Editorial
Challenges in studying the etiology of breast cancer subtypes

Melissa A Troester[1,2] and Theresa Swift-Scanlan[2,3]

[1]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[2]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[3]School of Nursing, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

Corresponding author: Melissa A Troester, troester@unc.edu

See related research by Kwan *et al.*, http://breast-cancer-research.com/content/11/3/R31

Abstract

Research that classifies breast tumors into homogenous sub-groups could ultimately help to define public health prevention strategies for aggressive breast cancer subtypes. However, etiologic research on molecular breast cancer subtypes must overcome several challenges. Stratifying breast cancers into subgroups can reduce statistical power and, therefore, may require non-traditional analytical methods. Integrating results across studies is hampered by varying definitions of molecular subtypes, with some studies using triple negative status and others using specific markers to define basal-like cancers. In addition, triple negative and basal-like breast cancers appear to show strong associations with race, so the varied racial and ethnic composition of different datasets can make comparison across studies challenging. In spite of these challenges, some strong and consistent associations between triple negative or basal-like breast cancer and demographic variables are emerging, and there are hints that prevention strategies for this aggressive subtype of breast cancer may also be attainable.

Growing evidence supports the concept of breast cancer as a group of diseases with distinct etiologies. In a recent issue of *Breast Cancer Research*, Kwan and colleagues [1] report a pooled analysis of molecular subtypes in two prospective breast tumor datasets. Their analysis of 2,544 invasive breast cancers is a well-powered, case-only analysis of associations between established breast cancer risk factors and molecular subtypes of breast cancer [2,3]. Together with similar findings from other studies, the results suggest racial disparities in the etiology of breast cancer subtypes and indicate that prevention strategies may be possible for aggressive breast cancers.

Evidence suggesting that distinct molecular subtypes of breast cancer have different etiologies has developed primarily from case-control studies. Case-control studies compare risk factor distributions among breast cancer cases to distributions among controls. In case-control analyses, breast cancer cases can be taken as a whole or can be stratified by molecular subtype [4-7], but as the cases are divided into subgroups, the power to discover associations can be reduced. To address this challenge, a few studies have used an alternative study design, the case-only design, to study breast cancer subtypes [1,4,8-10]. Instead of defining controls as the referent group, case-only analyses compare risk factor distributions or pathology characteristics across subtypes (for example, comparing HER2-enriched, basal-like, or luminal B tumors to luminal A tumors). In case-only analyses, risk factor distributions that differ significantly across subgroups of cases are interpreted as etiologic heterogeneity. Case-only studies have a history of use for investigating gene-environment interactions and can be an efficient and powerful approach [11,12]. Thus, by pooled analysis of two studies and an efficient case-only design, Kwan and colleagues [1] have replicated previous significant findings and report novel interactions in the etiology of breast cancer subtypes.

Several previously reported interactions between risk factors and breast cancer subtypes are confirmed in the report by Kwan and colleagues [1]. Their work supports interactions between parity and breastfeeding in triple negative breast cancer (TNBC) and/or basal-like breast cancer (BBC) [4,5]. Collectively, the data suggest that breastfeeding may be particularly important for preventing aggressive subtypes of breast cancer. The analyses by Kwan and colleagues [1] also support previously reported links between TNBC/BBC and obesity [4,6]. However, there have been divergent results for different obesity measures within and across studies and in interaction with other exposures [4,6,9,10]. While associations between obesity and molecular subtypes remain uncertain, there is agreement across several studies regarding patterns of breast cancer incidence by age and race.

BBC = basal-like breast cancer; TNBC = triple negative breast cancer.

Many of the studies linking race and aggressive breast cancer subtypes have included a substantial percentage of African Americans, allowing strong inference about subtype and racial disparities. The pooled analysis of the Life After Cancer Epidemiology (LACE) and Pathways studies included just 6% African-American cases [1], but a significant association was still observed between TNBC and race. In fact, the association between African American race and TNBC was among the strongest associations observed in their study. While evidence for a shift toward more aggressive subtypes in African-American women is consistent across several studies and has been observed in Surveillance, Epidemiology, and End Results (SEER) data [13], uncertainty remains about the genetic and environmental factors that contribute to the associations between race and breast cancer subtype. Complex relationships between biology, socioeconomics, and race exist, and future research in racially diverse populations will be needed to disentangle these factors and address disparities in breast cancer.

In integrating findings across studies, perhaps the greatest challenge is a lack of agreement about molecular definitions for subtypes. TNBCs include not only tumors that are truly BBCs, but also tumors for which there were false negatives for one or more of the three markers (estrogen receptor, progesterone receptor, or HER2). BBCs are measured with specificity by adding positive markers such as cytokeratin 5/6 or epidermal growth factor receptor [14] and future studies may take advantage of a 50-gene quantitative PCR assay for molecular subtypes [15]. Nonetheless, specific markers for BBC are not presently available in most studies and the resulting misclassification can be substantial. In the Carolina Breast Cancer Study and the Polish Women's Health Study, only 60% and 66%, respectively, of TNBCs were BBCs when specific, positive markers were included in the definition of that subtype. Distinguishing true BBC from TNBC has had important implications for clinical prognosis [14], and while it has been argued that these distinctions may be more important in clinical than etiologic research [6], misclassification can be an important bias and deserves careful consideration. Kwan and colleagues [1] have emphasized the importance of these molecular definitions and propose future research on their dataset using specific BBC markers.

Research on the epidemiology of molecular subtypes is challenging. Ideal studies include biomarker data that allow classification of breast cancers into homogeneous subgroups, have large and racially diverse samples, and utilize efficient analytic strategies. This analysis by Kwan and colleagues [1] has contributed important data confirming that different breast cancer subtypes have unique profiles by reproductive factors, age, and race. This study and others like it hold the promise of defining novel prevention strategies for breast cancer subtypes that contribute to racial disparities.

Competing interests
The authors declare that they have no competing interests.

References
1. Kwan ML, Kushi LH, Weltzien E, Maring B, Kutner SE, Fulton RS, Lee MM, Ambrosone CB, Caan BJ: **Epidemiology of breast cancer subtypes in two prospective cohort studies of breast cancer survivors.** *Breast Cancer Res* 2009, **11:**R31.
2. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Børresen-Dale AL, Brown PO, Botstein D: **Molecular portraits of human breast tumours.** *Nature* 2000, **406:**747-752.
3. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lønning P, Børresen-Dale AL: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.** *Proc Natl Acad Sci USA* 2001, **98:**10869-10874.
4. Millikan RC, Newman B, Tse CK, Moorman PG, Conway K, Dressler LG, Smith LV, Labbok MH, Geradts J, Bensen JT, Jackson S, Nyante S, Livasy C, Carey L, Earp HS, Perou CM: **Epidemiology of basal-like breast cancer.** *Breast Cancer Res Treat* 2008, **109:**123-139.
5. Phipps AI, Malone KE, Porter PL, Daling JR, Li CI: **Reproductive and hormonal risk factors for postmenopausal luminal, HER-2-overexpressing, and triple-negative breast cancer.** *Cancer* 2008, **113:**1521-1526.
6. Phipps AI, Malone KE, Porter PL, Daling JR, Li CI: **Body size and risk of luminal, HER2-overexpressing, and triple-negative breast cancer in postmenopausal women.** *Cancer Epidemiol Biomarkers Prev* 2008, **17:**2078-2086.
7. Yang XR, Sherman ME, Rimm DL, Lissowska J, Brinton LA, Peplonska B, Hewitt SM, Anderson WF, Szeszenia-Dabrowska N, Bardin-Mikolajczak A, Zatonski W, Cartun R, Mandich D, Rymkiewicz G, Ligaj M, Lukaszek S, Kordek R, García-Closas M: **Differences in risk factors for breast cancer molecular subtypes in a population-based study.** *Cancer Epidemiol Biomarkers Prev* 2007, **16:**439-443.
8. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S, Deming SL, Geradts J, Cheang MC, Nielsen TO, Moorman PG, Earp HS, Millikan RC: **Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study.** *JAMA* 2006, **295:**2492-2502.
9. Stead LA, Lash TL, Sobieraj JE, Chi DD, Westrup JL, Charlot M, Blanchard RA, Lee JC, King TC, Rosenberg CL: **Triple-negative breast cancers are increased in black women regardless of age or body mass index.** *Breast Cancer Res* 2009, **11:**R18.
10. Trivers KF, Lund MJ, Porter PL, Liff JM, Flagg EW, Coates RJ, Eley JW: **The epidemiology of triple-negative breast cancer, including race.** *Cancer Causes Control* 2009 [Epub ahead of print].
11. Begg CB, Zhang ZF: **Statistical analysis of molecular epidemiology studies employing case-series.** *Cancer Epidemiol Biomarkers Prev* 1994, **3:**173-175.
12. Khoury MJ, Flanders WD: **Nontraditional epidemiologic approaches in the analysis of gene-environment interaction: case-control studies with no controls!** *Am J Epidemiol* 1996, **144:**207-213.
13. Anderson WF, Jatoi I, Devesa SS: **Distinct breast cancer incidence and prognostic patterns in the NCI's SEER program: suggesting a possible link between etiology and outcome.** *Breast Cancer Res Treat* 2005, **90:**127-137.
14. Cheang MC, Voduc D, Bajdik C, Leung S, McKinney S, Chia SK, Perou CM, Nielsen TO: **Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype.** *Clin Cancer Res* 2008, **14:**1368-1376.
15. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, Quackenbush JF, Stijleman IJ, Palazzo J, Marron JS, Nobel AB, Mardis E, Nielsen TO, Ellis MJ, Perou CM, Bernard PS: **Supervised risk predictor of breast cancer based on intrinsic subtypes.** *J Clin Oncol* 2009, **27:**1160-1167.