Whole Genome Analysis of High Risk Breast Cancer Families

K Lee[1], J Matthew[2], R Bloom[1], A Collins[1], C Bizon[3], Charles Schmitt[3], K Wilhelmsen[1], P Sullivan[1], JP Evans[1] and JS Berg[1]

[1] Department of Genetics, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
[2] School of Medicine, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
[3] The Renaissance Computing Institute, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

BACKGROUND: Highly penetrant Mendelian cancer predisposition syndromes are responsible for approximately 5-10% of human cancers, and delineation of the genes associated with these syndromes has led to fundamental advances in understanding tumorigenesis. Germline mutations in *BRCA1* and *BRCA2* account for the majority of hereditary breast and ovarian cancer families, yet clinical genetic testing fails to elucidate an underlying genetic etiology for the cancer in some families with histories strongly suggestive of a Mendelian predisposition to cancer. We hypothesize that these families harbor rare mutations in novel genes associated with increased risk for breast cancer.

PLAN: Utilize a tiered approach to whole genome analysis; array-comparative genomic hybridization (aCGH) analysis to look for copy number variants (CNVs) in all probands and whole genome sequencing (WGS) of families with the highest risk and most informative relatives. We developed a variant database to facilitate storage, annotation, and analysis of variants identified through WGS. All suspected deleterious variants are being confirmed and tested for co-segregation in families. Follow up studies, including sequencing of candidate genes in other unrelated probands, are planned. Our goal is to identify the genetic risk factor for breast cancer in a portion of our families, and be able to offer testing to unaffected relatives once a causative mutation is identified. In addition, the discovery of novel genes associated with Mendelian breast cancer susceptibility will provide further insight into this disease.

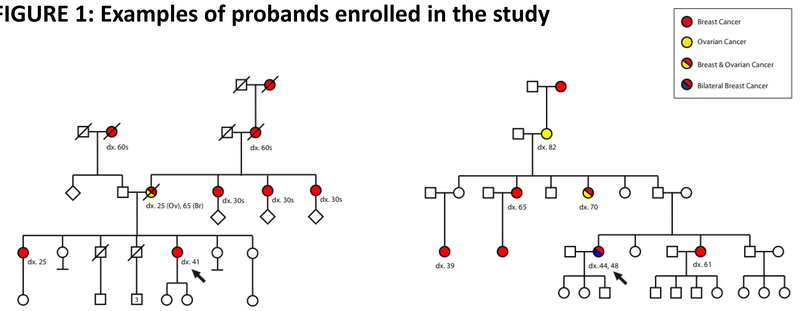
METHODS: We used Progeny to query clinical data from patients seen in the UNC Cancer Genetics Clinic since 1996. We identified patients with negative *BRCA1* and *BRCA2* test results despite a high pre-test probability of a mutation. Some patients also underwent genetic testing for p53 or PTEN. Two geneticists and one genetic counselor reviewed the pedigrees and prioritized eligible participants based on clinical suspicion and the number of informative family members available to study. We used Nimblegen 2.1M SNP aCGH to analyze for rare CNVs that disrupt candidate tumor suppressor genes. Array data was analyzed using segMNT (Nimblegen), and segments with log2 ratios > 0.3 and < -0.5 were defined as representing gains or losses, respectively.

Selected probands with a very high likelihood of a discrete genetic lesion were subsequently analyzed by WGS through Complete Genomics (Mountain View, CA) or the UNC high-throughput sequencing core facility.

PARTICIPANTS: We have recruited 67 probands from breast cancer families with negative genetic testing (figure 1). We have also enrolled 48 informative relatives to aid in our analysis. Characteristics of the cohort include:

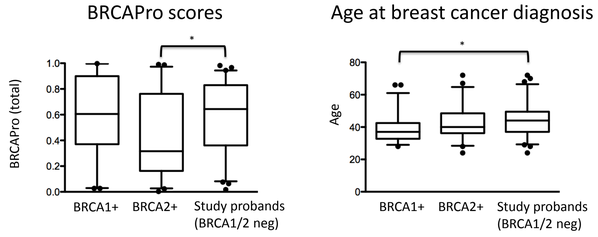
- 12 (17.9%) probands had bilateral breast cancer.
- 9 (13.4%) probands had a second primary other than breast; most commonly ovarian (4 probands) and skin (3 probands) cancers.
- 8 (11.9%) probands have 3 or more primary cancers.
- 2 probands have ovarian cancer with a family history of breast cancer.
- 2 probands with breast cancer are male.
- There are an average of 3 relatives with breast cancer per family.
- Probands have similar BRCAPro scores but slightly later age of onset compared to BRCA1+ patients in our clinic; probands have much higher BRCAPro scores but similar age of onset compared to BRCA2+ patients in our clinic (Figure 2)

FIGURE 1: Examples of probands enrolled in the study



Two examples of participants in the study are depicted. When available, we are also enrolling informative relatives in order to perform segregation analysis of the suspicious variants identified in probands. In families that will undergo WGS, two family members (such as the sisters in the family on the left, or the proband and one of her 1st cousins in the family on the right) will be analyzed simultaneously.

FIGURE 2: Comparison of clinic patients with BRCA1/2 mutations and probands enrolled in the study



Probands with breast cancer and positive for BRCA1 (n=50) or BRCA2 (n=48) mutations were identified through the clinic database. Probands in this study (n=67) had a mean BRCAPro score indistinguishable from BRCA1+ probands (0.587 vs. 0.586, P=0.99) but significantly different than BRCA2+ probands (0.43, P<0.01). Conversely, our probands were slightly older at their first diagnosis (mean age 44.7) compared to BRCA1+ probands (mean age 39.1, P<0.01) but similar to BRCA2+ probands (mean age 43.3, P=0.46). Shown are box-whisker plots (box = 25th – 75th percentile, whiskers 5th – 95th percentile; dots depict outliers). P values are based on unpaired two-tailed T-tests. * P < 0.01

TABLE 1: Summary of array-CGH results

	386702	386708	386757	387806	388030	388120	388130	388133	390879	433952	433954	433956
Segments with Log2 < -0.5	26	69	28	22	18	31	34	15	17	21	15	23
Mean size	38800	13653	26536	33486	31645	42046	21361	41072	30490	28255	21514	34141
Median size	22738	6557	17985	19874	21105	19004	15833	19766	21638	17460	16217	14585
Min size	10179	1599	9625	9625	9824	9625	10112	9552	10330	10355	9625	9625
Max size	166703	254805	92233	253192	118007	366194	117334	135420	158254	111387	70475	183367
# in DGV	25	46	28	18	16	30	27	13	17	19	15	23
# Novel	1	23	0	4	2	1	7	2	0	2	0	0
Genes in novel segments	0	14	0	3	1	0	4	1	0	1	0	0
Segments with Log2 > 0.3	55	54	50	28	56	86	54	65	279	45	39	42
Mean size	36585	31148	1071759	71499	68630	62549	46821	348165	46459	227766	62819	200908
Median size	18086	11195	76850	43633	26473	24045	26815	38068	21938	37383	20130	36758
Min size	9437	654	9718	10592	9428	9548	9819	10122	9354	10259	10183	11789
Max size	292000	278100	16845265	278100	830558	469079	288902	12993829	264454	3839153	602664	2006326
# in DGV	44	39	48	26	49	81	50	64	196	44	34	40
# Novel	11	14	2	2	7	5	3	1	83	1	5	2
Genes in novel segments	18	7	1	2	8	8	3	1	104	1	2	1

Complete array-CGH data is available for 12 probands. Our analysis is focused on detecting rare CNVs that disrupt candidate tumor suppressor genes. Novel putative losses were identified in 8 probands and novel putative gains were observed in all probands. No obvious deleterious CNVs were identified in our initial analysis. One individual had an inordinately high number of gains, which will be further clarified.

TABLE 2: Summary of our first 10 whole genome sequences

	Pt. 1	Pt. 2	Pt. 3	Pt. 4	Pt. 5	Pt. 6	Pt. 7	Pt. 8	Pt. 9	Pt. 10
Variants	$3.3x10^6$	$3.4x10^6$	$4.1x10^6$	$3.3x10^6$	$3.5x10^6$	$3.4x10^6$	$3.3x10^6$	$3.3x10^6$	$4.1x10^6$	$3.3x10^6$
Novel	140,368	324,178	803,480	151,624	202,970	164,676	151,863	146,498	423,892	145,047
Missense	9429	8955	10,140	9589	10,088	10,173	9753	9483	11,896	9608
Frameshift	232	235	262	277	265	275	275	245	309	293
Nonsense	99	91	105	85	79	92	79	85	113	85

To date, we have obtained WGS on 10 individuals from 6 pedigrees (2 individual probands, 2 pairs of sisters, and 2 pairs of 1st cousins). WGS on the 1st cousin of one of the individual probands is pending. Several candidate disease-causing variants have been identified, including novel variants in genes involved in the Fanconi Anemia DNA damage repair pathway. Further work is underway to analyze the segregation of these variants in family members and to determine whether other unrelated probands have mutations of these novel genes.

SUMMARY AND FUTURE DIRECTIONS: Patient and family enrollment is underway and the participation rate in our experience has been high (~80%). Patients and their families are eager to determine the cause of the apparently high-penetrance genetic variant responsible for their family's cancer susceptibility.

Array-CGH on 12 probands reveals a number of putative variants, including novel gains and losses, but no clearly disease-causing CNVs were revealed by the initial analysis. Given than large deletions might account for 5-10% of disease causing mutations, we may need to survey a much larger number of probands in order to identify a rare deleterious CNV.

Further array-CGH analyses are underway for additional members of our cohort and selected variants will be followed up in other family members for cosegregation. If a probable disease-causing CNV is identified, genes within that interval will be screened for mutations in other members of the cohort. We are also in the process of comparing WGS and array-CGH results to confirm the gains/losses identified by each technique

WGS on 10 individuals reveals a large number of possibly damaging variants. Bioinformatics analyses and family segregation studies are ongoing and are expected to clarify the significance of these candidates. Laboratory investigations of these initial hits are also underway in order to generate evidence of functional effects *in vitro*. Once confident that a particular candidate gene is associated with an elevated cancer risk, we will performed sanger sequencing of that gene in the remainder of our cohort to estimate its prevalence and possibly provide additional families with an answer for the cancer in their family. We will also offer genetic testing to unaffected relatives as part of the families' participation in the study once the familial mutation has been identified.

ACKNOWLEDGEMENTS: We thank our colleagues for helping us identify potential participants and our research assistants, Lydia Stickel and Minna Wiley, for recruiting our participants. We are also indebted to our families for their participation.

This study was supported by the University Cancer Research Fund and the North Carolina Translational and Clinical Sciences Institute - NC TraCS.