 1000 Patients. *J Mol Diagn* **17**, 533-544, doi:10.1016/j.jmoldx.2015.04.009 (2015).

126 Mu, W., Lu, H. M., Chen, J., Li, S. & Elliott, A. M. Sanger Confirmation Is Required to Achieve Optimal Sensitivity and Specificity in Next-Generation Sequencing Panel Testing. *J Mol Diagn* **18**, 923-932, doi:10.1016/j.jmoldx.2016.07.006 (2016).

127 Vysotskaia, V. S. *et al.* Development and validation of a 36-gene sequencing assay for hereditary cancer risk assessment. *PeerJ* **5**, e3046, doi:10.7717/peerj.3046 (2017).

128 Buske, O. J., Manickaraj, A., Mital, S., Ray, P. N. & Brudno, M. Identification of deleterious synonymous variants in human genomes. *Bioinformatics* **29**, 1843-1850, doi:10.1093/bioinformatics/btt308 (2013).

129 Supek, F., Minana, B., Valcarcel, J., Gabaldon, T. & Lehner, B. Synonymous mutations frequently act as driver mutations in human cancers. *Cell* **156**, 1324-1335, doi:10.1016/j.cell.2014.01.051 (2014).

130 Kuchenbaecker, K. B. *et al.* Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers. *Journal of the National Cancer Institute* **109**, doi:10.1093/jnci/djw302 (2017).

131 Lin, P. H. *et al.* Multiple gene sequencing for risk assessment in patients with early-onset or familial breast cancer. *Oncotarget* **7**, 8310-8320, doi:10.18632/oncotarget.7027 (2016).

132 Mannan, A. U. *et al.* Detection of high frequency of mutations in a breast and/or ovarian cancer cohort: implications of embracing a multi-gene panel in molecular diagnosis in India. *J Hum Genet* **61**, 515-522, doi:10.1038/jhg.2016.4 (2016).

133 Konrad, K. *et al.* Current practice of diabetes education in children and adolescents with type 1 diabetes in Germany and Austria: analysis based on the German/Austrian DPV database. *Pediatric diabetes* **17**, 483-491, doi:10.1111/pedi.12330 (2016).

134 Yorczyk, A., Robinson, L. S. & Ross, T. S. Use of panel tests in place of single gene tests in the cancer genetics clinic. *Clin Genet* **88**, 278-282, doi:10.1111/cge.12488 (2015).

135 Ricker, C. *et al.* Increased yield of actionable mutations using multi-gene panels to assess hereditary cancer susceptibility in an ethnically diverse clinical cohort. *Cancer Genet* **209**, 130-137, doi:10.1016/j.cancergen.2015.12.013 (2016).

136 Susswein, L. R. *et al.* Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients referred for next-generation cancer panel testing. *Genet Med* **18**, 823-832, doi:10.1038/gim.2015.166 (2016).

137 Nanda, R. *et al.* Genetic testing in an ethnically diverse cohort of high-risk women: a comparative analysis of BRCA1 and BRCA2 mutations in American families of European and African ancestry. *Jama* **294**, 1925-1933, doi:10.1001/jama.294.15.1925 (2005).

138 Network, N. C. C. Genetic/Familial High-Risk Assessment: Breast and Ovarian (Version I.2018.

139 National Heart, L. a. B. I. Exome Sequencing Project (ESP).

140 Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic acids research* **44**, D862-868, doi:10.1093/nar/gkv1222 (2016).

141 Sim, N. L. *et al.* SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res* **40**, W452-457, doi:10.1093/nar/gks539 (2012).

142 Schwarz, J. M., Rodelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* **7**, 575-576, doi:10.1038/nmeth0810-575 (2010).

143 Turner, S. A., Rao, S. K., Morgan, R. H., Vnencak-Jones, C. L. & Wiesner, G. L. The impact of variant classification on the clinical management of hereditary cancer syndromes. *Genet Med*, doi:10.1038/s41436-018-0063-z (2018).

144 Thusberg, J., Olatubosun, A. & Vihinen, M. Performance of mutation pathogenicity prediction methods on missense variants. *Hum Mutat* **32**, 358-368, doi:10.1002/humu.21445 (2011).

145 Calo, V. *et al.* The Clinical Significance of Unknown Sequence Variants in BRCA Genes. *Cancers (Basel)* **2**, 1644-1660, doi:10.3390/cancers2031644 (2010).

146 Chatterjee, N., Shi, J. & Garcia-Closas, M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet* **17**, 392-406, doi:10.1038/nrg.2016.27 (2016).

147 Sokolenko, A. P. *et al.* High prevalence of GPRC5A germline mutations in BRCA1-mutant breast cancer patients. *Int J Cancer* **134**, 2352-2358, doi:10.1002/ijc.28569 (2014).

148 Thiffault, I. *et al.* Germline truncating mutations in both MSH2 and BRCA2 in a single kindred. *British journal of cancer* **90**, 483-491, doi:10.1038/sj.bjc.6601424 (2004).

149 Cybulski, C. *et al.* Germline RECQL mutations are associated with breast cancer susceptibility. *Nat Genet* **47**, 643-646, doi:10.1038/ng.3284 (2015).

150 Kurian, A. W. *et al.* Clinical evaluation of a multiple-gene sequencing panel for hereditary cancer risk assessment. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **32**, 2001-2009, doi:10.1200/JCO.2013.53.6607 (2014).