**The Carolina Breast Cancer Study 3,**

**Biospecimen Resources and Priorities**

The Carolina Breast Cancer Study consists of three phases, each of which included unique biospecimen collection protocols. The following describes the biospecimen resources included in Carolina Breast Cancer Study Phase 3 (Section A) and details priorities for the use of these resources (Section B).

1. **DESCRIPTION OF RESOURCES**

**CBCS Population and Study Design.** The CBCS Phase III is a population-based resource for studying variation in tumor gene expression among breast cancer cases in relation to etiology, disease progression and survival. Additional aims are investigating various health care access and utilization factors and survivorship outcomes. The CBCS was initiated in 1993 to address disparities in breast cancer risk by age and race[1](#_ENREF_1). This population-based case-control study based in North Carolina employed a sampling schema that deliberately oversampled AA and young women (<50) with breast cancer to address risk factors specific to those generally underrepresented groups. Phases I-II accrued over 4000 women ending in 2001, and has been responsible for over 90 scientific papers examining breast cancer genetic and environmental risk, characteristics, and behavior.

The CBCS Phase III (also known as the Jeanne Lucas Study after a beloved African-American state senator who died of Breast Cancer and was a strong advocate for research into breast cancer disparities), is a population-based, case-control study of invasive breast cancer conducted in 44 of 100 counties across North Carolina. Newly-diagnosed incident cases are identified using the UNC Lineberger Rapid Case Ascertainment Core in cooperation with the North Carolina Central Cancer Registry. During phases I and II, controls were selected from Division of Motor Vehicles (women younger than 65 years old) and United States Health Care Financing Administration lists (women 65 years old or greater) and frequency matched to cases based upon age and race (± 5 years). Phase III does not include controls due to the large number of control data available and the desire to focus on a larger number of cases. A recent study supports our case-only approach[2](#_ENREF_2)); in that study, theoretical calculations of the statistical power of a test applied to the problem of dealing with case-control differences in genetic ancestry related to population isolation or population admixture revealed that statistical correction for admixture differences is feasible even in studies that carry over control data from a different region/population. In this case, there is even less concern given that the populations included in CBCS Phases I, II, and III have considerable regional overlap presumably reflecting shared admixture proportions (for which we can account).

Recruitment of Study Participants**.** CBCS Phase III recruited cases between May 2008 and October 2013, with 2998 cases enrolled. The study includes approximately 1500 incident cases of invasive breast cancer in AA women: equally divided between cases <50 and ≥50. The study also recruited approximately 1500 incident cases of invasive breast cancer in non-AA women following the same age-based sampling scheme. Contact rates are 95%, cooperation rates are 76%, and overall response rates are 72%. Response rates are similar for African American (70%) and white (74%) women. CBCS uses similar procedures to conduct participant interviews and collect blood samples and tumor blocks, and has added the prospective collection of detailed treatment and outcome data that is unique to this study.

In-person and follow-up interviews.In-person interviews are used to collect information on risk factors for breast cancer. Blood samples are obtained at the time of interview, as well as permission to obtain formalin-fixed, paraffin-embedded tumor blocks and medical records. The interview includes information on family history and other risk factors, as well as quality of life and initial treatment information. Follow-up interviews have been conducted with patients at 9 months (follow-up #1, N = 2861) and 18 months (follow-up #2, N = 2524) with less than 5% loss to follow-up. Consent for medical records is obtained and information is collected on quality of life and other factors as well as adherence to hormonal therapy.

Medical Record Abstraction. Pathology reports are used to confirm the breast cancer diagnosis, and to request the "best diagnostic tissue block" and accompanying H&E slide(s) chosen at the discretion of the referring pathologist. Charts are reviewed for TNM stage, surgical and oncology treatments; ER, PR and HER2+ status (when available), and other tumor characteristics. Currently, we have collected medical record information on 100% of cases that includes a complete assessment of primary and adjuvant treatments.

Biologic Samples. The study collects germline DNA to evaluate inherited susceptibility genes for breast cancer. Pre-treatment tumor blocks are requested for each case, and used to create Tissue Microarrays (TMAs). The original H&E slides used to make the diagnosis of invasive breast cancer as well as recut H&E slides from the TMAs are reviewed by the study pathologists. Currently we have collected viable DNA samples on 98% of cases and have permission to obtain tumor blocks and/or slides on over 99% of cases. More than 2000 tumors have been processed. In addition to the tumor cores incorporated into the TMAs, separate tissue cores or slides are collected for more than 99% of the cases in CBCSIII.

Histology data. Digital images have been obtained using an Aperio Scanner for all CBCS cases. For most cases, H&Es were used to select regions for 1.0-mm coring, and up to 8-9 cores were obtained. A bottom section was also stained to evaluate presence of tumor in the cores. Four cores were used to construct tissue microarrays (TMAs) and four to five cores were preserved for molecular use.

IHC Data for Tumor Subtyping. Six markers are being assayed on CBCSIII TMAs to establish immunohistochemical tumor subtypes. ER, PR, HER2, EGFR, Ck5/6 and Ki67 are being quantitatively assessed using digital analysis with Tissue Studio and/or Genie. Positive/negative status at published thresholds are being used to categorize tumors into Luminal A (ER+ or PR+, Ki67 low), Luminal B (ER+ or PR+, Ki67 hi), HER2-enriched (HER2+), Basal-like (Ck5/6 or EGFR positive, hormone receptor negative).

Nanostring Data for Tumor Subtyping. Recent technological advancements have occurred and robust gene expression profiling is in progress using Nanostring methods (<http://nanostring.com/>)[3](#_ENREF_3),[4](#_ENREF_4); in this technology, fluorescently labeled and bar-coded probes are created and allowed to hybridize to specific RNA molecules/genes coming from fixed or fresh frozen materials. Next, hybridization is allowed to occur and the now labeled RNA target molecules are captured on a solid surface and then imaged and literally counted. Thus, the great advantage of this technology is that there is no enzymatic amplification or manipulation, and in the end the assay simply counts how many RNA molecules of a given gene were present. We have already developed a PAM50 Nanostring assay and applied it to 1200 tumors. These same tumors have been assayed for race-associated gene expression (using a set of genes identified by Troester & Perou), kinome reprogramming, p53 signaling, hypoxia signature, and a claudin-low predictor.

RNA Seq data. RNA-sequencing is in progress for 200 tumor-derived paraffin embedded samples, using 1-2 cores or microdissected slides from CBCSIII. We will age and stage match AA and Cau patients selected to minimize confounding by these variables. We plan to identify 50 basal-like AA and 50 basal-like Cau tumors, and 50 luminal AA and 50 luminal Cau tumors for RNA-sequencing.

Oncoarray data. In collaboration with the National Cancer Institute, GWAS genotyping is being performed on CBCS 1, 2, and 3 white and black cases and controls from CBCS 1 and 2 using the Illumina Oncochip array: <http://products.illumina.com/products/infinium-oncoarray-500K-beadchip-kit.html> <http://www.nature.com/icogs/primer/cogs-project-and-design-of-the-icogs-array/>. Another custom GWAS panel is also being applied for all African American cases through the AMBER consortium.

1. **BIOSPECIMEN PRIORITIES**

All biospecimen use for the Carolina Breast Cancer Study Phase 3 will be evaluated by the CBCS Steering Committee. The committee will use the following criteria to evaluate collaborative requests:

1. **Preliminary data.** Investigators must show that preliminary data from public data sources or previous literature are available to support the proposed investigation. For example, if biopsecimens are sought for molecular analyses, public data such as METABRIC or TCGA data should be used to first evaluate the specific markers proposed. Such analyses should be presented at the time of application for biospecimens.
2. **Impact.** The investigators must describe the impact of the research findings, clinically or for public health. In addition, investigators must indicate how the data collected will benefit the CBCS. Specific attention should be given to the aims of the CBCS and the unique characteristics of the CBCS study population, with the goal of leveraging the unique characteristics of the study.
3. **Funding.** The investigators must indicate the existing or planned funding sources for the project and show sufficient funding or plans to complete the proposed work. If the data collection will be phased (i.e. with part funded via one mechanism and later phases to be contingent upon preliminary data), power calculations or other justificaitons for the pilot study sample size should be provided.
4. **Investigator Productivity.** All investigators applying for use of CBCS biospecimens should detail their previous collaborations with epidemiologic studies and indicate their publication record relevant to these collaborations.

REFERENCES

**1.** Newman B, Moorman PG, Millikan R, et al. The Carolina Breast Cancer Study: integrating population-based epidemiology and molecular biology. *Breast Cancer Res Treat.* 1995;35(1):51-60.

**2.** Chen GK, Millikan RC, John EM, et al. The Potential for Enhancing the Power of Genetic Association Studies in African Americans through the Reuse of Existing Genotype Data. *PLoS Genet.* 2010;6(9).

**3.** Geiss GK, Bumgarner RE, Birditt B, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol.* Mar 2008;26(3):317-325.

**4.** Malkov VA, Serikawa KA, Balantac N, et al. Multiplexed measurements of gene signatures in different analytes using the Nanostring nCounter Assay System. *BMC Res Notes.* 2009;2:80.

**5.** Wang Y, Xia XQ, Jia Z, et al. In silico estimates of tissue components in surgical samples based on expression profiling data. *Cancer Res.* Aug 15;70(16):6448-6455.

**6.** Camp JT, Elloumi F, Roman-Perez E, et al. Interactions with fibroblasts are distinct in Basal-like and luminal breast cancers. *Mol Cancer Res.* Jan 2011;9(1):3-13.

**7.** Elloumi F, Hu Z, Li Y, et al. Systematic Bias in Genomic Classification Due to Contaminating Non-neoplastic Tissue in Breast Tumor Samples. *BMC Med Genomics.* 2011;4:54.