2014

Investigation of DNA and RNA markers by novel technologies demonstrates DNA content intratumoral heterogeneity and long non-coding RNA aberrations in breast tumors

Zhouwei Zhang
University of Vermont

INVESTIGATION OF DNA AND RNA MARKERS BY NOVEL TECHNOLOGIES DEMONSTRATES DNA CONTENT INTRATUMORAL HETEROGENEITY AND LONG NON-CODING RNA ABERRATIONS IN BREAST TUMORS

A Thesis Presented

by

Zhouwei Zhang

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements
for the Degree of Master of Science,
Specializing in Pathology

October, 2014

Accepted by the Faculty of the Graduate College, The University of Vermont, in partial fulfillment of the requirements for the degree of Master of Science, specializing in Pathology

Thesis Examination Committee:

_____ Advisor
Mark F. Evans, Ph.D.

_____ Advisor
Donald L. Weaver, Ph.D.

_____
Nicholas H. Heintz, Ph.D.

_____ Chairperson
Christopher S. Francklyn, Ph. D.

_____ Dean, Graduate College
Cynthia J. Forehand

July 29th, 2014

Abstract

BACKGROUND:

Breast cancer is the most commonly diagnosed cancer and second leading cancer death cause among females in the U.S.A. About 1 in 8 women in U.S will develop invasive breast cancer over the course of her lifetime. In 2013, 234,580 new invasive breast cancer cases are expected to occur in women within the US and approximately 64,640 non-invasive carcinomas *in situ* were diagnosed in 2013, most of which were ductal carcinoma *in situ* (DCIS). Along with technological advances, a wide variety of candidate biomarkers have been proposed for cancer diagnosis and prognosis, including DNA content and non-coding RNA. Current techniques for detecting DNA content abnormalities in formalin-fixed, paraffin-embedded (FFPE) tissue samples by flow cytometric analysis have used cells recovered from ≥50μm whole tissue sections. Here, in our first study, a novel core punch sampling method was investigated for assessing DNA content abnormalities and intratumoral heterogeneity in FFPE specimens. Secondly, long non-coding RNAs (lncRNAs) has been examined. LncRNA participates in a broad spectrum of biological activities by diverse mechanisms and its dysregulation is associated with tumorgenesis. Some lncRNAs may function as oncogenes (O) and others as tumor suppressor genes (TSG). To date, lncRNA has been investigated primarily by qRT-PCR and RNA sequencing. This study has examined the relationship of lncRNA expression patterns to breast tumor pathology by chromogenic *in situ* hybridization (CISH).

METHODS:

Firstly, FFPE breast carcinoma specimens were selectively targeted using 1.0 mm diameter punch needles. Extracted cores were assayed by flow cytometry using a modified-Headley method. Secondly, the lncRNA expression levels of 6 lncRNAs: HOTAIR, H19, KCNQ1OT1, MEG3, MALAT11 and Zfas1, was examined by RNAscope® CISH using FFPE breast tissue microarrays (TMAs) comprising normal adjacent epithelia (NA), DCIS, and invasive carcinoma (IC) from 46 patients. LncRNA associate polycomb complex protein EZH2 was evaluated by immunohistochemistry (IHC). LncRNA data was also compared to standard breast tumor data including ER, PR, Her2 and Ki67 IHC. SYSTAT version 11 statistical package was used to perform for all the tests.

RESULTS:

Following optimization experiments of the core punch flow cytometric approach, DNA index and percent S-phase fraction intratumoral heterogeneities were detected in 10/23 (44%) and 11/23 (47%) specimens respectively. The lncRNA CISH study utilized a TMA that contained 36 spots of NA breast tissues, 34 DCIS spots and 43 IC spots. HOTAIR CISH staining was significantly stronger in IC than DCIS ($p<0.001$) and NA spots ($p<0.001$). In DCIS, HOTAIR was correlated with Her2 ($p=0.03$) IHC. And in IC, the data suggest HOTAIR is a marker for high histological grade ($p=0.026$). H19 was rarely expressed in normal adjacent epithelial or tumor cells but was strongly expressed especially in inter-lobular stromal cells around invasive growths ($p<0.001$). H19 correlated with Ki67 IHC expression in DCIS, ($p=0.047$). KCNQ1OT1 expressed stronger in IC and DCIS than in NA ($p<0.001$), and was associated with Her2 ($p= 0.032$) in IC. No significant expressional difference was found in MEG3. MALAT1 stained strong universally and Zfas1 was very faint in all samples; as such neither of these was analyzed statistically. Polycomb protein EZH2 expressed differently among tissues but did not correlate with lncRNA levels.

CONCLUSION:

Core-punching is an effective alternative to whole specimen sectioning and shows that macro-level genomic heterogeneity is common even within a single FFPE block. The interrelationship of DNA content heterogeneity to other forms of heterogeneity requires further study. RNAscope CISH supports bright-field microscopy investigations of lncRNA expression in FFPE tissue specimens. HOTAIR, H19 and KCNQ1OT1 may be potential breast cancer biomarkers, both HOTAIR and H19 may be a marker for DCIS at increased risk of progression to invasive cancer. HOTAIR, in particular, may be a predictor for invasive cancer grade.

Citation

Material from this thesis has been published in the following form:

Zhang, Z., Weaver, D.L., Munjal, K., Evans, M.F.. (2014) Intratumoral DNA content heterogeneity in breast carcinomas demonstrated by core punch tissue sampling and flow cytometry. Journal of Clinical Pathology, 67, 821-824.

Acknowledgements

First and foremost, I would like to thank Dr. Mark Evans and Dr. Donald Weaver for offering me the opportunity to be your students to complete the graduate program in the Department of Pathology. I'd like to give a huge thank you to Dr. Evans, who is my primary advisor for my last two years. He is a patient, supportive and creative mentor who provides generous help to me to finish my study in UVM and I will be always proud of be his student. Many thank yous to Dr. Donald Weaver for being dedicated and encouraging advisor. He always gave me hand with professional support for my projects. I also would like to thank you to Dr. Nicholas Heintz and Dr. Christopher Francklyn for being excellent mentors throughout my graduate education. I could not have a better committee to head my defense.

I'd like to express my thanks to Mr. Scott Tighe and Dr. Roxana del Rio for training me with flow cytometry, Dr. Takamaru Ashikaga for insight into the statistical analyses of our results, Dr. Doug Taatjes and the Microscopy Imaging Center for all their help, and residents in Fletcher Allen Health Care for collecting tissue samples. Many thanks to Brenda Shinosky for helping me familiarize with the department, and Erin Montgomery for offering me an opportunity to be a teaching assistant for CMB program. Thank you to Mrs. Stephanie Phelps for helping me "head off danger" every time I had trouble speaking with our lovely undergraduate students as a teaching assistant. Special thanks to Dr. Alan Rubin for his great suggestions when I felt perplexed about my life and my future. He is always an excellent listener and mentor for me.

Last but not least, I shall give a huge thanks to my lab crew. Thank you Zhihua for speaking Chinese with me when I first got here and knew nothing about here, and teaching me experimental techniques, and to Chris for helping me for my lab computer, printer, finding everything I need and answering every question I have about U.S, and to Vanitha for teaching me IHC and sharing ideas at lunch time.

Dedication

This thesis could not have been completed without unwavering support from my family and friends. It is with much respect that I dedicate my graduate work to my parents, who always support me spiritually and physically for the previous 23 years, my girlfriend, who has been accompanying with me over four years since we were at college and she always encourages me to move forward, and couple of my friends and relatives, Xuejun, Qiuming, Xiao, Yang, who provide me generous help to get used to my new life here and shows me around.

Table of Contents

Page

Alternative Tissue Sampling Method to Detect Genetic Instability of Breast Carcinoma with Flow Cytometry

Long non-coding RNA Chromogenic *in situ* Hybridization Signal Patterns Correlate with Breast Tumor Pathology

List of Tables

Page

List of Figures

Page

Comprehensive Literature Review

Introduction

Breast cancer is the most commonly diagnosed cancer and second leading cause of cancer death among females in the U.S (CDC website). In 2013, 234,580 new invasive breast cancer cases are estimated to occur in women in U.S, whereas 2,240 cases are expected in men (Siegel, Naishadham et al. 2013). In addition to invasive breast cancer, around 64,640 new cases of non-invasive carcinoma *in situ* in women breast, most of which are ductal carcinoma *in situ* (DCIS), are predicted to be newly diagnosed in 2013 (Siegel, Naishadham et al. 2013). About 14% female cancer deaths in 2013 are estimated to result from breast cancer (Siegel, Naishadham et al. 2013). With the advancement of classification systems, screening techniques and therapeutics, more and more females have been diagnosed with breast cancer at younger ages and have received better treatment. Breast cancer caused female deaths have decreased steadily since 1990, with a dramatic decrease in women aged 20-69 years old (Lacey, Devesa et al. 2002; Bray, McCarron et al. 2004). Rather than invasive breast carcinoma, which is able to metastasize to regional lymph nodes or distant sites, DCIS is confined within the basement membrane of ducts and lobules and often co-exists with invasive cancer cells; it can eventually develop to invasive cancer. The treatment for DCIS has become a widespread medical issue because of malignant uncertainty (Ernster and Barclay 1997; Duffy, Agbaje et al. 2005; Virnig, Tuttle et al. 2010) and most patients get treated with surgery, breast irradiation and endocrine therapy (Baxter, Virnig et al. 2004), however, about 70% of DCIS have been found not to have a devastating impact on a patient's life (Page, Dupont et al. 1982; Eusebi, Foschini et al. 1989; Eusebi, Feudale et al. 1994).

Chromosomal instability (CIN), resulting from elevated rate of chromosome missegregation during mitosis, is a hallmark of cancer. DNA ploidy changes are one consequence of CIN and typically result in abnormal DNA content (Fridlyand, Snijders et al. 2006). Sporadic tumors likely

develop through processes involving chronic sub-clonal diversification and accumulation of genomic aberrations resulting in intratumoral heterogeneity, which is being unveiled by large-scale massively parallel sequencing and paving the path for personalized medicine(Campbell, Pleasance et al. 2008; Campbell, Yachida et al. 2010; Michor and Polyak 2010).

Long non-coding RNA (lncRNA), transcribed RNA molecules greater than 200 nucleotides in length, has been demonstrated to be involved in multiple biological activities including epigenetics regulation, transcriptional regulation, interaction with small non-coding RNA, post-transcriptional regulation, nuclear compartment formation and cell cycle control (Gupta, Shah et al. 2010; Hung, Wang et al. 2011; Tsuiji, Yoshimoto et al. 2011; Schor, Lleres et al. 2012; Han, Liu et al. 2013; Luo, Li et al. 2013; Tripathi, Shen et al. 2013). Both DNA content abnormalities (Pinto, Monteiro et al. 2005; Bremmer, Brakenhoff et al. 2011) and lncRNAs (Dhanasekaran, Barrette et al. 2001; He, Bao et al. 2014; Zhao, Guo et al. 2014) have already been suggested as biomarkers in some cancer types, but this has not yet been investigated in DCIS. Understanding the potential role of CIN, intratumor heterogeneity and lncRNAs in breast tumor pathology would be valuable to science and medicine by creating better diagnostic methods, screening options and treatments.

Breast cancer symptoms

Early breast cancer doesn't cause any significant symptoms, so regular breast test is required. Although patients are not able to notice a lump on breast, which is a very typical symptom for a breast tumor, people can still pay attention to subtle changes to the body to get some clues for early disease. These signs include subtle changes in breast size and shape, some morphological or color changes in skin, for example, dimpling, swelling or becoming red. As a tumor grows, apparent symptoms can be seen, including a breast lump or lump in the armpit that are hard and have uneven edges, more visible changes in breast size, shape, or a recent nipple change, as

example, nipple inversion, nipple ulceration, retraction, bloody discharge as well as peeling, However, more than 80% breast cancer case are detected by breast lump (Therapy 2003). When cancer advances, there are a more symptoms that can be noticed by both patients and clinicians. Bone pain, breast discomfort, skin ulcers, swelling of one arm and weight loss are important features at that stage. According to different presentations of breast cancer in different stages, a complete test that helps doctors determines the exact stage that is necessary. Physical exams include breasts, armpits, and the neck and chest area and a variety of methods are combined. For instance, magnetic resonance imaging (MRI) scan can provide clinicians with better identification of a breast lump or an abnormal mammogram changes, ultrasound can be utilized to show whether the lump is solid or not, and biopsies help pathologists to stage or come up with a treatment decision. Computerized tomography (CT) scans can contribute to analyzing the tumor spread. To test a very subtle change in breast contour, the test must contain one assessment of the breast with patients upright with arms raised. If some symptoms like contour change, skin tethering, dilated veins, ulceration, or Paget's disease are found then clinicians become more concerned about the patient. Paget's disease, whose symptoms include eczematoid skin, is diagnosed in almost a quarter of the female patients (NCI 2005). Additionally, tests on axillae and supraclavicular fossae areas and additional abdominal and neurologic examination should be included to make a complete evaluation because there might be some symptoms indicating metastasis occurring in such places. These symptoms include breathing difficulties, bone pain, symptoms of hypercalcemia, abdominal distention, jaundice, localizing neurologic signs, and altered cognitive function.

Though a lump is an indicator for breast tumor, it is quite indiscernible and hard to detect inflammatory breast cancer, which can pose a substantial diagnostic challenge. For diagnostic purposes, people who may have this type of cancer should pay attention to symptoms that

resemble a breast inflammation as well as itching, pain, swelling, nipple inversion, warmth and redness throughout the breast. Moreover, *peau d'orange*, which is an orange-peel texture on skin could be another sign. Occasionally localized breast cancer cells metastasize to other locations through the body including bone, brain, liver and lung. Unexplained weight loss and fever sometimes presage occult breast cancer and pains in the bones or joints could also manifest metastatic breast cancer, however all these symptoms are non-specific, meaning they could also be indications of other diseases.

Breast cancer histology features and classification

The accurate diagnosis and pathological assessment are key steps undertaken by pathologists, who need to differentiate benign breast tissues from early and established breast cancers. Then, an evaluation of pathological features and suggestions for treatment should be provided by pathologists (Russo and Russo 1992). Carcinomas are most common malignant tumor types for breast tissues and can be divided into two groups: carcinoma *in situ* and invasive carcinoma. And since most breast carcinomas are derived from epithelium of ducts or lobules, we also define them as mammary ductal carcinoma and lobular carcinoma. By histopathological evaluation, the three most frequent carcinoma types are invasive ductal carcinoma (55%), DCIS (13%) and invasive lobular carcinoma (5%) comprising about approximately 75% of all incidence (Eheman, Shaw et al. 2009). Invasive carcinoma usually presents with a large breast mass and sometimes with nipple discharges or breast pain. When observed by microscopy, invasive carcinoma is more fixed asymmetrically and not well circumscribed. However, nearly all triple negative tumors present solid architectures without forming tubules and have a large amount of tumor cells with little stromal area between these cells, demonstrating very detailed morphological difference. In addition, above half of basal-like tumors exhibit a pushing border and have stromal lymphocytic infiltration at the edge of tumor to some extent (Livasy, Karaca et al. 2006). DCIS, with the exception of neoplastic epithelial proliferation within the ducts basement membranes shares some

symptoms with invasive cancer and surprisingly, many similarities between DCIS cells and invasive carcinoma cells are found, on both cellular and molecular levels (Ma, Salunga et al. 2003; Hannemann, Velds et al. 2006; Kuerer, Albarracin et al. 2009).

Immunohistochemistry (IHC) and fluorescent *in situ* hybridization (FISH) are valuable tools for histological assessment to identify estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (Her2) expression status. Sentinel lymph node biopsy is another method to aid histological assess, especially for metastatic status (Veronesi, Paganelli et al. 1997).The nearest single lymph node to primary breast tumor is removed and examined in detail to see if there are some tiny clusters of metastatic cells. If no tumor cells are found, pathologists can predict a very low chance to get metastatic cells in axillary lymph nodes, and therefore, full lymph node excision is not necessary. The classification of the invasive breast cancer aims to classify breast cancers into various categories according to different criteria in order to primarily help select the best therapeutics and provide a better prognosis. In 2003, the World Health Organization recommended a comprehensive breast cancer classification system including both non-malignant and malignant tumors to help clinicians to better identify tumor types and specific treatments (Tavassoli, Devilee et al. 2003). Histopathology, tumor grade, tumor stage and receptor expression status are most commonly used criteria for classification. Tumor grade relies on the similarity and differences between cancer cells and normal cells and thus divides tumor into three major groups: low grade (well-differentiated), intermediate grade (moderately differentiated) and high grade (poorly differentiated). Usually well-differentiated tumor has a better prognosis due to its similar appearance as normal cells. The overall grade is assessed by the Nottingham system (Genestie, Zafrani et al. 1998; Simpson, Gray et al. 2000), which generates the overall scores by totalizing scores for nuclear features, tubule formation and mitotic activity, each of which is scored from 1 to 3 points. Nuclear feature is used to assess how

much variation tumor cell nuclei have compared with normal breast cells nuclei, while tubule formation is a parameter that evaluates how much tumor has the normal structure of ducts. Uncontrolled cell division is one of the hallmarks of cancer, so mitotic count is a parameter to assess how many dividing cells can be seen in 10x microscope fields. The sum of three points can lead to a conclusion: 3-5 points means grade 1 tumor which is the best among the three, 6-7 represents grade 2 tumor and 8-9 points is grade 3 tumor. Tumor stage is a criterion which determines how severe the cancer is and often come along with estrogen/ progesterone receptor expression levels, Her2 status and menopausal status for a diagnostic decision. At present, the tumor, nodes and metastases (TNM) staging system is the most widely used in world. Tumor value is based on the primary cancer and lymph node value depends on cell number, size in regional lymph nodes, and metastases value refers to the information of metastatic cancer. Both grade and staging units contribute to The Nottingham Prognostic Index (NPI) (Galea, Blamey et al. 1992). NPI= (0.2 x tumor size (cm)) + tumor grade + stage, is an equation to stratify patients into different prognostic group, due to which, favorable prognosis has scores below 3.4 and medium gets scores between 3.41 to 5.4. However, poor prognosis has a score greater than 5.41 (Galea, Blamey et al. 1992).

IHC and FISH tests for ER, PR, and Her2 are the most common methods to help classify breast cancers and of great importance as a guide for therapeutics. Based on different receptor expression, most breast cancers are categorized to five groups including luminal A/B, Her2+, triple-negative Claudin-low and normal basal-like (Perou, Sorlie et al. 2000). Estrogen receptor positive (ER+) cancer cells can be treated with tamoxifen (Jordan and Koerner 1975) or aromatase inhibitors (AIs) (Smith and Dowsett 2003)to reduce estrogen effect or decrease ER expression level because these cells need estrogen to grow, while monoclonal antibody trastuzumab can precisely target Her2 positive cancer cells and thus highly improve breast cancer

therapeutics combining with traditional chemotherapy (Romond, Perez et al. 2005). In terms of DCIS, the current pre-mammography classification is predominately relied on microscopic features: comedo, solid, papillary, micropapillary and cribriform (Allred 2010). The grading system refers to the degree DCIS cells resemble normal cells, the level DCIS cells differentiate, conveying criteria from existing invasive breast cancer histology grading pattern since we know a positive relationship between tumor cell differentiation and cancer aggressiveness (Elston and Ellis 1993).

Genetics

BRCA1 and BRCA2 are most important hereditary genes in breast cancer genetics. More than 200 germline mutations in BRCA1 and over 100 mutations in BRCA2 shows clear associations with breast cancer susceptibility and have been registered in the Breast Cancer Information Core Database (Szabo, Masiello et al. 2000). Mutations of BRCA1 and BRCA2 have been reported to account for almost 20-25% familial breast cancer cases but less than 10% of overall breast cancers. The majority of known mutations in BRCA1 and BRCA2 are predicted to result in premature translation termination (Easton 1999). BRCA1 encodes a 220kD nuclear protein and its molecular functions are to control cell cycle (i.e., functional BRCA1 missing will cause cell cycle arrest) and responses to DNA damage (Gowen, Avrutskaya et al. 1998; Scully and Livingston 2000). In the meantime, BRCA1 is also a component of the RAD51-MRE11-p95 complex that works for DNA repair system (Scully, Chen et al. 1997). BRCA2 encodes a bigger protein which also involves in DNA double-strand break repair and maintains chromosome integrity (Chen, Silver et al. 1999). Approximately 15-20% women have been found with BRCA1 mutation if someone had breast cancer in their family history and the number is 60-80% when their families showed both breast and ovarian cancer (Couch, DeShano et al. 1997). These mutations carriers have a chance of 60-80% to get breast cancer through their lifetime, with a median diagnosed age of 20 years as compared to those women who do not have BRCA1

mutations (Easton, Narod et al. 1994; Struewing, Tarone et al. 1996). Females who carry BRCA2 mutation will also have 60-80% chance to get cancer eventually, while men with BRCA2 mutations will surprisingly develop a 6% lifetime breast cancer risk. It has been found those BRCA associated cancers appears in patients' at younger ages and are more dangerous. BRCA1 mutations have been shown to associate with triple negative breast cancer (TNBC), while mutations in BRCA2 are more likely to be found in post-menopausal breast cancer.

Scientists have also found several germline genetic mutations that contribute to breast cancer risk by analyzing familial data, including STK11/LKB1, phosphatase and tensin homolog (PTEN), androgen receptor (AR) and p53 (Malkin, Li et al. 1990; Wooster, Mangion et al. 1992; Liaw, Marsh et al. 1997; Boardman, Thibodeau et al. 1998). However, all these genetic factors mentioned above are rarely found across the population so that they can only account for a very small portion of heritability of human breast cancer. As techniques develop, the secrets of breast cancer genetic world are rapidly getting dissected. With comparative genomic hybridization (CGH), patterns of chromosome gain and loss have characterized breast cancer of different grades (Roylance, Gorman et al. 1999). Losses of 11q, 8p, 13q, gains on 1q, 8q, 17q are high grade invasive breast tumor signatures whereas low grade cancers commonly show gains on 1q, 16p, 8q and loses of 16q (Buerger, Otterbach et al. 1999; Roylance, Gorman et al. 1999). Noticeably, DCIS shows high similarities to genetic abnormalities found in invasive cancers (Buerger, Otterbach et al. 1999). Advanced high-throughput sequencing have identified many susceptible loci in candidate genes, single nucleotide polymorphisms (SNP) and copy number variations (CNV) in somatic cells that are highly associated with breast cancer. MAP3K1, AKT2 and CDKN1B are newly found genes that contribute to breast cancer susceptibility (Stephens, Tarpey et al. 2012). Together with meta-analysis, results from different genome–wide association studies can by synthesized to better identify some widely-accepted loci in different race, e.g., 6q14 and 20q11 (Siddiq, Couch et al. 2012).

Screening

Screening for breast cancer is for apparent healthy women to get an early diagnosis, which will largely improve the final outcomes. Several tests have been utilized including self and clinical breast exams, mammography, genetic screening, MRI and ultrasound. In 2003, a review demonstrated that breast cancer examination did not associate with lower death rates and so did not recommend self or clinical exams (Kosters and Gotzsche 2003). Mammography utilizes a specialized X-ray machine which emits a small amount of ionizing radiation to get the X-ray image, which will be on plain digital mammography or photographic film on a computer screen, interpreted by radiologists and now it becomes a very common breast cancer screening test among women ages 40 to 74 because of its quickness and wide availability (Mandelblatt, Cronin et al. 2009). Overall, mammography saves a small number of lives and is more effective in older aged women but it has no benefits for predicting the outcome of that cancer. Recommendations about what time and what frequency it is best to undergo mammography screening vary across the world. Although mammography offers a small, but statistically significant benefit, it also creates some criticism because many patients overestimate the effect that mammography provides and this often results in heavy psychological and financial burden. MRI is another tool in screening breast cancer with a very high negative predictive value and it can also diagnose benign proliferative changes, fibroadenomas and some other benign findings. In terms of shortcomings, it is much less specific than mammography, more expensive and less available in some developing countries. The third screening method is genetic test, which can only reveal a susceptibility to develop cancer rather than detect cancer. Due to significant role of BRCA mutation in promoting breast cancer, US government published a clinical practice guideline to recommend women for BRCA mutation test (Nelson, Huffman et al. 2005). About 2% of American women have family histories have been found with an increased risk of having a medically significant BRCA mutation after genetic test (Nelson, Huffman et al. 2005).

Therapeutics

The mainstay of breast cancer treatment is surgery if the tumor is localized and the following treatments are chemotherapy, radiotherapy and hormonal therapy. There is no general treatment for all kinds of breast cancer so it is very important to identify what specific molecular subtype the patient has beforehand. TNBC has been given first priority currently by clinicians around the world because it has the worst prognosis among all breast cancer subtypes. Multiple clinical trials have been tried on TNBC patients and it has already been found that several chemotherapeutic drugs and biological agents are very promising. DNA-damaging compounds like platinate agents should be useful because platinum compounds function in developing a covalent bifunctional cross-linked adducts, which blocks DNA double-strand break caused by replication forks (Helleday, Petermann et al. 2008; Bosch, Eroles et al. 2010). Epidermal growth factor receptor (EGFR) is another therapeutic target because it is expressed in 45-70% TNBC cases, but target antibody or drugs including Cetuximab still await assessment (O'Shaughnessy 2007). Angiogenesis is critical for tumor growth so drugs that specifically target neoangiogenesis is another treatment option. Bevacizumab was approved by FDA to treat metastatic Her2-negative breast cancer in combination with paclitaxel due to results of the phase III clinical trial E2100 (Miller, Wang et al. 2007). Another clinical trial that tested Bevacizumab detected a statistically significant difference in progression-free survival between women taking bevacizumab and docetaxel (Miles D 2008) .Last but not at least, PARP inhibitors are compounds that work to inhibit already defective DNA repair, which could be potentially used.

Lumina A type, which is a low grade breast cancer that is ER-positive and Her2-negative, has the best prognosis among all the subgroups. It seems there is no standard treatment for luminal A and it depends on an individual's traits, tumor size, histological grade and lymph node conditions. Luminal B type is a higher grade breast cancer that is ER-positive and Her2-negative, and a potential target is insulin-like growth factor signaling, to which clinicians adopt IGF-1R antibody

or small molecule inhibitors to block this pathway (Atzori, Traina et al. 2009). In addition, endocrine therapy plus PI3K inhibitors dramatically improve treatment efficiency for luminal B breast cancer, while highly specific PI3K inhibitors are in development to reduce unexpected side effect (Creighton, Fu et al. 2010). Patients that are Her2-positive are likely to have poorly differentiated tumors with a relatively high level proliferation rate and associated with increasing risk of recurrence and death. Trastuzumab, which is an engineered antibody with a high affinity for Her2 extracellular transmembrane protein, has had a major impact in the treatment of Her2-positive metastatic breast cancer with cytotoxic agents (Hortobagyi 2005). In terms of treating DCIS, there are three major options: 1) Breast-conserving surgery and radiation therapy with or without tamoxifen; 2) Total mastectomy with or without tamoxifen; 3) Breast-conserving surgery without radiation therapy, but the results are pending because of poor accuracy.

Biomarkers for breast cancer

Cancer biomarker refers to any molecule secreted by tumors or responses of body that are indicative of the presence of the cancer. Genetics, epigenetics, proteomics, metabolomics, glycomic as well as imaging biomarkers can be applied to cancer diagnosis, prognosis and other aspects. For breast cancer, although classical histopathological features play important roles in diagnosis, prognosis and treatments, in the era of big data, several novel biomarkers have been proposed and validated to improve clinical prediction and have become clinical routine tests for breast cancer. Genetic alternations in either BRCA1 or BRCA2 seem to account for about 20-25% of familial cancer incidences, which consists approximately 5-10% all breast cancer cases (Easton 1999). Particularly the 3 major mutations are: 185delAG, 5382insC in BRCA1 and 6174delT in BRCA2. BRCA1 mutations carrier have a dramatically lower short-term and long-term survival rates and a significantly lower progression-free survival rate, while BRCA2 mutation does affect neither short-term nor long-term survival rate but associates with an increased risk of pancreatic cancer (Lee, Park et al. 2010). Estrogen receptor (ER) is the most important biomarker in breast

cancer and its status directs the therapy option to endocrine therapies because all the ER-positive tumors use steroid hormone estradiol as their primary growth stimulus. Progesterone receptor (PR) expression is dependent on the present of ER and there is <1% of breast cancer cases express only PR but not ER. Considering breast cancer early detection, serum tumor marker like carcinoembryonic antigen (CA) has not been demonstrated to be sensitive for early detection (Hayes 1996), while mammaglobin and maspin seems to be promising for that purpose (Maass, Nagasaki et al. 2002; O'Brien, Maguire et al. 2002). From the point of view of diagnosis, cytologic examinations for nipple duct fluid has used for decades and one study which applied protein chip method to analyze nipple fluid revealed some potential biomarkers, e.g., a 15940-Da protein was found with a 80% sensitivity and 100% specificity (King, Chew et al. 1983; Sauter, Zhu et al. 2002). Immunohistochemical staining for other glycoproteins such as B72.3, a-lactalbumin and milk fat globule, showed their potential in identifying metastasis in breast cancer (Lee, DeLellis et al. 1984; Hilborne, Cheng et al. 1986). Furthermore, the functions of DNA ploidy and S-phase fraction in predicting prognosis have been put forward a long time ago but vary greatly among studies. Currently, neither the American Society of Clinical Oncologists (Bast, Ravdin et al. 2001) nor the College of American Pathologists (Hammond, Fitzgibbons et al. 2000) recommend ploidy status and S-phase fraction as standalone prognostic marker. Ki-67 staining is a common method to reveal cell proliferation and it is more consistent than ploidy and S-phase fraction measurement alone. The epidermal growth factor receptor (EGFR) are overexpressed widely among breast cancers and famous Her2/neu is a member of EGFR and amplified in 10-34% invasive breast cancer (Ross and Fletcher 1999). Both molecular techniques and morphology techniques are applied to measure Her2/neu status clinically. Immunohistochemistry staining, fluorescent *in situ* hybridization, southern blot, RT-PCR, and enzyme-linked immunosorbent assay (ELISA) are common methods to test Her2/neu but none of these methods is perfect (Hillig, Thode et al. 2012). Additionally, Her2/neu is also an important marker in monitoring treatment

for trastuzumab (Esteva, Cheli et al. 2005). Besides EGFR, other growth factor like transforming growth factor a (TGF-α), TGF-β, insulin-like growth factor (IGF) - I and –II, platelet-derived growth factor (PDGF), fibroblast growth factor (FGF) and vascular endothelial growth factor (VEFG) are all associated with breast cancer prognosis to some extent (Bonneterre, Peyrat et al. 1990; Castellani, Visscher et al. 1994; Yiangou, Gomm et al. 1997; Dumont and Arteaga 2000; Shao, Nguyen et al. 2000; Linderholm, Lindahl et al. 2001).

Genomic instability and flow cytometry

Genomic instability indicates an increased rate of genomic aberrations of a cellular lineage and it is believed to be necessary for carcinogenesis. Those alternations include nucleic acid sequence changes, chromosomal rearrangements and aneuploidy. Genomic instability was first characterized in colorectal cancers in two groups: microsatellite instability is due to defects in DNA mismatch repair, causing mutations in gene sequences at simple repeats; the other type, which is much more common, is called chromosome instability (CIN) (Lengauer, Kinzler et al. 1997; Lengauer, Kinzler et al. 1998). The underlying mechanism is not well known but it likely results from elevated rate of chromosome missegregation during mitosis. Telomere dysfunction has also been presented as another mechanism for genomic instability (Artandi and DePinho 2000). Traditionally, many breast genomic alternations were described as gains and losses, including DNA amplifications at 11q13, 17q12, 8p12 and 8q24 from cytogenetic analyses, FISH and CGH (Gray, Collins et al. 1994). The advent of array CGH improves the analysis and contributes to identify copy number variation, single nucleotide polymorphism at a high-resolution genomic profile level (Loo, Grove et al. 2004; Bergamaschi, Kim et al. 2006; Staaf, Jonsson et al. 2011; Krepischi, Achatz et al. 2012). Aneuploidy, a type of chromosomal abnormality, represented by numerical changes in whole chromosomes, is usually indicated by abnormal DNA content. Aneuploidy consistently occurs in all cancers but is less well studied than structural chromosome alternations. Since genes and pathways which are deregulated by

aneuploidy are still unknown, it is possible that these genomic alternations have a more complex effect on carcinogenesis. In aneuploid chromosomes, candidate genes are more difficult to identify due to big genomic regions and potential gene-gene interactions (Gordon, Resio et al. 2012). Flow cytometric measurement of nuclear DNA content renders reliable information about DNA ploidy and estimation of tumor cell genetic instability. DNA index, which is the ratio of tumor sample / standard DNA fluorescence channel measured by flow cytometer, is frequently applied to show DNA content difference among tumors, has become a prognostic and diagnostic marker in cancers (Pradhan, Abeler et al. 2012; Giaretti, Monteghirfo et al. 2013). Initially, dyes that binds double strand DNA must be added into single cells suspension. The scheme is that the stained cells have incorporated an amount of dye proportional to the amount of DNA measured in the flow cytometer and the emitted fluorescent signal yields an electronic pulse that is proportional to the total fluorescence emission from the cell. Considering the scatter provided by the cytometer, forward scatter is designed to measure cell size, while side scatter is used to indicate cellular complexity and granularity. This combination of scattered and fluorescent light is picked up by the detectors, and, by analyzing fluctuations in brightness at each detector, it is then possible to collect various types of information about the physical and chemical structure of each individual particle. The fluid rate should not be too high in order to yield a good signal of discrimination between singlets or doublets and the final collected cell number should be more than 5000. As DNA content could be not shown directly by flow data and thus reference cells for example human red blood cells, chicken red blood cells or some external control, should be included every time to help identify the positions of normal diploid DNA.

Long non-coding RNA

Long non-coding RNA (lncRNA) is a transcribed non-coding RNA molecule greater than 200 nucleotides in length. LncRNA is an emerging focus in biomedical research. Studies have now demonstrated or implicated lncRNAs as important participants in a wide spectrum of processes in

the normal or abnormal development in organisms even though it was termed as “junk” or “transcriptional noise” several decades ago. Scientists had found some lower order animals have much larger genomes than that of higher animals including humans. The C-value paradox: why the amount of DNA in the haploid genome does not correspond with organism size or developmental complexity, was firstly solved by the discovery of non-protein-coding transcripts (Thomas 1971). In 1970s, with the advancement of biotechnique, e.g., DNA-RNA hybridization, it has been suggested that human would not have >20,000 genes and the rest space of human genome were mostly possessed by noncoding genes, which was called “junk DNA” at that time (Comings 1972). However, massively distributed non-coding genes also brought scientists another hypothesis: junk DNA may be useful, and a number of early hypothesized functions were also proposed, including genome integrity, gene regulation, mRNA processing, etc. (Britten and Davidson 1971; John and Miklos 1979; Lewin 1982). More transcripts that do not account for coding sequences have been found, e.g., tRNAs, rRNAs, snoRNAs in 1970s and 1980s but scientists could not elaborate non-coding genes until late 1990s, with the invention of whole-genome sequencing. Now, it has been estimated that approximately 70-90% of human genome will be transcribed at some points during development in spite of low inter-species conservations and some low expression transcripts (Wang, Zhang et al. 2004; Djebali, Davis et al. 2012; Mercer, Gerhardt et al. 2012); conserved lncRNAs number only a few thousand (Ponjavic, Ponting et al. 2007). It might be because lncRNAs don't require very much nucleotide sequence conservation to maintain their functionality, as compared with protein coding genes which are under strong selection restraints to maintain codons or open reading frames (Ponting, Oliver et al. 2009). However, one recent paper calculated and found a higher sequence conservation in lncRNA promoters than that in protein coding genes promoters in mice, indicating although lncRNA sequences might not be highly conserved, the level of their transcription is (Carninci, Kasukawa

et al. 2005). To classify a large magnitude of lncRNAs, one of useful manners is to sort according to their genomic locations, by which we can categorize them into five distinct groups:

1) Stand-alone lncRNAs, which are located in small or large interspace of protein-coding genes (Cabili, Trapnell et al. 2011).
2) Pseudogenes, which are extra genes copies that have lost their protein-coding potential. Only a small portion of pseudogenes are transcribed and also have acquired function during resurrection (Bekpen, Marques-Bonet et al. 2009).
3) Natural antisense transcripts (NAT), which are derived from the opposite strand to sense DNA strand. About 70% of sense transcripts have been found with some antisense counterparts (He, Vogelstein et al. 2008). Sense-antisense pairs can be formed by two coding mRNAs, coding/non-coding RNAs or dual non-coding RNAs.
4) Intronic non-coding RNAs, which are harbored within introns.
5) Transcription elements associated transcripts, which are non-coding RNAs produced or processed within or near the sequence of promoter, enhancer or transcription start site in both sense and antisense direction.

From various studies in lncRNAs in last decade, it is now obvious that lncRNAs contribute to a variety of developmental processes and diseases, but mostly the molecular mechanism details that lncRNAs employ have not been demonstrated or verified. Based on a few relatively well-studied examples, we can still distill the functions into several types though many lncRNAs may be implicated in various mechanisms. One of the major themes is about the role of lncRNAs in epigenetics. LncRNAs of this group can both interact with chromatin in *cis*, regulating genes nearby; or in *trans*, functioning towards distant genes. Chromatin-modifying elements such as polycomb repressive complex 2 (PRC2) recently have been found to interact with multiple lncRNAs recently (Khalil, Guttman et al. 2009; Tsai, Manor et al. 2010; Aguilo, Zhou et al. 2011;

Guil, Soler et al. 2012). HOTAIR, an lncRNA transcribed from HOXC cluster, is one of the best-studied lncRNAs that dictates methylation of H3 on K27 and thus repress genes expression by interacting with PCR2 (Rinn, Kertesz et al. 2007). Not only PRC2, HOTAIR was also found to bind a second complex containing CoREST, REST and LSD1 in a different location to demethylation H3 on K4 and thereby inhibit target gene activation (Tsai, Manor et al. 2010). Similarly, lncRNAs KCNQ1OT1 and Air recruit chromatin-modifying complex to silence multiple gene in their imprinted gene domains, respectively. KCNQ1OT1 interacts with PRC2 and G9a, both of which are histone methlytransferases, in *cis* to repress gene expression, while Air only brings G9a to its target promoter to form a repressive domain (Nagano, Mitchell et al. 2008; Mohammad, Mondal et al. 2009).

X chromosome inactivation (XCI) inactivates one X in female cells to equalize gene expression between males and females. The XCI process is largely controlled by a lncRNAs enriched cluster known as the X-inactivation center. Xist, a 17kb transcript, "coats" X chromosome and recruit silencing complex including PRC2, resulting in a chromosome-wide gene inhibition (Zhao, Sun et al. 2008).

Another major function that lncRNAs have is to regulate transcription directly. LncRNA can act as a molecular decoy, by which lncRNA binds to specific transcription factor or other regulatory factors, which is necessary for target gene transcription initiation. PANDA, a p53 dependent lncRNA in cell cycle, works as a decoy to sequester transcription factor NF-YA away from its target genes to undergo cell cycle arrest (Hung, Wang et al. 2011).. Another subtype lncRNA of this class works directly with Pol II. An upstream minor promoter of dihydrofolate reductase (DHFR) produces a lncRNA that inhibits assembling of pre-transcription elements at the major promoter via a mechanism of forming DNA: ncRNA complex or binding general transcription factor II (Martianov, Ramadass et al. 2007).

LncRNAs can also affect the nuclear compartment. Nuclear enriched abundant transcript 1(NEAT1) associates with multiple paraspeckle proteins to stabilize paraspeckles (Chen and Carmichael 2009; Clemson, Hutchinson et al. 2009). Metastasis-associated lung adenocarcinoma transcript 1 (MALAT1), another very abundant lncRNA localized in nucleus, binds serine/arginine splicing factors. MALAT1 accounts for relocating those splicing factor to transcription start sites, where mRNA precursors get cleaved (Tripathi, Ellis et al. 2010). LncRNAs could also exert their functions in post-transcriptional steps. MALAT1, Gomafu/MIAT and some natural antisense transcript all play roles in mRNA splicing, as examples of post-transcriptional regulation (Sone, Hayashi et al. 2007). In addition to process mRNA, lncRNA may even be able to impact mRNA stability and protein translation (Gong and Maquat 2011; Yoon, Abdelmohsen et al. 2012). To date, lncRNA expression analyses predominately studied by qPCR have reveled association between lncRNA dysregulation and diseases, most notably cancer. Although most lncRNAs resides in the nucleus, a proportion of lncRNAs localize within or are transported to cytoplasm to regulate proteins translation, localization or mRNA stability (Kapranov, Cheng et al. 2007). LncRNA NRON prevents the traffic of nuclear factor of activated T-cells (NFAT) transcription factor to nucleus from cytoplasm where it activates target genes in response to calcium-dependent signals (Willingham, Orth et al. 2005) . Cytoplasmic lncRNAs also are capable to base pair with mRNA and thereby regulate mRNA transcription and protein translation. For example, UCHL1 mRNA antisense, which is a lncRNA, complements to UCHL1 AUG initiation codon and combined inverted SINEB2 domains to enhance UCHL1 protein synthesis in cytoplasm(Carrieri, Cimatti et al. 2012); PTENP1, a pseudogene gene of tumor suppressor gene PTEN, has the same microRNA binding sequences in 3'UTR as that of PTEN, leading to a microRNA binding competition between PTEN and PTENP1, resulting in PTEN gene activation and translation (Poliseno, Salmena et al. 2010).

Table 1. Current cancer-related lncRNAs

lncRNA	Cancer	Molecular Function	Reference
HOTAIR	Multiple cancers	Epigenetic regulation	(Gupta, Shah et al. 2010; Kogo, Shimamura et al. 2011; Yang, Zhou et al. 2011; Chen, Sun et al. 2013; Zhang, Han et al. 2013)
MALAT1	Multiple cancers	mRNA splicing	(Ji, Diederichs et al. 2003; Luo, Ren et al. 2006; Yamada, Kano et al. 2006; Fellenberg, Bernd et al. 2007)
H19	Multiple cancers	Epigenetic regulation	(Hibi, Nakamura et al. 1996; Berteaux, Lottin et al. 2005; Fellig, Ariel et al. 2005; Tsang, Ng et al. 2010)
MEG3	Multiple cancers	P53 activation	(Zhang, Gejman et al. 2010; Braconi, Kogure et al. 2011; Jia, Wei et al. 2013; Lu, Li et al. 2013; Sun, Xia et al. 2013)
KCNQ1OT1	Breast, colon	Epigenetic regulation	(Tanaka, Shiota et al. 2001; Rodriguez, Weng et al. 2011)
Zfas1	Breast	NA	(Askarian-Amiri, Crawford et al. 2011)
Gas5	Breast, pancreas	Decoy of glucorticoid receptor(GR)	(Mourtada-Maarabouni, Pickard et al. 2009; Lu, Fang et al. 2013)
HULU	Multiple cancer	Decoy of microRNA	(Panzitt, Tschernatsch et al. 2007; Matouk, Abbasi et al. 2009; Zhao, Guo et al. 2014)
ANRIL	Prostate, leukemia	Epigenetic regulation	(Yu, Gius et al. 2008; Yap, Li et al. 2010)
CCND1	NA	Induced by DNA damage	(Wang, Arai et al. 2008)
BC200	Multiple cancers	Protein binding	(Chen, Bocker et al. 1997; Iacoangeli, Lin et al. 2004)
PTENP1	Prostate	Decoy of microRNA binding sites	(Poliseno, Salmena et al. 2010)
PCA3	Prostate	NA	(Bussemakers, van Bokhoven et al. 1999)
PCATs	Prostate	NA	(Prensner, Iyer et al. 2011)

Different expression of lncRNAs between cancer and normal tissue indicates the potential use of lncRNA as a diagnostic and prognostic tool, and expanding understanding of its molecular mechanisms also suggest avenues to treat cancer with lncRNA molecules. PCA3, for an instance, has been tested in large controlled clinical settings, confirming the previous discovery of PCA3 differential expression pattern, about 60 to 100 times higher in cancer tissue than benign prostate (Hessels, Klein Gunnewiek et al. 2003) . The effectiveness of PCA3 in diagnosing prostate cancer

is about the same as that of prostate-specific antigen (Day, Jost et al. 2011). HOTAIR, as another example, was found to be increased by hundreds of times in metastatic breast cancer tissues and also shows a robust association with patient prognosis (Gupta, Shah et al. 2010). In addition to breast cancer, HOTAIR was subsequently found to be associated with multiple cancers, including colon cancer and hepatocellular carcinoma (Kogo, Shimamura et al. 2011; Yang, Zhou et al. 2011). Scientists also identified lncRNA MALAT1 as a prognostic factor for patients with non-small cell lung cancer and then validated it in endometrial stromal sarcoma of uterus and liver cancer afterwards(Ji, Diederichs et al. 2003; Yamada, Kano et al. 2006; Lin, Maeda et al. 2007). Ideally biomarkers should be sampled easily, for example, from body fluids. Previous investigations have suggested some microRNAs are stable and detectable in blood, urine and sputum, while very few lncRNAs examples have been elucidated in body fluids. HULC (highly upregulated in liver cancer) is detectable in the blood of liver cancer and colon cancer patients by conventional PCR and PCA3 is also a good biomarker in urine for prostate cancer as mentioned above (Panzitt, Tschernatsch et al. 2007; Matouk, Abbasi et al. 2009). In addition to cancer, lncRNAs aberrations were also found in other diseases including Alzheimer's diseases, psoriasis and heart diseases (Sonkoly, Bata-Csorgo et al. 2005; Faghihi, Modarresi et al. 2008; Korostowski, Sedlak et al. 2012). LncRNA Beta-secretase 1 antisense (BACEAS), which regulates Beta-secretase 1sense, exhibit enhanced expression in several regions of brains in several patients with Alzheimer's diseases (Faghihi, Modarresi et al. 2008). Moreover, lncRNA psoriasis susceptibility-related RNA Gene Induced by Stress (PRINS) was named by its potential relationship with psoriasis by recent study (Sonkoly, Bata-Csorgo et al. 2005).

LncRNAs might be also useful for therapeutics. Since lncRNA can recruit chromatin-modifying complexes to silence gene expression, it might be possible to deliver specific lncRNAs by gene therapy delivery systems though it could be risky. Moreover, lncRNA could also be recognized

by synthetic siRNAs or microRNA. For example, H19 was successfully targeted by a plasmid-based RNAi system to advance treatment of bladder cancer and pancreatic cancer (Smaldone and Davies 2010; Sorin, Ohana et al. 2012). Knockdown of prostate cancer noncoding RNA 1 (PRNCR1), which is upregulated in some prostate cancer, weakens the viability of prostate cancer cells and androgen receptor transactivation activity (Chung, Nakagawa et al. 2011). Not only being potential biomarkers or therapeutic agents, lncRNA could also be used to develop novel therapeutic strategies. Synthetic lncRNAs might work as decoy of transcription factor or microRNA, thus reduce transcription activity or microRNA expression, as Gas5 and PTENP1 do (Mourtada-Maarabouni, Pickard et al. 2009; Poliseno, Salmena et al. 2010). To date lncRNA has mostly been looked at by qPCR based and by RNA sequencing tests. However, a novel *in situ* hybridization assay allows us to investigate lncRNA expression by microscopy.

Principle of RNAscope® chromogenic *in situ* hybridization assay

RNAscope® chromogenic *in situ* hybridization assay is novel RNA detection system use with intact cells on either fresh or preserved tissues. Two independent probes (double Z probes) are designed to hybridize to target sequence in tandem for signal amplification, which ensure specificity because it has extremely low possibility that two probes will bind to non-specific target simultaneously. The lower part of Z probe is 18 to 25 base region that is complementary to target sequence and the higher part is a 14 base tail sequence that is designed to bind pre-amplifier. Two Z probes are connected by a spacer sequence and double Z probes are designed specifically for each RNA target. Usually, 20 double Z probe pairs can cover 1 KB length of target sequence. After double Z probes hybridize to target RNA sequence, a cascade of steps is employed to amplify signals. Preamplifier first binds to higher region of double Z probes and then amplifier bind to the 20 binding sites on each pre-amplifier. Label probes containing chromogenic enzymes (horseradish peroxidase) are then applied to conjugate with the 20 binding site on each amplifier. Finally, colored RNA molecules could be visualized by bright field microscope after

substrate application. The general scheme is displayed in Figure 1. Comparing with other RNA in situ hybridization systems, RNAscope® technology holds several advantages: 1) it is a highly sensitive platform because each RNA sequence only requires three double Z pairs to bind, so, 20 double Z probe pairs should target RNA molecule robustly even it is somehow degraded. 2) it is highly specific because only simultaneous binding of two independent Z probes provide binding site for pre-amplifier, while singe Z will not produce any binding site that prevent non-specific amplification.

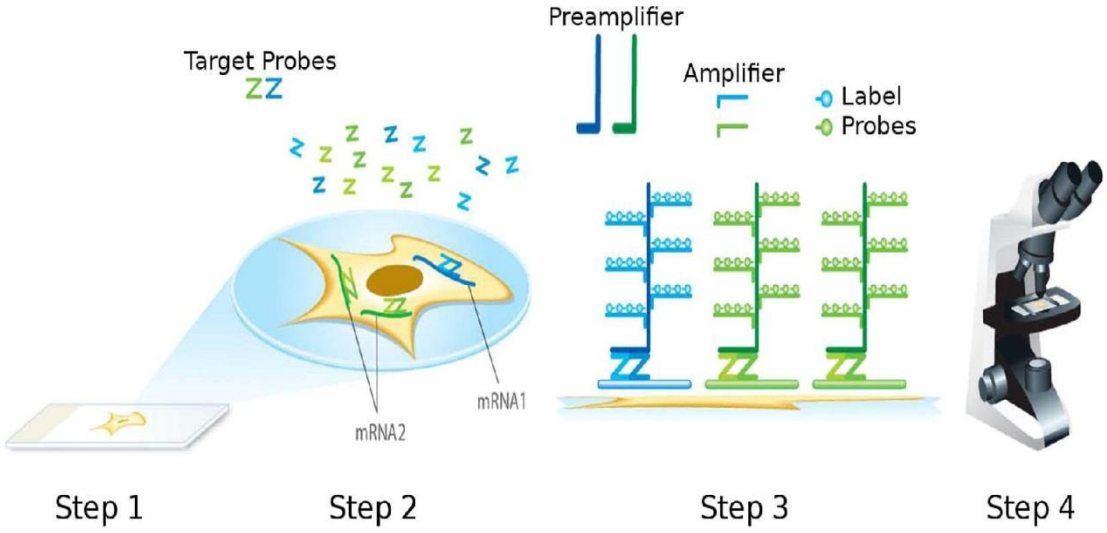


Figure 1. Work scheme of RNAscope® *in situ* hybridization assay

Conclusion

As a major cause of morbidity and mortality in women, breast cancer is one of the most critical diseases internationally. Tremendous effort has been made in basic research and clinical trials to investigate the molecular science of breast cancer and improve its diagnosis, prognosis and treatment. However, there are some barriers remaining for this complex disease. With advanced technologies, we need to pay more attention to discover useful biomarkers which can accurately

classify subtypes of breast cancer, especially carcinoma *in situ*. Though DNA ploidy change has been studied in FFPE samples previously, subtle intratumoral heterogeneity have not been studied, which might contribute to the identification of various cancer signatures in combination with multiparametric flow cytometry. In addition, very few investigators have used CISH to detect and quantify lncRNA expression in breast cancer archived samples and very few prior studies have looked at the connection between lncRNAs expression levels and breast histopathological markers, which may be valuable for the purpose of breast cancer diagnosis and prognosis.

Alternative Tissue Sampling Method to Detect Genetic Instability of Breast Carcinoma with Flow Cytometry

Aim:
To test a novel approach for investigating DNA ploidy and intratumoral heterogeneity within solid tumors by flow cytometric analysis

Hypothesis:
Core punch tissue sampling allows the targeted recovery of tumor tissues from FFPE specimens for DNA aneuploidy and intratumoral heterogeneity analyses of breast carcinomas by flow cytometry.

Material and Methods:

Tissue acquisition
FFPE tissue blocks of breast carcinoma were provided courtesy of Dr. Kavita Munjal (Department of Pathology, Sri Aurobindo Institute of Medical Sciences, Indore, Madhya Pradesh, India) and normal placenta specimens were retrieved from Fletcher Allen Health Care (FAHC), Burlington, VT. Each breast carcinoma FFPE block was cut into 5µm section on microtome (Leica, Allendale, NJ) subsequently stained with hematoxylin and eosin (H&E) in the FAHC histology department. All H&E stained slides were reviewed and tumor regions were marked by Dr. Donald L. Weaver from Department of Pathology, University of Vermont.

Tissue Sampling
Breast cancer FFPE blocks were placed in tissue arrayer (Beecher Instrument, Silver Spring, MD) and 1mm diameter tumor tissue cores were punched from the blocks guided by the marked H&E slides. 1mm normal placental tissue cores were punched from FFPE placental tissue blocks. Cores were put on a piece of 50 micron nylon mesh provided by Advanced Genome Technology Core (AGTC) in the University of Vermont and cut by knife to remove excess paraffin wax. Then, tweezers were used to wrap the tissue core in nylon mesh that was then put it into

micromesh biopsy cassettes (Thermo electron cooperation, Pittsburgh, PA). The mesh was clamped down securely by the cassette lid to prevent the tissue cores from coming out of the nylon mesh. Post-punch FFPE block tissue sections were recut for H&E staining in order to confirm that core was correctly punched from tumor enriched regions.

Single cell suspension preparation

Several approaches were applied to deparaffinize tissue core and disaggregate cells in order to make single cell suspensions. The overall schemes are similar including xylene deparaffinization, ethanol rehydration and enzymatic treatment, but distinct treating durations and enzymes were tested to generate the most effective approach. According to the optimal procedures, extracted tissue core were placed into biopsy cassettes and then immersed into two changes of xylene solution (Fisher scientific, Pittsburgh, PA) for two hours, respectively. Subsequently, two changes of 100%, 95%, 70%, 50% ethanol (Pharmco-aaper, Brookfield, CT) were sequentially applied to rehydrate tissue core in cassettes for twenty minutes each. After rehydration, the cassettes were left in MilliQ DI water for one and a half hours to get rid of remaining ethanol. After that, cassettes were opened carefully and tweezer was used to clamp mesh wrapped tissue core and place it into a clean 1.5ml centrifuge tube. Subtilisin Carlsberg solution was prepared by adding 0.1%(w/v) Sigma Protease XXIV (Sigma, St.Louis, Mo) , 0.1M Tris, 0.07M NaCl to make a pH 7.2 solution, or by using a Protease XXIV kit (Biocare medical, Concord, CA); 500µl of the solution was added into tubes for disaggregating cells of the tissue core at 37°C water bath overnight. Next morning, Orbit LS shaker (Labnet International, Edison, NJ) was applied to shake tubes for 20 mins after removing that tube from 37°C water bath. Then, Subtilisin Carlsberg solution now containing the digested cells was filtered through fresh 50 micron nylon mesh into a clean dark centrifuge tube for staining.

Staining

For staining purpose, propidium iodide (PI), Sytox Green, SYBR Green and 4', 6-Diamidino-2-Phenylindole, Dihydrochloride (DAPI) were used for comparison. PI solution (Sigma, St.Louis, Mo) was diluted from 1.0mg/ml into 1.0µg/ml by adding 10mM Tris, 10um Nacl and 1ul Nonidet P-40 water solution. And the prepared PI solution was mixed with Substilin Carlsberg solution in equivalent amount in 1.5ml dark centrifuge tube to bind both double strand and single strand nuclei acid. Appropriate Ribonuclease A (10µg RNase A/ 1ml PI solution -100µg RNase A /1ml PI solution) was added up in order to get rid of single strand RNA. The tubes were then placed in the dark fridge to incubate at 4 centigrade for 1.5 hours. Both SYBR Green (Life Technology, Foster City, CA) and Sytox Green (provided by Dr. Yvonne Janssen-Heininger's lab) were required to add in 1X as the final concentration and the incubation situation was set at 4 centigrade for 30min in dark. DAPI (Life Technology, Foster City, CA) was firstly made 5mg/ml stock solution in DMSO, which was diluted to 3µM in DAPI staining buffer (100 mM Tris, pH 7.4, 150 mM NaCl, 1 mM CaCl2, 0.5 mM MgCl2, 0.1% Nonidet P-40) and the specific incubation situation for DAPI was 4 centigrade and 15 mins in the dark. After staining, all solution was transferred from 1.5ml centrifuge tube to flow cytometry tube (either provided by AGTC or the Department of Immunology at the University of Vermont).

Flow Cytometry

Two flow cytometers were utilized. Coulter EpicsXL-MCL (Beckman Coulter, Inc, Brea, CA) is a flow cytometer located in the AGTC and was applied to cells suspension stained with PI, Sytox Green and SYBR Green. Procedures recommended by operator manual were followed. The other flow cytometer that was used is BD LSRII (Becton Dickinson, Franklin lakes, NJ) that locates in Department of Immunology in University of Vermont. Sample tubes stained with Sytox Green, SYBR Green and DAPI were detected by this cytometer. Procedures recommended by operator manual were followed. At least 5000 intact nuclei were collected for plotting a DNA histogram.

The diploid placental cells obtained from placental FFPE block were used as an external standard which was analyzed firstly to adjust the voltage of the photomultiplier in order to fix the signal of the diploid standard at channel position 50 and it was also measured after all the tumor samples.

Data analysis

FCS EXPRESS 4 FLOW CYTOMETRY – RUO [De Novo Software, Los Angeles, CA] was used to analyze all EpicXL- MCL LMD file and BD LSRII FCS file. The DNA index (DI) was defined as the ratio of the mean fluorescence channel number of the G0/G1 in tumor sample to the mean fluorescence channel number of the G0/G1 in external control and the coefficient of variation (CV) of the G0/G1 and G2/M peaks was defined as the normalized standard deviation obtained by Gaussian curve fitting, both of which were analyzed by FCS EXPRESS 4 FLOW CYTOMETRY – RUO. Besides, other parameters like G1 phase portion, G2 phase portion, S phase portion and percentage of background aggregates debris (B.A.D) were also measured. The best fitting curve was selected from six analysis models that already exist in software. For this study, a DI between 0.95-1.05 was defined as DNA diploid, and any DI < 0.95 or > 1.05, as aneuploidy (Corver, Ter Haar et al. 2011). To assess intratumoral DNA content heterogeneity, multiple cores were punched from individual FFPE tissue blocks identified by H&E review as containing sufficient tumor enriched regions. Within-assay coefficient of variation, which was 2% for DNA index analysis and 7% for %SPF analysis, was calculated with 10 cores from a placental FFPE block to confirm the limits for valid intratumoral heterogeneity. The assessment of intratumoral heterogeneity was defined by either a difference of the DNA-ploidy pattern or a variation of >8% (±2 x CV) in the DNA index among aneuploid patterns between tumor subpopulations in separate regions, or S-phase fraction difference >28% (±2 x CV) regarding lowest DNA index among all samples as the reference. DNA content intratumoral heterogeneity statistical analyses (Fisher's exact test) were performed using GraphPad InStat (GraphPad Software, Inc., La Jolla, CA).

Results:

The optimization of single cell suspension

To determine which enzymes is better to disaggregate single cells from tissue core, 0.5% trypsin solution and 0.1% Subtilisin Carlsberg solution were applied for 1mm tissue cores that was punched from archived breast carcinoma FFPE tissue blocks in Experimental Pathology lab, and compared, following same deparaffinization and rehydration procedures. PI was used as DNA binding dye in both tests. Figure 2 shows the better applicability of Subtilisin Carlsberg for cell releasing.

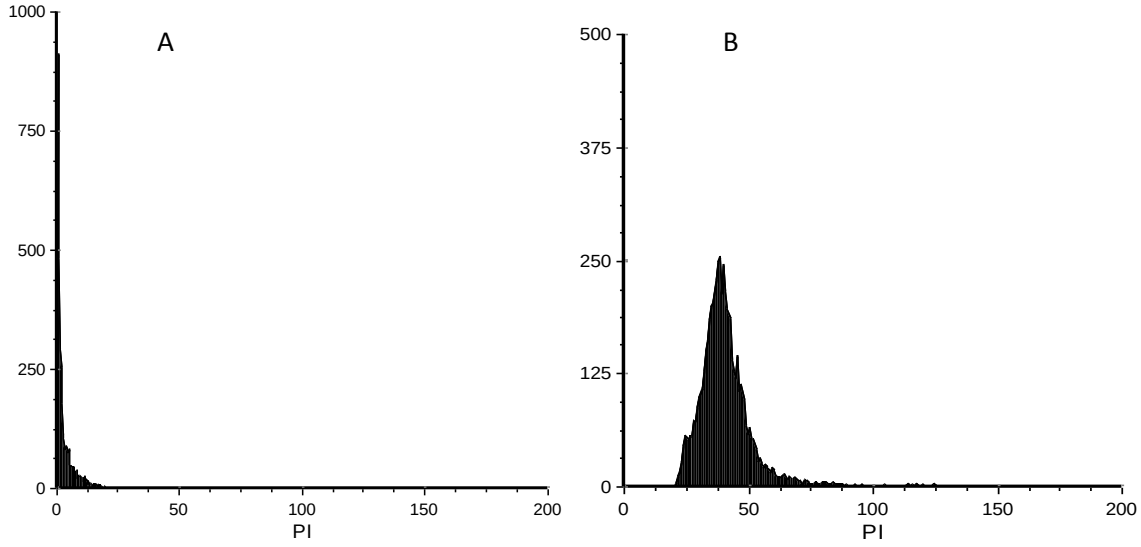


Figure 2. Comparison of different digestion enzymes

Tissue core were treated with 0.5% trypsin (A) and 0.1% Subtilisin Carlsberg (B) and subsequent flow cytometric analysis were compared. PI was used as binding dye for all tests and RNase A was applied to eliminate RNA binding.

Then, we assessed different combinations of deparaffinization and enzymatic digestion to optimize the time for disaggregating cell. 1mm placental tissue core were used as our sample and PI was used for DNA binding dye. Figure 3 shows de-wax and digestion time data. In brief, a 4h xylene dewaxation step plus overnight digestion rinse resulted in a histogram with a distinct and

narrower peak with a lower CV than a 2h xylene dewaxation plus shorter time of incubation (data not shown).

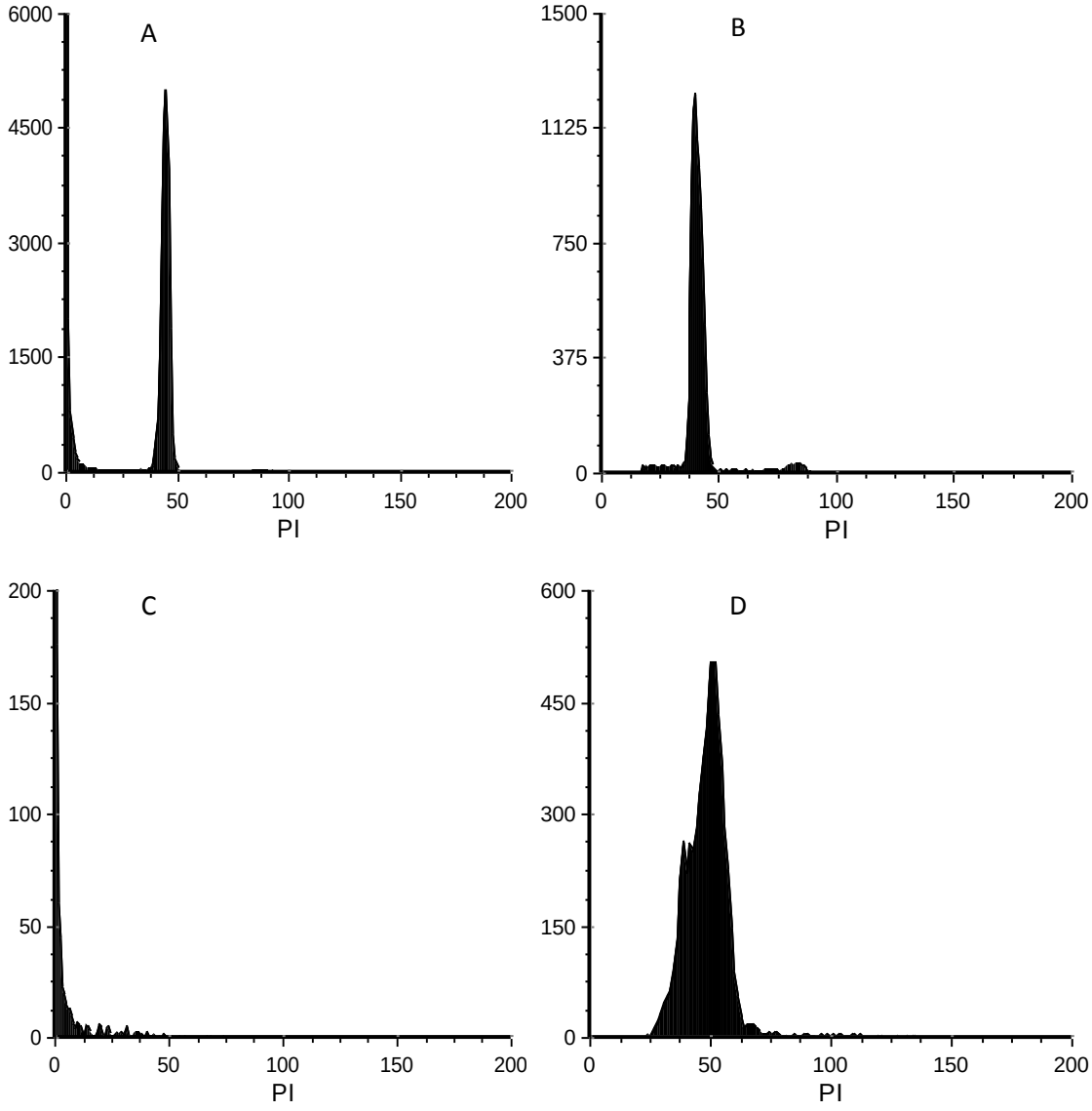


Figure 3. Comparisons between deparaffinization time and enzymatic digestion time

Four combinations of xylene diparaffination and subtilisin carlsberg digestion, differing in length, were tried and compared: 4 hours dewaxing plus overnight enzymatic digestion (A); 2 hours dewaxing plus overnight enzymatic digestion (B); 2 hours dewaxing plus 2 hours enzymatic digestion (C); 4 hours dewaxing and 2 hours enzymatic digestion (D). Propidium iodide (PI) was used as binding dye for all tests and RNase A was applied to eliminate RNA binding.

DNA Binding affinitiy is another factor that influences flow cytometry. Four binding dyes: PI, SYTOX Green, SYBR Green and 4',6-diamidino-2-phenylindole (DAPI) were compared and two flow cytometers were used. Coulter EpicsXL-MCL was applied to only assess affinity of DNA binding with only PI, SYTOX Green and SYBR Green as it lacks apporporite detector for DAPI , as Figure 4 upper panel shows, suggesting a better applicablity of SYTOX Green and SYBR Green, while BD LSRII flow cytometry was subsequently used to compare SYTOX Green, SYBR Green and DAPI. Figure 4 low panel disaplys DAPI is the best among three.

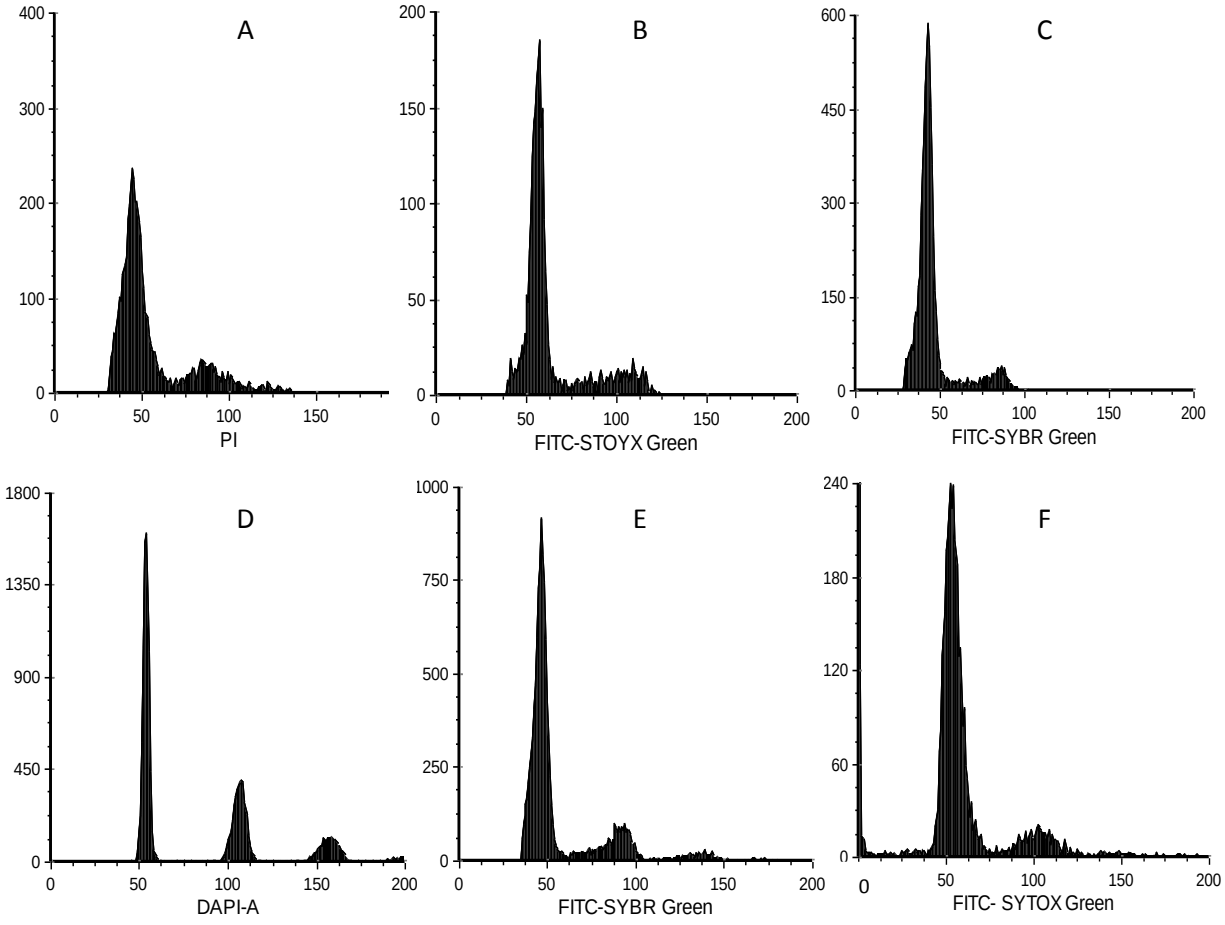


Figure 4. Comparisons of different DNA binding dyes

PI (A), SYTOX Green (B), SYBR Green (C) DNA binding affinity was tested with Beckman Coulter EpicsXL-MCL; DAPI (D), SYBR Green (E), SYTOX Green (F) DNA binding affinity was assessed by BD LSRII. RNase I was added with PI to eliminate RNA binding

To verify our temporary conclusion that DAPI was the most appropriate dye, we set up another test to compare the results that SYBR Green produced with two flow cytometers to exclude any bias on machines. Figure 5 demonstrates that BD LSRII generated result is much better than that of Beckman Coulter EpicsXL-MCL with the same DNA binding dye, suggesting DAPI would be the best option for our subsequent analysis.

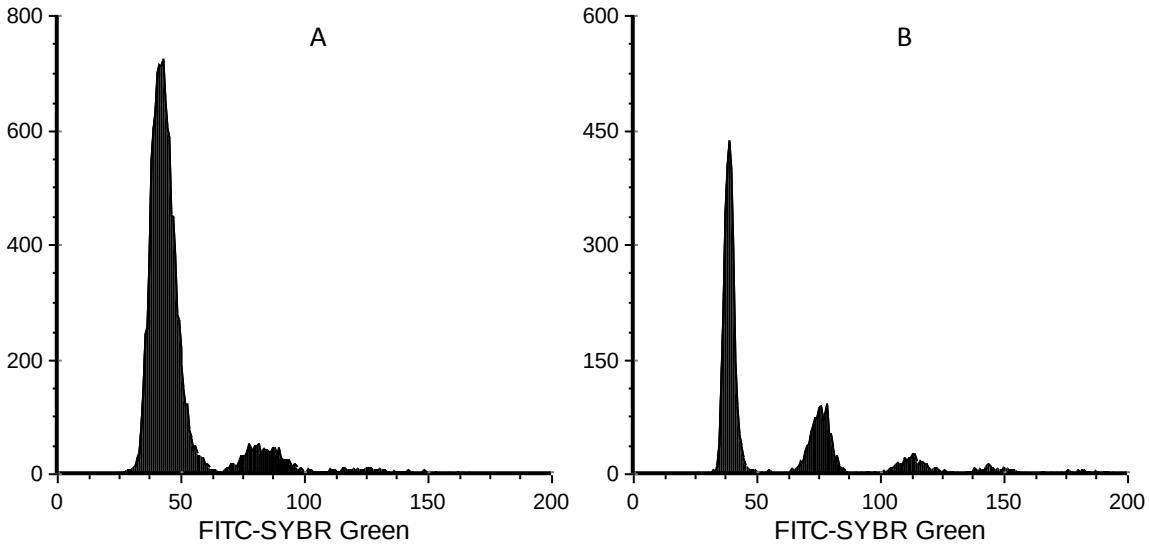


Figure 5. Comparisons of two flow cytometers

SYBR Green was used as DNA binding dye to compare flow cytometry results generated by each flow cytometers: Beckman Coulter EpicsXL-MCL (A) and BD LSRII (B)

The application of novel core punch sampling method to show DNA ploidy

Of the 31 breast carcinoma specimens (1 DCIS, 1 medullary, 1 invasive ductal carcinoma/invasive lobular carcinoma, 2 pure invasive lobular carcinoma, 26 pure invasive ductal carcinoma), 23 (74%) were shown as aneuploidy by our novel sampling method, while 8 (26%) were diploidy. Pre-punched and post punched H&E stain slides was made to confirm proper tissue selection, as Figure 6 showed. Figure 7 exhibits various aneuploidic examples.

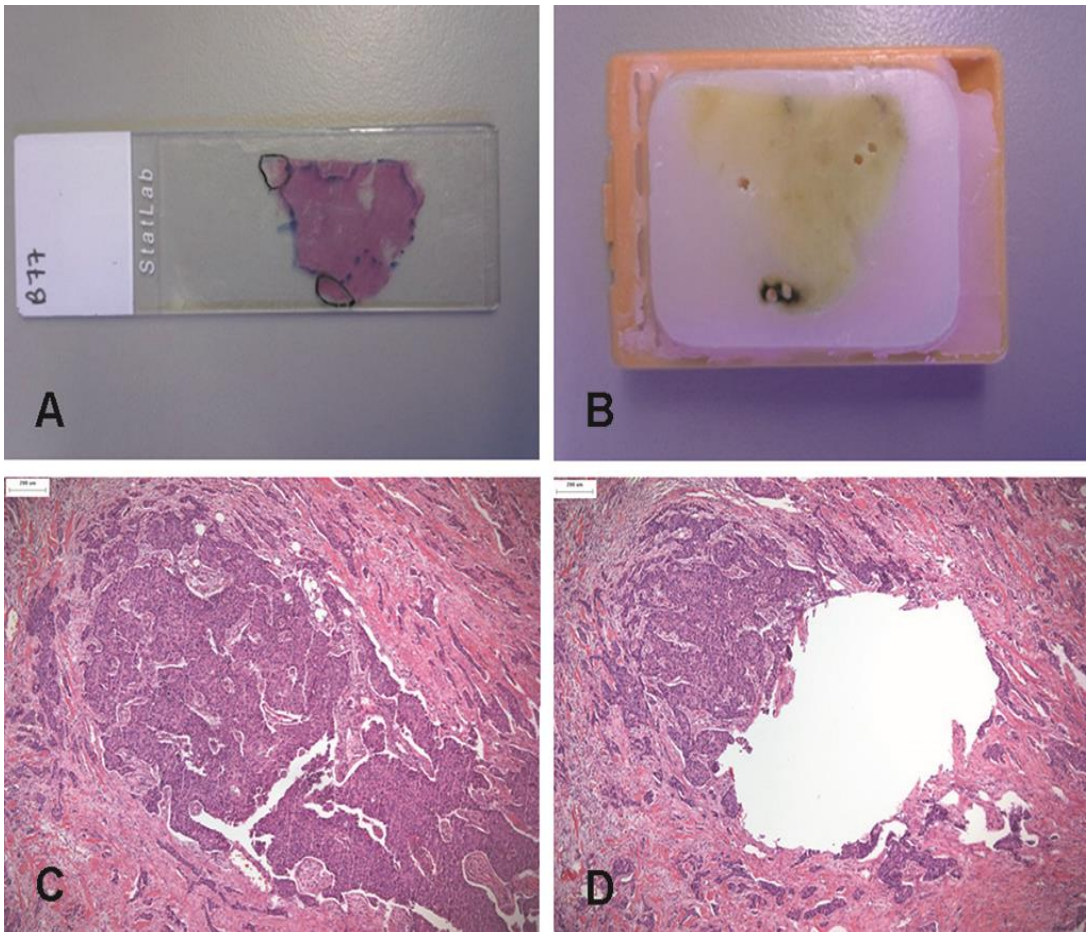


Figure 6. Identification and sampling of tumor rich regions in FFPE tissue blocks

(A) Hematoxylin and eosin (H&E) stained slide marked for tumor (blue line) and normal cell regions (black line). (B) Post-punch FFPE block. (C) Pre-punch H&E stained tumor (D) Post–punch H&E to confirm accurate sampling

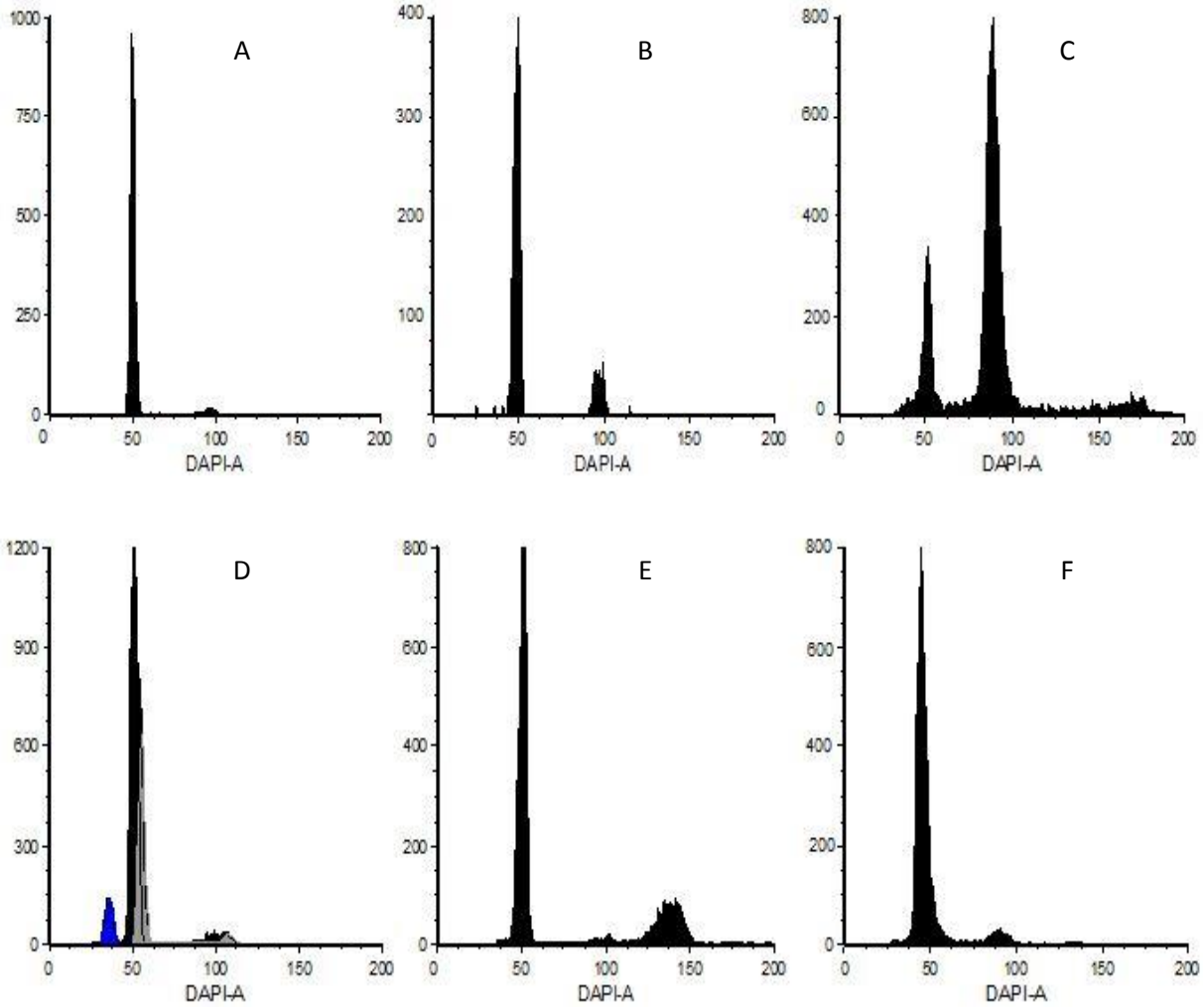


Figure 7. DNA content profiles from selected individual core punch samples

(A) FFPE normal placental cells with DI=1.00. (B) Tumor adjacent normal cells DI=1.00, showing the correspondence of normal DNA content profiles from breast and placenta tissues. (C) DNA content histogram showing a hyperdiploid tumor with DI=1.72. (D) Multiploid tumor showing hyperdiploid (DI=1.08 [grey]) and hypodiploid (DI=0.72 [blue]) populations. (E) Hypertetraploid tumor with DI=2.67. (F) Diploid tumor DI=1.00.

The application of our novel core punch sampling method: to detect subtle intratumoral heterogeneity

Among all 31 specimens, 23 tumors had sufficient tumor for 2 or more core-punches, 3 (13.0% [all IDC: 1 moderate, 1 poorly differentiated]) showed diploidy only; 20 (87.0% [3 moderately, 17 poorly differentiated]) were aneuplodic: 10 (43.5%) showed no significant variation in DI and 10 (43.5%) showed DI intratumoral heterogeneity; 11 (47.8%) showed %SPF intratumoral heterogeneity, 7 (30.4%) showed both DI and %SPF heterogeneity. Figure 8 displays 2 cases with intratumoral DNA index heterogeneity and 1 case with intratumor S-phase fraction heterogeneity.

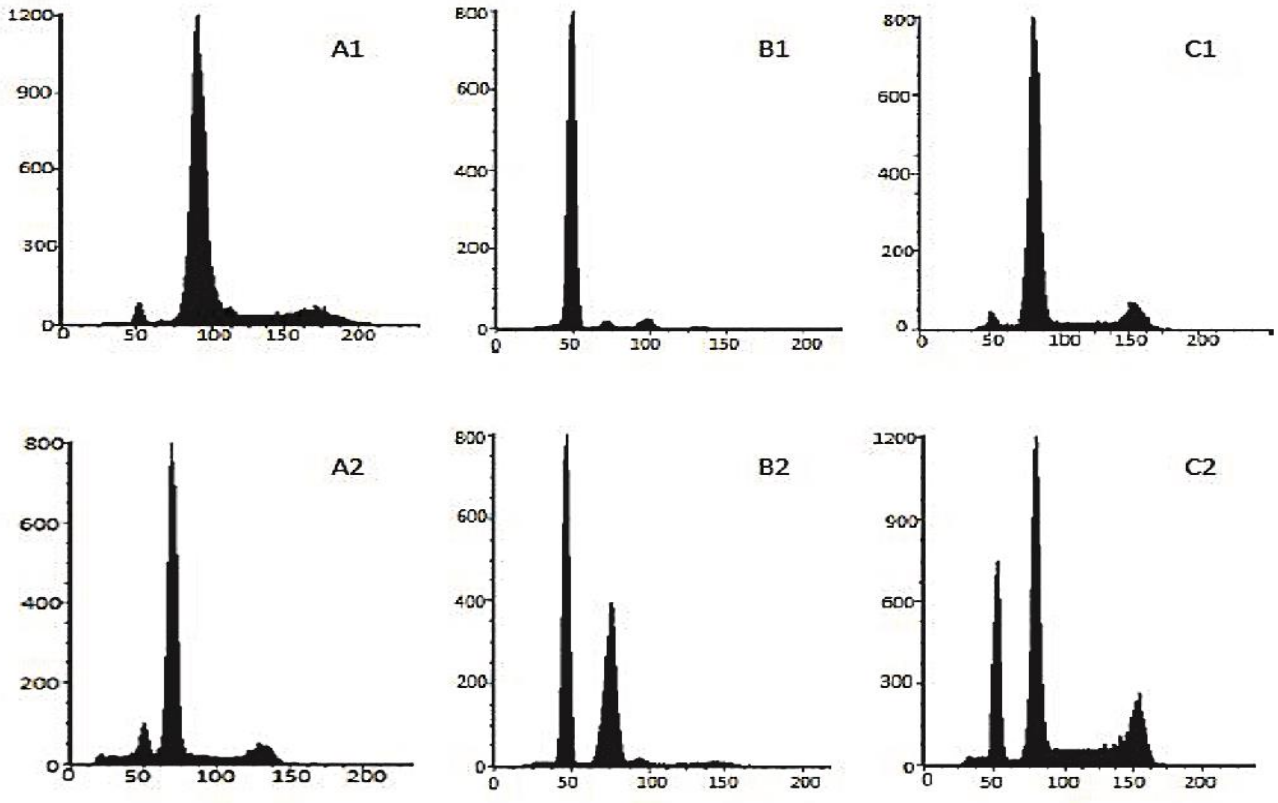


Figure 8. Intratumoral DNA heterogeneity demonstrated by core punch tissue sampling

Histogram group A and B generated by cores in the same tissue block showing DNA index heterogeneity with (A1) DI=1.71 vs. (A2) DI=1.39; (B1) DI=1.47 vs. (B2) DI=1.62. Intratumor percentage S-phase fraction heterogeneity (C1) %SPF=12.61 vs. (C2) %SPF=27.01.

Statistical analysis

Patient ages ranged from 35 to 95 years (mean 52.1); tumor size from 10 to 80 mm (mean 37.0); tumor stage 1 (n=7), 2 (n=15), 3 (n=5), 4 (n=4); ER positive 17 (54.8%), ER negative 14 (45.2%); PR positive (64.5%), PR negative 11(35.5%); HER2 positive 10 (32.3%), HER2 negative 21 (67.7%). There were no significant differences ($P>0.05$) of DI and/or %SPF intratumoral heterogeneity with reference to any of these parameters.

Discussion:

The main findings of this study are two-fold: firstly, that core-punching is an effective method for sampling cells from FFPE specimens for flow cytometric DNA content analysis, and secondly that DI and %SPF intratumoral heterogeneity are relatively common events in breast tumors even within focal tumor localized within a single surgical block.

Previous FFPE specimen studies have assessed DNA content from intact nuclei recovered from ≥50 μm whole tissue sections prepared by microtomy. Thick sections are required as DI index and the proportion of cells scored as aneuploidic increases significantly comparing sections cut to a thickness of 5, 10, 20, 40, or 80 mm and the thinner the section the greater the cellular debris generated (Kallioniemi 1988). Punching recovers 1 mm diameter cores through the depth (up to several millimeters) of a tissue block and as such is a highly effective approach for recovering intact nuclei. Besides, with the margined tumor rich regions on tissue blocks, punching targets much more specifically on tumor cells instead of including abundant normal adjacent cells around by cutting whole sections that possibly lower or remove comparative tumor cell population proportion in our scale (Figure 9). The potential disadvantages are the assumption that tumor identified at the block surface is present through the depth and the lack of 'internal control' normal diploid cells in a tumor rich region; whole sections are more likely to contain sufficient control reference cells; however, the very presence of these cells also impair DNA

content/aneuploidy assessment (Hedley, Friedlander et al. 1983). The tissue blocks available for this study lacked high cellularity normal adjacent tissues. FFPE normal placental tissue was used instead and its status as a viable diploid control confirmed by comparing multiple (10) cores to confirm consistent DNA content measurement. The core approach was then used to investigate intratumoral DI and %SPF values within a surgical block: heterogeneity was found in 43.5% and 47.8% of breast tumors respectively.

Several previous studies have investigated DNA content heterogeneity comparing different samples of gross dissected fresh tumors or serial FFPE sections or sections from alternative blocks (Table 2). Assay of multiple samples is recommended to detect aneuploidy; for example, Bergers and his colleagues (Bergers, van Diest et al. 1996) reported that at least six separate areas required measurement in order to fully detect heterogeneity. In the present study, tumor within an FFPE block was sampled at 2-3 tumor rich sites. The heterogeneity detected is within the range found in previous studies (Table.15). Practically, this shows that selective core punching of tumor regions requires less extensive tissue usage to reveal heterogeneity than a whole section approach. Biologically, these data show that even DNA content heterogeneity can be a highly localized event within a tumor.

With the advent of improved dyes and instrumentation, there has been a resurgence of interest in the use of flow cytometry for the analysis of FFPE specimens. Multiparametric techniques enable combined DNA content and protein biomarker assay through the combined use of labeled antibodies (Corver and ter Haar 2011; Dayal, Sales et al. 2013). Multiparametric approaches in conjunction with core punching will likely provide more refined data than is possible from a whole section approach. Studies investigating the relationship of DNA content heterogeneity to other markers of genomic instability such as mutations, deletions, insertions, and translocations are warranted.

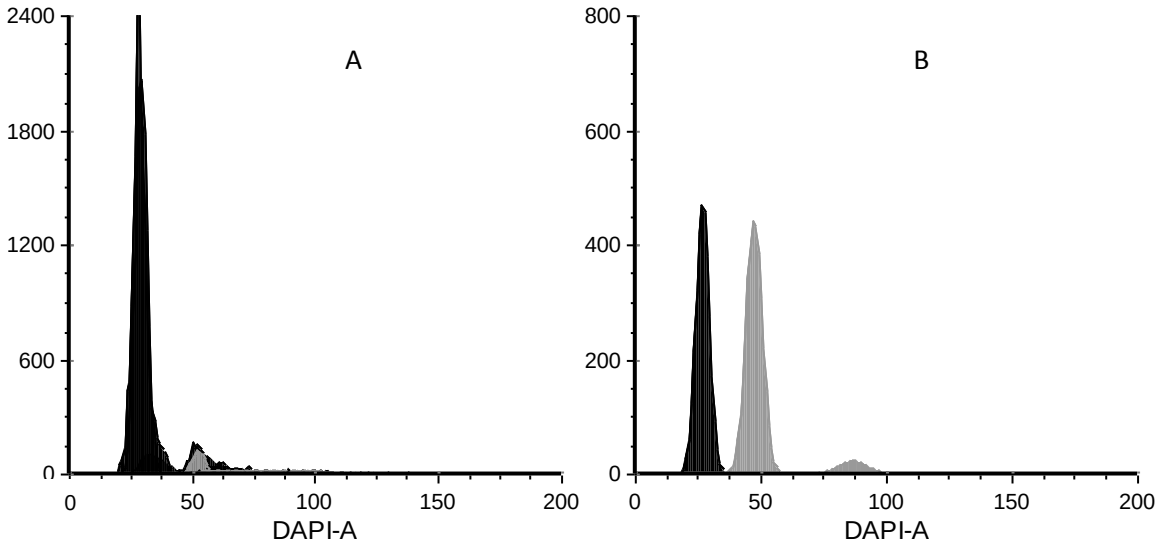


Figure 9. Comparison of core punching and whole section tissue sampling method

One FFPE tissue block was sampled by cutting whole section (A) and alternative punching method (B), stained by DAPI and compared by BD LSRII flow cytometer. Normal adjacent cells were drawn in black and tumor cells were depicted in grey.

Table 2. DNA content heterogeneity reported in previous studies

Author	Tumors n	Samples n	DNA index heterogeneity rate (%)	SPF heterogeneity rate (%)
(Prey, Meyer et al. 1985)	8	5–11	67	-
(Kallioniemi 1988)	104	3–10	13	36
(Meyer and Wittliff 1991)	61	2–31	26	-
(Bonsing, Beerman et al. 1993)	18	1–11	67	-
(Schvimer, Lash et al. 1995)	28	3	43	-
(Danesi, Spano et al. 1997)	102	Not given	28	-
(Arnerlov, Emdin et al. 2001)	48	4–5	44	71
The present study	23	2–3	44	47

Conclusion:

The first study demonstrates that a core-punching method is effective to release cells from FFPE tissue specimens for flow cytometric DNA content analysis, which is a standard techniques to detect abnormal cell in terms of either chromosome abnormalities or disordered cell cycle. The composite analysis of the results from this investigation well illustrates that this alternative tissue sampling method is able to release intact single cell for flow cytometry from one 1mm diameter tissue core punched in depth from preserved tissue blocks. We were able to perform flow cytometric analysis with this new method to reveal both normal and abnormal DNA ploidy. Comparing with typical whole tissue sections cell releasing strategy, this new technique can provide more accurate tumor cell ploidy status without taking much adjacent normal cells into account. More importantly, by using current method, we identified intratumoral heterogeneity

either in DNA content or S phase fraction, from different cores that were retrieved from single surgical blocks, indicating even within a single piece of solid tumor, tumor cell populations may come from distinct clones and undergo complicated mutations along tumor develops. Our findings can provide some insights not only for breast cancer biomarker discovery that looks for precise DNA abnormality pattern, but also for breast cancer therapeutics regarding intratumoral heterogeneity. Although internal control is lacking, external control normal placental cells seems to be a good surrogate.

Long non-coding RNA Chromogenic *in situ* Hybridization Signal Patterns Correlate with Breast Tumor Pathology

Aim

To apply a novel RNA *in situ* hybridization detection technology for the investigation of lncRNA expression in FFPE specimens and to assess the relationship between six lncRNAs and common breast tumor markers using tissue microarrays.

Hypothesis

RNAscope® CISH can be used to substantiate lncRNAs identified as potential markers of breast cancer by qPCR or RNA sequencing studies.

Material and Methods

Tissue microarray (TMA) preparation

All FFPE tissue blocks of breast carcinoma and DCIS were retrieved from FAHC archives. FFPE blocks were recovered for 52 patients identified by electronic record search by pathology residents Drs. Daniel Olsen and James deKay as having concurrent DCIS and invasive breast cancer (IC). Each FFPE block was sectioned, stained by H&E and subsequently reviewed by Dr. Donald L. Weaver. Regions of DCIS, IC and normal adjacent epithelia (NA) were marked on the slides and used to guide core punching (one core punch per NA, DCIS or IC region per patient) for the construction of the TMAs from the companion FFPE surgical blocks. A tissue arrayer (Beecher instrument, Silver Spring, MD) was used to prepare tissue microarray FFPE blocks. A one millimeter diameter receptor needle was used to extract cores from a FFPE block and relocate it into a recipient paraffin block. Tissue cores from same individuals were made in the same horizontal line, separated by tissue types. Five tissue microarray FFPE blocks were made by this technique and each block contained about ten invasive cancer spots, ten DCIS spots and ten NA tissue spots with one separated head and neck tumor tissue spot as location control (line 1, 3, 5, 7, 9), as shown in Figure 10. Five TMA blocks were placed in Thelco® laboratory oven (Jouan, Inc.,

Winchester, VA) to incubate overnight at 60℃. One section of five micrometer tissue section of each TMA block was cut, stained and reviewed by pathologists in FAHC.

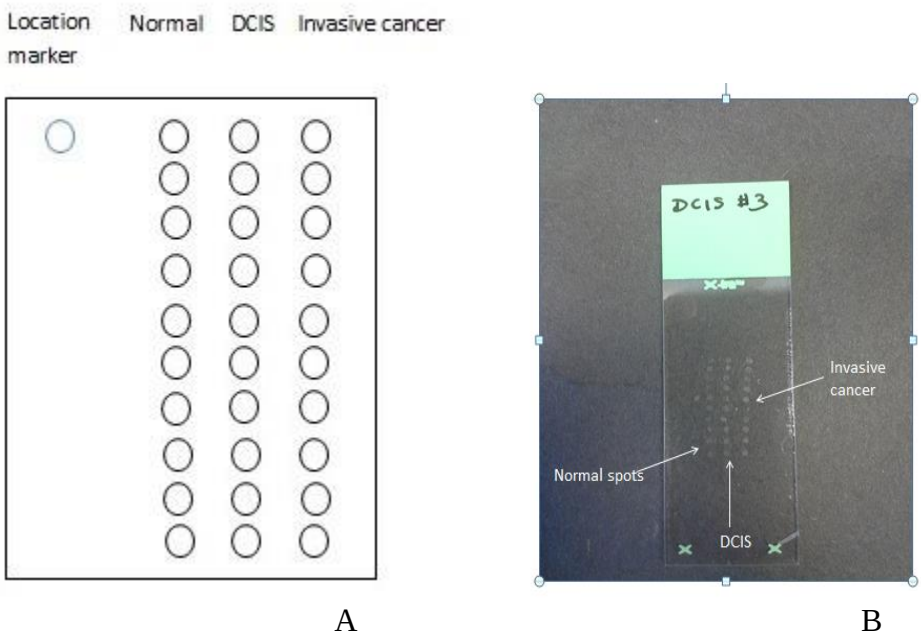


Figure 10. Tissue microarray design

A) diagram of tissue microarray block; B) actual TMA

LncRNA in situ hybridization process

Five micrometer tissue slides were made and baked in Thelco® laboratory oven (Jouan, Inc., Winchester, VA) overnight at 60℃. Generally, RNAscope® FFPE *in situ* hybridization assay platform (Advanced Cell Diagnostic, Hayward, CA) was applied in our study with minor modification on pre-treatment steps. The optimal procedures were listed in a flow diagram in figure 11. Positive control POLR2A (P/N 310451, Advanced Cell Diagnostic, Hayward, CA) and negative control DapB (P/N 310043, Advanced Cell Diagnostic, Hayward, CA) were tested for each lncRNA probe. To ensure the RNA targeting specificity of ACD lncRNA probes, we applied an additional RNase A (Sigma-Aldrich, St. Louis, MO) treatment step and DNase I (Sigma-

Aldrich, St. Louis, MO) treatment step right after pretreat 3 and compare the results to that of regular procedures. RNase A was diluted to 100µg/ml in phosphate buffered saline (PBS) (Sigma-Aldrich, St. Louis, MO) and then 100µl RNase A solution was applied to cover and incubate tissue section for thirty minutes at room temperature, followed by Milliq water washing for five minutes for three times. Similarly, 50µg/ml DNase I work solution was mixed by DNase I, 10x DNAse I reaction buffer (P/N y02340, Invitrogen, Carlsbad, CA) and Buffer AE (Qiagen, Valencia, CA) and 100µl DNase I work solution was added to cover and incubate tissue section for fifteen minutes at room temperature. Three times of Milliq water washing for five minutes, respectively were followed to get rid of excessive solution.

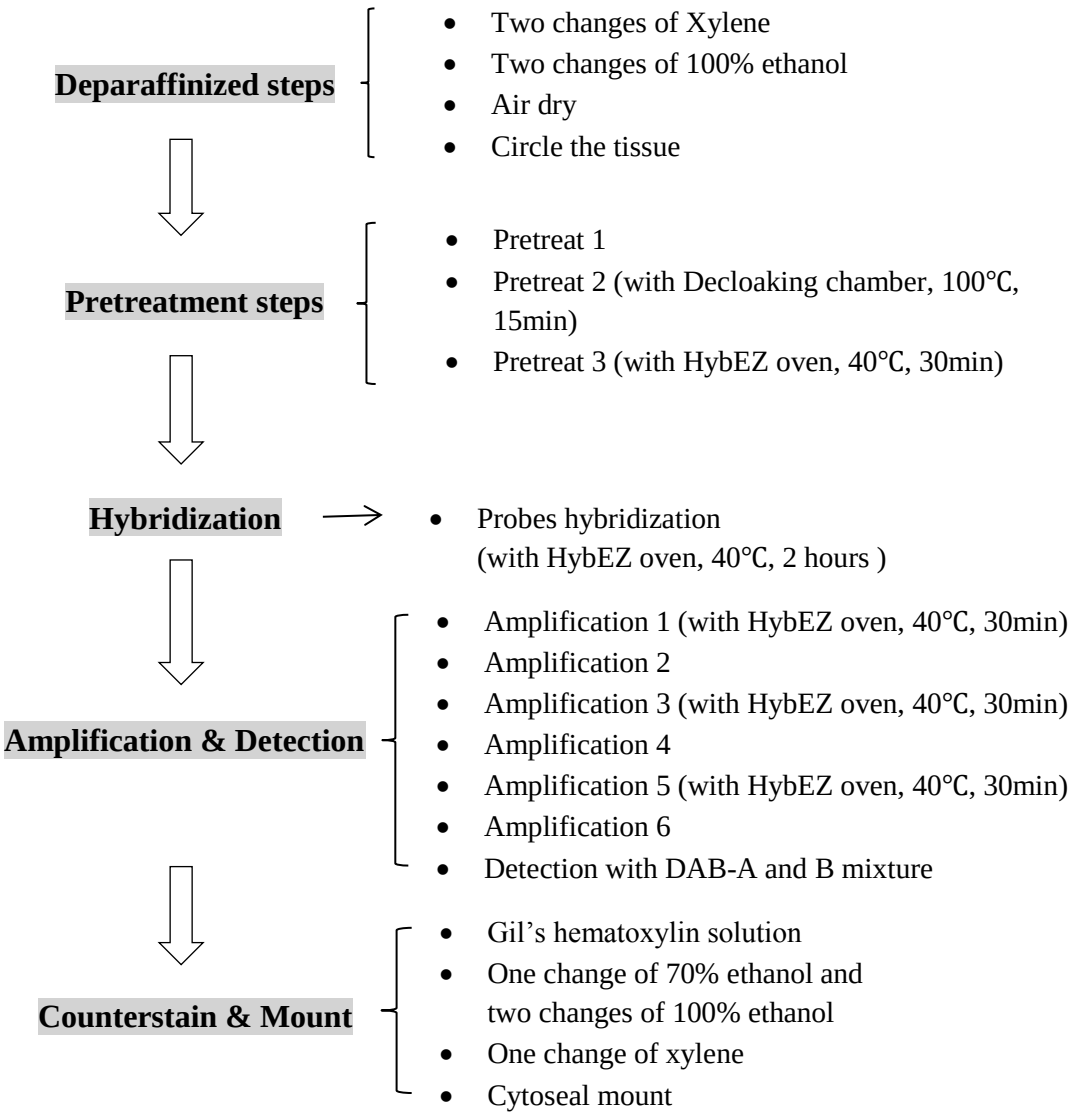


Figure 11. Work flow of modified RNAscope FFPE in situ hybridization assay

Immunohistochemistry

Three antibodies (EZH2 (D2C9) XP(R) Rabbit mAB: Cell Signaling, Danvers, MA; p53 (Y5): Thermo Fisher Scientific, Waltham, MA; CDKN1C/p57 antibody [EP2515Y], N-term: GeneTex, Irvine, CA) were tested on five tissue microarray slides. Several tests were done to optimize antigen retrieval time, antibody concentration and antibody hybridization duration, which usually varies depends on antigens. The overall protocol is similar to standard chromogenic immunohistochemistry with HRP protocol that was established by Experimental Pathology laboratory of the University of Vermont, including deparaffinization, antigen retrieval, non-specific antigen block, primary antibody hybridization, secondary antibody incubation, DAB staining and counterstain. The general processes are shown in figure 12.

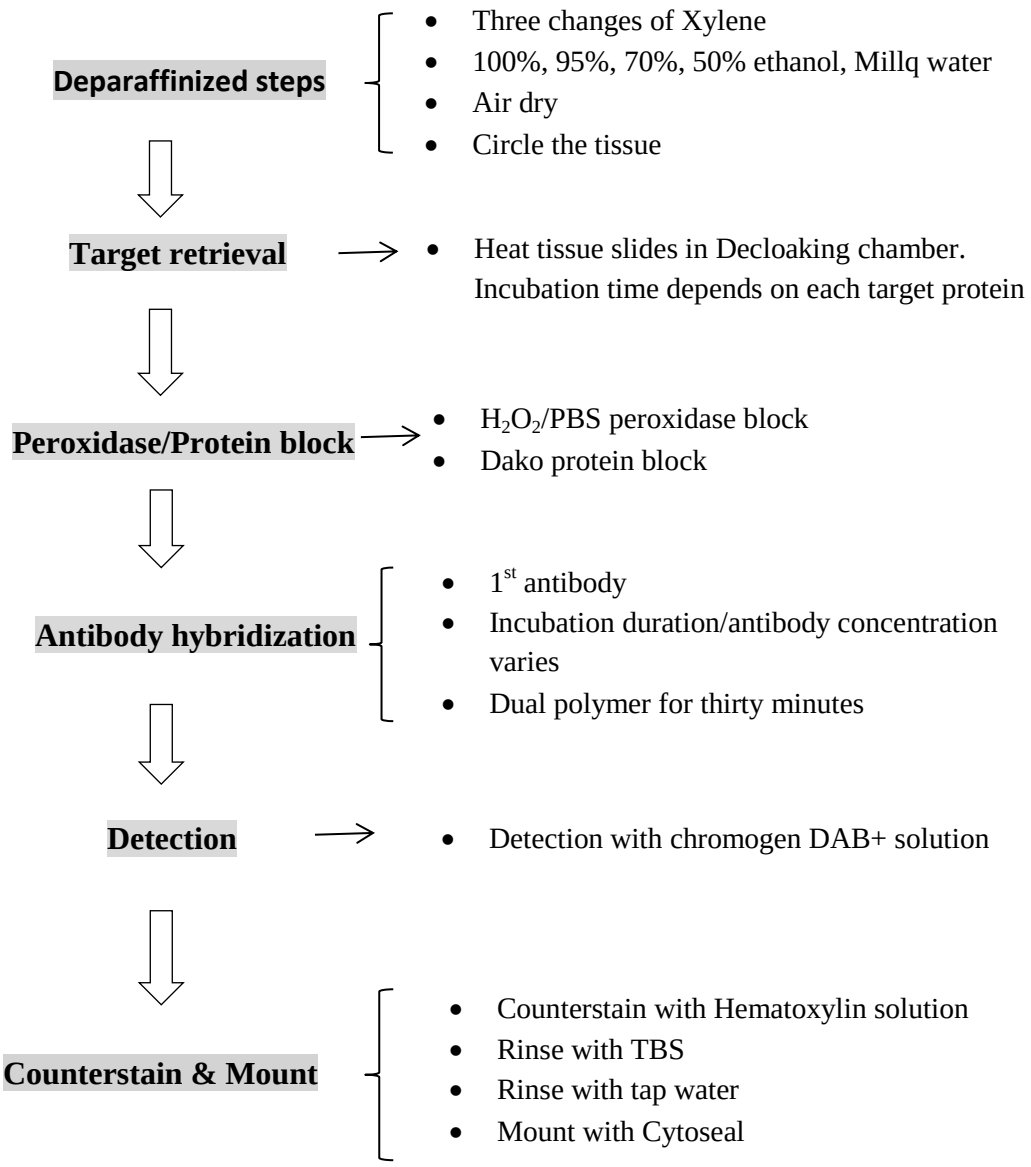


Figure 12. Work flow of immunohistochemistry

RNA in situ hybridization/ Immunohistochemistry evaluation criteria

Both RNA and protein markers expression level were scored. Two score systems were used: visual scoring system and RNAscope SpotStudio™ software scoring system. Visual scoring system has been well established for years to assess immunohistochemistry and *in situ* hybridization stain, so all our markers were scored by eye. RNAscope SpotStudio™ software is a automated, semi-quantitative scoring software that specializes in scoring RNAscope based images. Since it was designed to only recognize and analyze punctate dot, it was only used to provide score for lncRNA HOTAIR and KCNQ1OT1.

- Visual scoring system

  RNA markers

Bright field microscope was used to evaluate stain by experienced pathologist, Dr. Donald Weaver. According to various staining patterns, different scoring criteria were applied and an overall stain grade (SG) was generated by adding up all subgrades. Besides overall stain grade, we also categorized lncRNA and protein markers expressions into 1) four scale patterns, including negative, weak positive, moderate positive and strong positive and 2) two-tiered dichotomous score, which represents low and high expressional group. For HOTAIR, three variables: stain intensity, copy number per cell and stain proportion, were used. Stain intensity was divided into two levels, which are low to medium and medium to high, and was assigned score of 1, 2, respectively. Copy number per cell was stratified into low (1) and high groups (2) based on visual experience and three copies per cell was the cut-off value. Upon stain proportion, a generic rule was used for all RNAs and proteins except for lncRNA H19 because of its special staining format. We gave a case score of 0 once it was not stained or it only had positive stained cells less than 10%; score of 1 was given if one case had about 10-50% positive cells; 2 was assigned if it had positive cells ranging from 50-75% and 3 was for cases which had more than 75%

cells positive. So, by adding up three subgrades, the overall stain grade for HOTAIR ranged from 2 to 7. Similar to HOTAIR, we applied intensity, copy number per cell and stain proportion to assess KCNQ1OT1. The only difference was the cut off value of copy number per cell in KCNQ1OT1 was 2, which meant any cases with majority cell stained more than 2 copies was called high, and vice versa. Therefore, the overall stain grade of KCNQ1OT1 was ranged from 0 to 7. We assessed H19 by copy intensity and proportion. Copy density was given from 0 to 3, which represents negative, low, medium and high, and proportion was given from 0 to 3, showing four different portion levels: negative, dots, patchy and diffuse. Thus, the overall score of H19 was from 0 to 6. MEG3 was assessed by stain intensity, which we assigned 1, 2, 3 to low, moderate and high intensity, respectively, and stain proportion, which was the same as that of HOTAIR and KCNQ1OT1. In general, stain grade of MEG3 was from 0 to 6. Regarding four-tiered scoring pattern for HOTAIR and KCNQ1OT1, any case with at least two of three scoring variables in lowest end were weak positive, any cases with high copy number, strong stain, and more than 50% of cells stained were in strong positive. The rest cases were all in moderate positive category. For H19 and MEG3, any subjects with both two scoring variables scored no greater than 1 were weak positive, subjects scored more than 5 were strong positive, indicating one of the two variables must be highest and the other should be at least second highest, while the others were all moderate positive. However, KCNQ1OT1, H19 and MEG3 had another group of negative case. Additionally, we dichotomized all markers into low and high group to further reduce categories. To HOTAIR, we gave 0 to low HOTAIR group, in which cases with score of 2-3, and 1 to cases with score of 4-7, suggesting any case with at least two of three scoring variables in lowest end were low and the others were high. To KCNQ1OT1, the rationale of two-tiered system was both negative cases and any positive subject had all three scoring variables in lower end were given a 0, while others were 1. To H19, cases which either were negative or had low copy density with signal as dot, was in low and the rest were in high group. To MEG3, the

cut off value is 1, indicating a case that either were negative or had weak stain in less than 10% cells was 0.

Table 3. Visual scoring system for lncRNA

a) HOTAIR

	stain intensity		copy number per cell		stain proportion			
HOTAIR	low – medium	medium –high	0-3	>3	<10%	10%-50%	50%-75%	>75%
	1	2	1	2	0	1	2	3

b) H19

	copy intensity			stain proportion		
H19	low	medium	high	dot	patchy	diffuse
	1	2	3	1	2	3

c) KCNQ1OT1

	stain intensity		copy number per cell		stain proportion			
KCNQ1OT1	low – medium	medium -high	0-2	>2	<10%	10%-50%	50%-75%	>75%
	1	2	1	2	0	1	2	3

d) MEG3

	stain intensity			stain proportion			
MEG3	low	Medium	High	<10%	10-50%	50-75%	>75%
	1	2	3	0	1	2	3

Table 4. Four tiered pattern and dichotomous scoring system of lncRNA

lncRNA	Four tiered scoring system				Dichotomous scoring system	
	Neg	W	M	S	Low group (score=0)	High group (score=1)
HOTAIR	SG=0	SG=2-3	SG=4-5	SG=6-7	SG=2-3	SG=4-7
H19	SG=0	SG=1-2	SG=3-4	SG=5-6	SG=0-2	SG=3-6
KCNQ1OT1	SG=0	SG=2-3	SG=4-5	SG=6-7	SG=0-3	SG=4-7
MEG3	SG=0	SG=1-2	SG=3-4	SG=5-6	SG=0-1	SG=2-6

Neg= negative; W= weak positive; M= moderate positive; S= strong positive; SG=stain grade

- Protein markers

We applied similar scoring system to assess protein makers. EZH2 was assessed based on intensity and proportion as MEG3. Intensity was categorized into low, moderate and high with a score of 1, 2 and 3, respectively; Proportion was set as other lncRNAs. The overall stain grade of EZH2 ranged from 0 to 6 and pattern was segregated into negative (SG=0), weak positive (1-2), moderate positive (3-4), strong positive (5-6). For further simplification, stain grade of 1 was used as cut off value in EZH2 and rationale is same as MEG3. Other protein markers belong to clinical markers, including ER, PR, Her2, Ki67 and p53, so we scored them from a clinical perspective and dichotomized them into positive and negative. For ER, any case with more than 10% cells expressed ER was called ER positive, and so does PR; For Her2, score was given to each case from 0 to 3 based on Her2 expression level. Any case with a score of 0 or 1 was considered as Her2 negative, while cases with a score of 3 were called Her2 positive. Confirmation of Her2 status by dual color *in situ* hybridization was undergone once we found Her2 score of 2 by immunohistochemistry. For Ki67, 15% was the cutoff point, suggesting Ki67 positive was given to those cases with more than 15% cells expressed Ki67 protein. Ki67 protein was scored as 1 to 4 based on percentage of cells that were stained: <5%=1; 5-10%=2; 10-15%=3, >15%=4. P53 status was simply defined by whether it was stained or not.

Table 5. Visual scoring system for EZH2

	stain intensity			stain proportion			
EZH2	low	Medium	High	<10%	10-50%	50-75%	>75%
	1	2	3	0	1	2	3

- Other clinicopathological factors

Other clinical information including DCIS nuclear grade, Nottingham grade, invasive tumor size, invasive lymph nodes and lymphovascular invasion (LVI) status were recorded in original diagnosis. In order to simplify variables that are continuous or have multiple groups, we stratified them into fewer groups. We dichotomized DCIS nuclear grade to 1 that represents cases with an original DCIS grade of 3, and 0 that represents any cases with a score less than 3, instead of having original five grades (1, 1-2, 2, 2-3, 3). Nottingham grade was initially scored from 3 to 9, while we applied three-tiered histological grade on Nottingham scores based on the following way: grade 1 tumors have a total score of 3-5; grade 2 tumors have a score of 6-7; grade 3 tumors have a score of 8-9. In terms of invasive tumor size, we trichotomized in the following way: 1 represents in a tumor less than 2cm, 2 represents in a tumor between 2cm-5cm, 3 stands for all cases with a tumor greater than 5cm. Subsequently, we dichotomized tumor size with a cutoff value of 2: any invasive lesion smaller than 2cm was assigned as 0, while any lesion bigger than 2cm was assigned as 1. Invasive lymph node was also dichotomized based on its status: any case with a positive invasive lymph node was scored as one, and vice versa. LVI status was dichotomized originally.

RNAscope® SpotStudio™ software

In order to apply RNAscope software to provide semi-quantitative results for individual case, all TMA slides were firstly scanned by Ventana® iScan Coreo system (Ventana Medical System, Inc. Tucson, AZ) with high definition (HD) resolution. Scanned images of HOTAIR and

KCNQ1OT1 were then imported to RNAscope SpotStudio™ software as a JP2000 file for further analysis. The other lncRNAs were ineligible for this assessment for reasons of high (MALT1) or negligible (Zfas1) expression or stromal expression (H19 and MEG3). In terms of settings, for both lncRNAs, we kept default settings for image resolution that was 0.25, nucleus nucleus diameter that was 7.5μm and spot diameter which was 1.2μm since they all fit our cases well. Parameters of hematoxylin stain and spot stain were adjusted to better delineate cell regions and recognize brown spots. For HOTAIR, spot diameter of 1.2μm was selected after previews and kept for all cases to guarantee consistency for analysis. However, we optimized hematoxylin stain parameter for each sample due to different background hematoxylin stain. We applied lower hematoxylin stain level to those cases which had lighter nucleus stain and vice versa. For KCNQ1OT1, 1.2μm spot diameter was also selected to all our samples and we repeated optimizing hematoxylin stain parameter case by case. The range of hematoxylin stain value was from 0.05 to 0.15. Regions of interest, which was primarily regions that would be analyzed, were selected manually with the settings we optimized. An average of three to four regions of epithelia cells in each case, were delineated manually for software calculations. We also revised on regions of interest we selected by manually deleted those cells which were fake epithelia cells or have fake spots in that area once after batch run to avoid bias. Once we satisfied, we exported the results part and thus we were able to see what score of estimated spots per cell the software calculated.

Data analysis

All statistical tests were performed with SYSTAT version 11. The comparisons of lncRNAs expression level (lncRNA score, pattern RNAscope® software results) between different tissue types were analyzed by non-parametric Friedman two-way analysis of variance (ANOVA). It was firstly applied to test lncRNAs expression difference in cases containing all three tissue types. On the ground of P value < 0.05, subsequent paired comparisons between each two tissue types were

tested with the same method. The correlations between lncRNA expression and clinicopathological factors in both DCIS and invasive cancer were analyzed by Pearson correlation test and non-parametric tests, Mann-Whitney U test for two groups and Kruskal-Wallis one-way analysis of variance (ANOVA) test for groups ≥ 3. Unadjusted pearson correlation coefficient was tested on the null hypothesis that there is no correlation between variables in order to highlight potential associations ($p < 0.05$) and subsequent Kruskal-Wallis test or Mann-Whitney test were used to further analyze associations. Since Kruskal- Wallis evaluates only evaluate differences in mean ranks to assess the null hypothesis that the medians are equal across the group, Pearson Chi-square test and Fisher's exact test were used to compare proportional difference across groups, most for dichotomized variables. Variables with P value < 0.05 in univariate analysis were also used in the subsequent multivariate logistic regression test. Spearman rank correlation tests were used to assess whether RNAscope® SpotStudio™ software can produce consistent lncRNA score as eye scoring system.

Results

Sample size and clinical information

Tissue microarray contains total of 36 NA breast tissue, 34 DCIS tissue and 43 invasive breast tissue collected from 46 patients. 2 of 46 patients were diagnosed as lobular carcinoma *in situ* (LCIS), which were excluded in our study. Patient ages ranged from 30 to 86 years (mean 58.4); tumor size from 0.04cm to 13.5cm (mean 1.8cm); DCIS nuclear grade 1(n=1), 1-2(n=4), 2(n=26), 2-3(n=2), 3(n=11), invasive Nottingham histologic score from 4 to 9 (mean6.2). Invasive histological tumor grade was determined by Nottingham histologic total score based on criteria mentioned previously: grade 1 (n=17); grade 2 (n=12) and grade 3 (n=13). Thus, our sample contains invasive tumor grade 1(17), 2(n=15), 3(n=12). Positive lymph node was detected in 13/36(36.1%) and positive lymphovascular invasion (LVI) was found in 15/44(34.1%). In DCIS,

ER, PR and Her2 statuses were also evaluated in our study by IHC, while their statues in invasive cancer were recorded when patients received diagnosis. In DCIS, we had ER positive 26/30(86.7%), PR positive 24/30 (80%), Her2 positive 3/16 (18.8%), hormone receptor positive 26/30(86.7%), triple negative 1/16 (6.3%); In invasive cancer, we had ER positive 40/44(90.9%), PR positive 36/44(81.8%), Her2 positive 5/43(11.6%), hormone receptor positive 40/44(90.9%), triple negative 3/43(7.0%). Ki67 and p53 were stained by immunohistochemistry and dichotomized. Ki67 positive were 6/18(33.3%) and 13/35(37.1%) in DCIS and invasive cancer, respectively, while p53 positive were 2/20 (10%) in DCIS and 2/33(6.1%) in invasive cancer. A summary of patient clinicopathological characteristics is shown in Table 6.

Table 6. Summary of patient clinicopathological and molecular markers characteristics

Variables	Value
Age(years)	
Range	30-86
Mean	58.4
DCIS nuclear grade	
1	1 (2.2)
1-2	4 (9.1)
2	26 (59.1)
2-3	2 (4.5)
3	11 (25)
Invasive histologic grade	
1	17 (38.7)
2	15 (34.1)
3	12 (27.2)
Lymph node status	
Positive	13 (29.5)
Negative	23 (52.3)
NA	8 (18.2)
Tumor size (cm)	
Range	0.04-13.5
Mean	1.8
lymphovascular invasion status	
Positive	15 (34.1)
Negative	29 (65.9)
Molecular markers (DCIS)	
ER positive	26 (86.7)
PR positive	24 (80)
Her2 overexpression (IHC + CISH)	3 (18.8)
Triple negative ($ER^-/PR^-/Her2^-$)	1 (6.3)
Ki67 positive	6 (33.3)
P53 positive	2 (10)
Molecular markers (invasive cancer)	
ER positive	40 (90.9)
PR positive	36 (81.8)
Her2 overexpression (IHC + CISH)	5 (11.6)
Triple negative ($ER^-/PR^-/Her2^-$)	3 (7)
Ki67 positive	13 (37.1)
P53 positive	2 (6.1)

Demonstration of tissue cores on TMA

When construct TMA, normal adjacent breast, DCIS and IC tissue regions were marked on original H&E slides to guide core punch. To reconfirm tissue specificity on TMA, 5 microarray blocks were cut to make 5μm slides, stained with H&E, and then reviewed by pathologists (DL). Within NA tissue cores on TMA, several tissue cores only containing collagen and (or) adipocyte were excluded in HOTAIR and KCNQ1OT1 analysis but not in H19 and MEG3 because both H19 and MEG3 was stained mostly in stromal interspace while the rest two lncRNAs were all found in epithelia cells. And quite a few tissue cores which were supposed to be pure DCIS or IC ended up with a mixture of DCIS and IC, which were graded and recorded separately. Tissue core H&E stains were displayed in Figure 13.

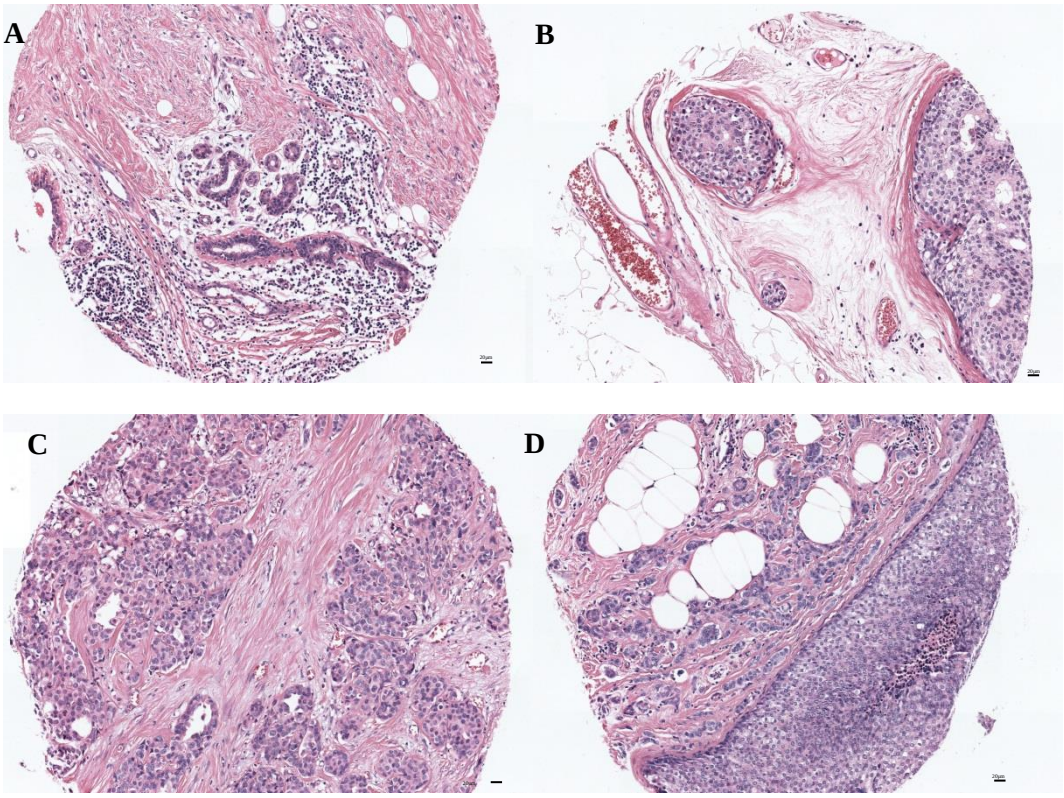


Figure 13. H&E stain of tissue cores

Image represents H&E stain of different tissue cores on TMA, (A)normal adjacet tissue; (B)DCIS tissue; (C) invasive cancer tissue; (D) mixture of DCIS and invasive cancer tissue(Left: invasive; Right: DCIS). All images were taken with 10X objective. [Scale bar: 20 μm]

Detection of lncRNA by RNAscope® chromogenic in situ hybridization assay (CISH)

To ensure the lncRNA targeting specificity of RNAscope® assay platform, RNase A and DNase I digestion steps were performed to all lncRNA probes and compared with standard staining procedures. We confirmed that RNAscope® in situ hybridization assay was able to detect all lncRNAs properly on FFPE tissue slides with standard staining procedures or even plus an addition of DNase I treatment, however, lncRNA cannot be stained after RNase A treatment step which intends to remove all ribonucleic acid. Two illustrations are shown in Figure.13.

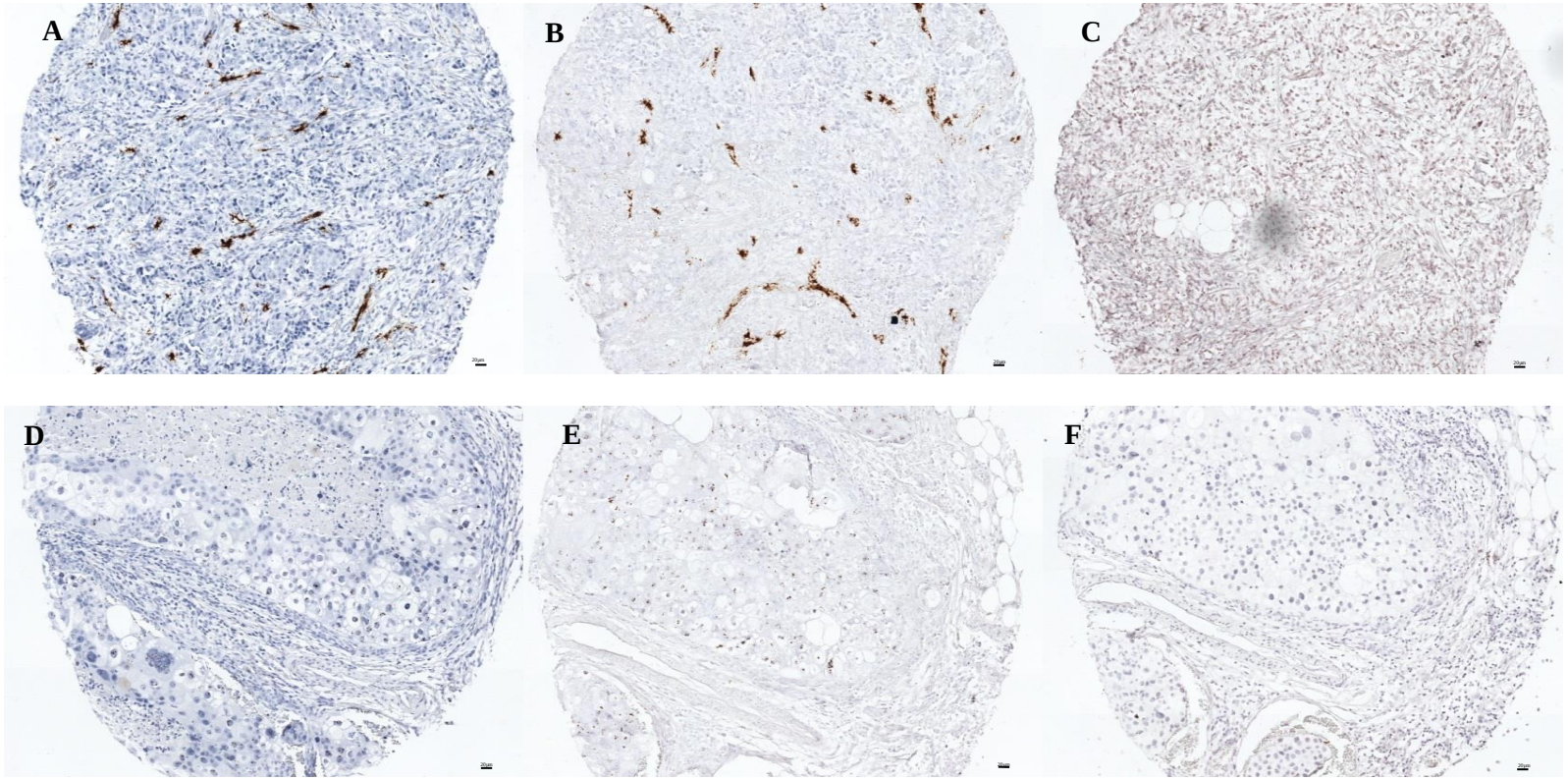


Figure 14. RNAscope® in situ hybridization assay specifically detect RNA molecules

Images (A-C) shows the detection of H19 in invasive caner tissue, by standard CISH(A) by standard CISH plus a step of DNase I digestion (B by standard CISH plus RNase A treatment (C). Images (D-F) shows the detection of KCNQ1OT1 in invasive cancer tissue, by standard CISH (D); by standard RNAscope® CISH plus DNase I digestion (E) by standard CISH plus RNase A treatment (F). All images were taken using a 10X objective. [Scale bar: 20 µm]

Long non-coding RNA staining results

- HOTAIR

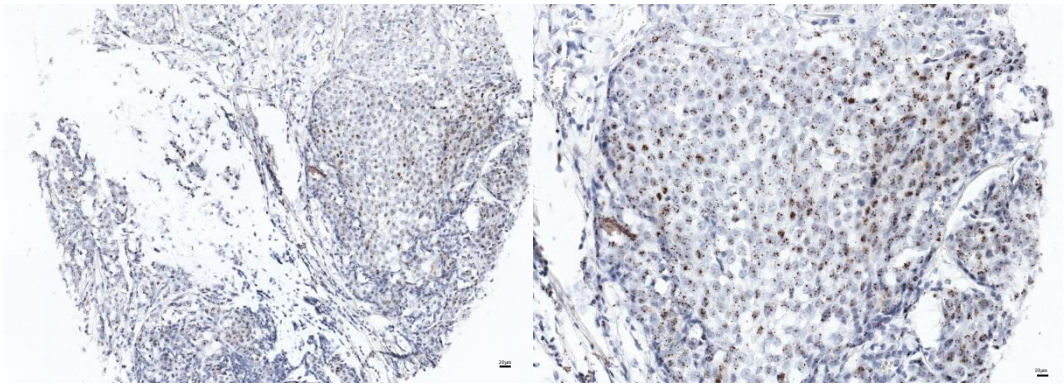
HOTAIR staining was widely present as single of multiple dots in epithelia cell nuclei in all three tissue types. Within TMAs probed with HOTAIR, 26 NA tissue regions, 25 DCIS regions and 32 invasive cancer tissue regions were scored by eye and RNAscope® SpotStudio™ software. HOTAIR expression level between different tissue types were compared within same individuals for patients who had two or three tissue spots. For HOTAIR, 11 patients had all three tissue types, 15 patients kept NA breast tissue and DCIS, 20 patients kept NA and invasive breast cancer tissues, and 17 patients had both DCIS and invasive cancer tissues. In NA breast tissue, HOTAIR was scored from two to five: 2(n=2), 3(n=21), 4(n=1), 5(n=2) with a mean of 3.42 by eye and in terms of pattern distribution of HOTAIR in NA breast tissue, 23 cases were weak positive and only 3 cases were moderate positive. By RNAscope® SpotStudio™ software, HOTAIR scores varied from 0.07 to 1.25 with a mean of 0.54 in NA breast tissue. In DCIS regions, HOTAIR was given a score by eye from 3 to 7: 3(n=4), 4(n=5), 5(n=7), 6(n=4), 7(n=5) with a mean of 5.04. Four of them were weak positive, twelve were moderate positive and the rest nine were strong positive. From the perspective of software, 0.1 was given as the minimum while 13.78 was given as the maximum and the mean value was 2.17. In invasive tissues, they had HOTAIR scores from three to seven: 3(n=7), 4(n=3), 5(n=5), 6(n=5), 7(n=12) and the mean is 5.38. Here, we had 17 cases in strong positive, 8 cases in pattern in moderate positive and 7 cases in weak postive. Summary of HOTAIR stain is listed in Table 7. However, the score assigned by software ranged more broadly, from 0.25 to 17.1 with a mean of 2.34. Figure 15 shows different HOTAIR expression levels in terms of eye based scoring criteria and their corresponding RNAscope SpotStudio™ results.

Table 7. Summary of HOTAIR stain

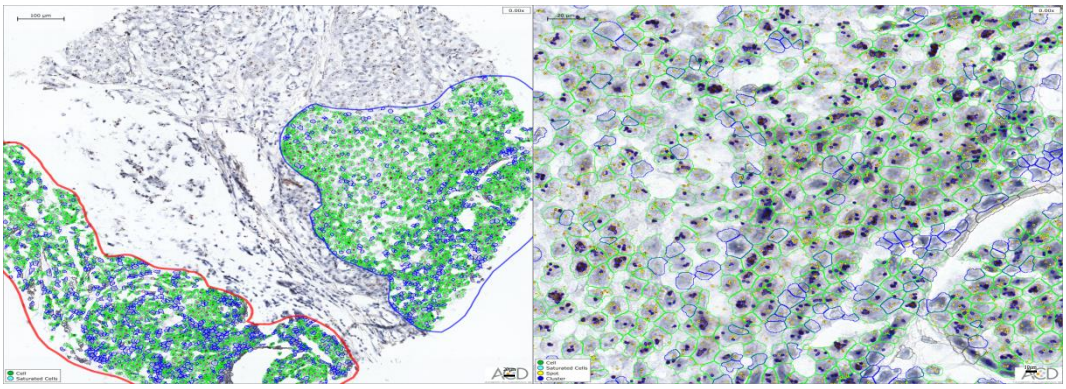
Tissue	n	mean stain grade	Four tiered pattern				Dichotomous system	
			Neg	W	M	S	Low	High
NA	26	3.42	0	23	3	0	23	3
DCIS	25	5.04	0	4	12	9	4	21
IC	32	5.38	0	7	8	17	7	25

Neg= negative; W=weak positive; M=moderate positive; S=strong positive

A

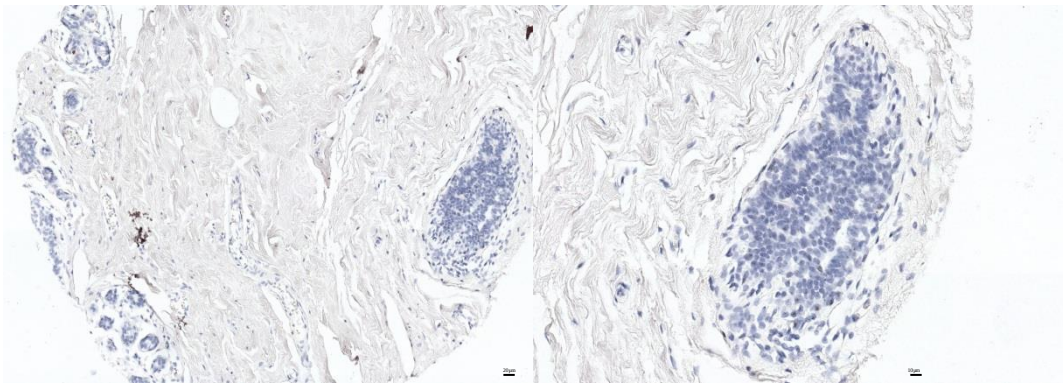


In this DCIS spot, HOTAIR signal was considered moderate to high intensity, with more than three copies per cell in more than 75% epithelia cells. Images were taken in 10X objective lens (left) [scale bar: 20 µm] and 20X lens (right) [scale bar: 10 µm] for same region.

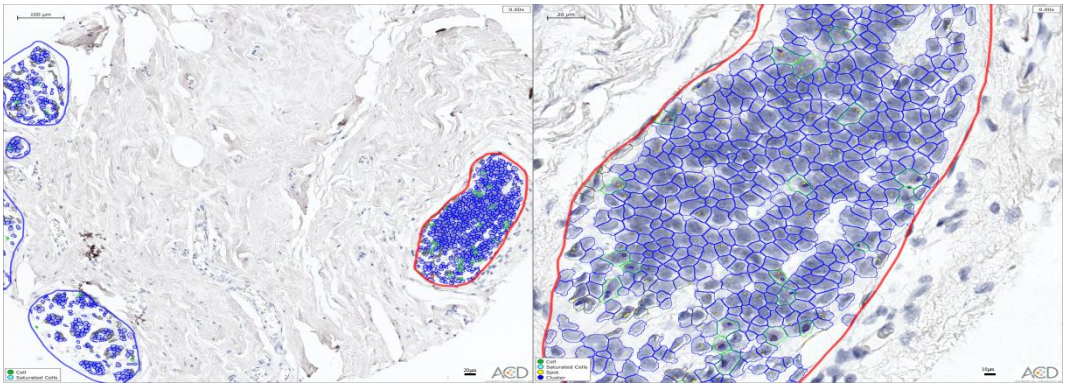


This set of images show results of case shown above, analyzed by RNAscope® SpotStudio™ software package. Green lined cells represent cells which were estimated by software to have more than three copies of RNA per cell, while blue circled cells displays cells which have less than three copies per cell. However, black circled cells were unwanted cells which have been eliminated manually. Yellow dot represents single copy of HOTAIR, and blue shows HOTAIR clusters. In this case, 3526 epithelia cells were selected to have an estimated 8.32 HOTAIR copies per cell. Images were taken in 10X objective lens (left) [scale bar: 20 µm] and 20X lens (right) [scale bar: 10 µm] for same region.

B



In this normal ajacent tissue spot, HOTAIR signal was recorded low to moderate intensity, with less than three copies per cell in about 10-50% epithelia cells. Both left images were taken with 10X objective [scale bar: 20 µm] and right images were same regions as left, taken by 20X microscope lens[scale bar: 10 µm]



In this case, RNAscope® SpotStudio™ software was used to analyzed 951 epithelia cells. The overall estimated HOTAIR copies per cell was 0.82. Images were taken in 10X objective lens (left) [scale bar: 20 µm] and 20X lens (right) [scale bar: 10 µm] for same region

Figure 15. The expression patterns of HOTAIR by CISH

- H19

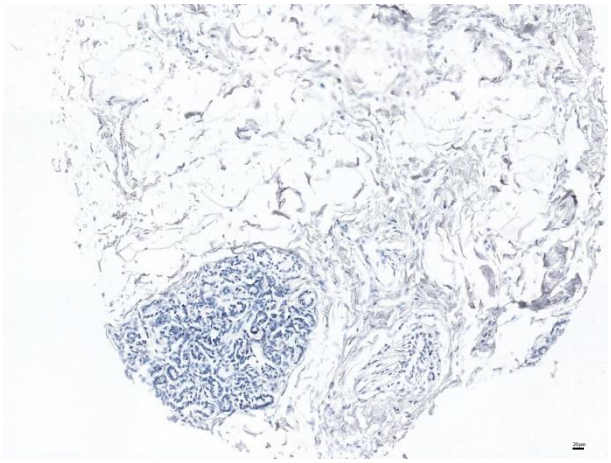
H19 was stained mostly in the stromal component around breast ducts with some possibilities of epithelia cell stain as concentrated dots. Within TMAs probe with H19, 40 adjacent tissue spots including normal adjacent epithelia cells and other normal spaces, 25 DCIS tissue spots and 35 invasive cancer tissue spots were scored by eye. To compare H19 expression between different tissue types within same individuals, patients who had no less than two different tissue types were sorted out. For H19, 19 patients kept all three tissue spots, 23 patients had both NA and DCIS tissue spots, 33 patients had both NA and invasive tissue spots and 20 individuals owned DCIS and invasive cancer regions. In NA tissue regions, H19 was scored from 0 to 3: 0(n=22), 1(n=1), 2(n=14), 3(n=3) with a median of 0.95, and the 22 of them was recorded as negative and 15 cases were weak positive and 3 cases were in moderate positive category. In DCIS spots, fewer cases (n=9) were scored as 0. More cases were assigned to a higher score: 2(n=8), 3(n=3), 4(n=3), 5(n=2). The mean score of H19 in DCIS was 1.88. In terms of pattern distribution, 8 cases were negative; 9 cases were weak positive; 6 cases were pattern moderate positive and last 2 cases fit in strong positive group. In invasive cancer spots, H19 expressed broader: 0(n=4), 2(n=1), 3(n=4), 4(n=10), 5(n=7), 6(n=9) and the mean value was 4.08. Here, we had 16 cases in strong positive, 14 cases in moderate positive, 1 case in weak positive and 4 were negative. We summarized H19 stain in Table 8. Figure 16 shows different H19 expression levels in terms of different scoring criteria.

Table 8. Summary of H19 stain

Tissue	n	mean stain grade	Three tiered system				Two tiered system	
			Neg	W	M	S	Low	High
NA	40	0.95	22	15	3	0	37	3
DCIS	25	1.88	9	8	6	2	17	8
IC	35	4.08	4	1	14	16	5	30

Neg= negative; W=weak positive; M=moderate positive; S=strong positive

A



This image represents H19 negative in NA tissue [scale bar: 20 µm]

B

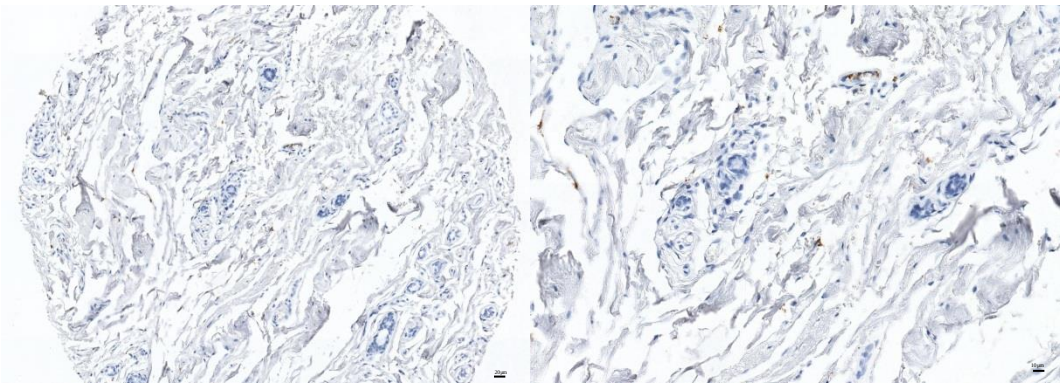
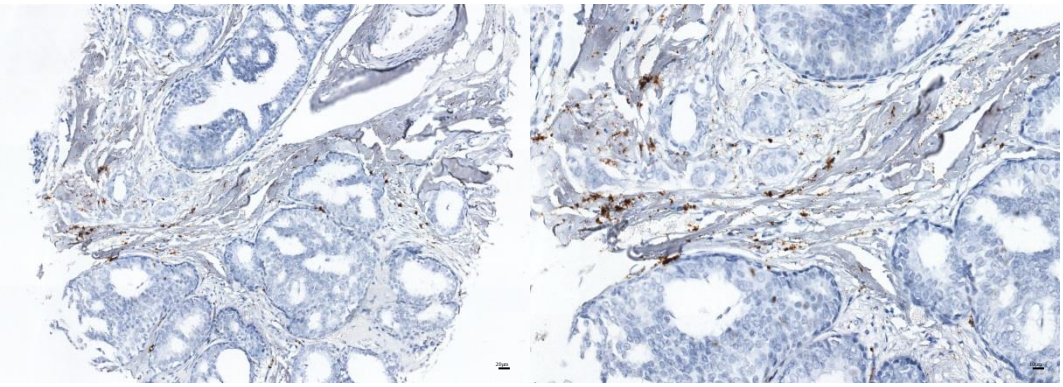


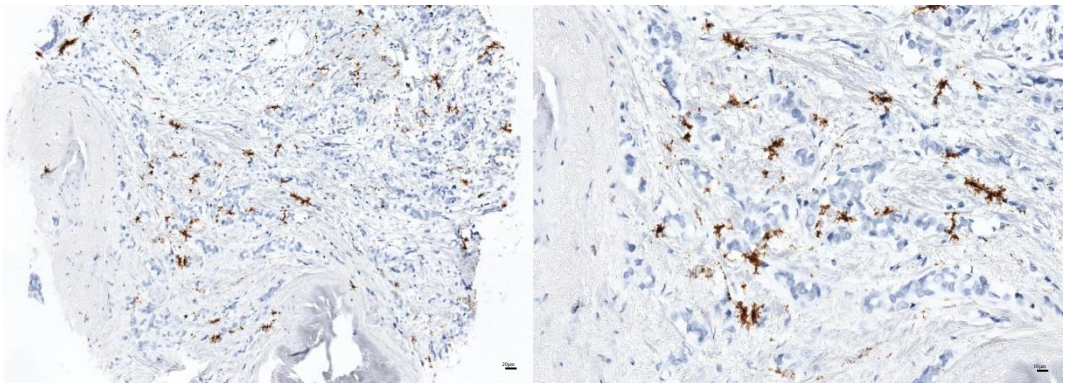
Image shows H19 staining mainly in dots of low intensity in one NA tissue. Images were taken in 10X objective lens (left) [scale bar: 20 µm] and 20X lens (right) for same regions [scale bar: 10 µm]

C



H19 singal was detected in patch pattern with a moderate stain intensity in this DCIS spot. Images were taken in 10X objective lens (left) [scale bar: 20 µm] and 20X lens (right) for same case [scale bar: 10 µm]

D



In this invasive cancer spot, H19 signal was diffusely found in strong intensity. Images were taken in 10X objective lens (left) [scale bar: 20 µm] and 20X lens (right) for same region [scale bar: 10 µm]

Figure 16. Expression illustrations of H19 by CISH

- KCNQ1OT1

Similar to staining pattern of HOTAIR, KCNQ1OT1 was mostly found in cancer epithelia cell nuclear as punctate dots in all kinds of tissue. Within TMAs probed with KCNQ1OT1, 30 NA cell spots, 25 DCIS spots and 32 invasive cancer spots were scored by both eye and RNAscope® SpotStudio^{TM} software. 10 individuals who kept all three types of tissue, 17 individuals who had both NA and DICS tissues, 21 patients who had both NA and invasive cancer tissues and 16 patients who had DCIS and invasive cancer tissues were sorted out for further tests. Regarding to KCNQ1OT1 scores and pattern, all tissue types were assigned a broad range of score. For NA spots, scores were from 0 to 6: 0(n=5), 2(n=7), 3(n=13), 4(n=1), 5(n=3), 6(n=1) with a mean of 2.6. To fit in our four tiered score system, 5 were negative, 20 cases were in weak positive, 4 cases were moderate positive and 1 strong positive. By software, KCNQ1OT1 scores varied from 0 to 1.73 with a mean of 0.45. In DCIS, we had KCNQ1OT1 score from 0 to 7 (mean=3.76): 0(n=1), 2(n=3), 3(n=8), 4(n=5), 5(n=5), 6(n=2), 7(n=1). In terms of pattern in DCIS, weak positive (n=11) and moderate positive (n=10) had more number than negative (n=1) and strong positive (n=3). However, score given by software was narrower, from 0 to 3.46 with a mean of

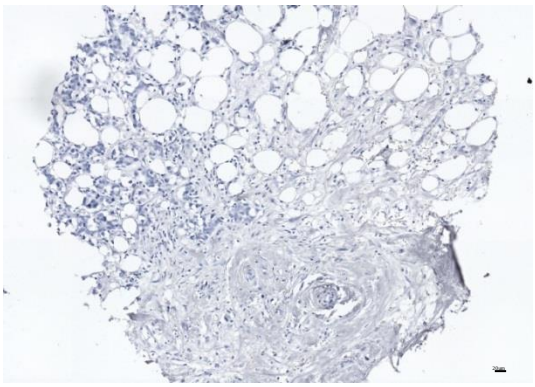
0.86. In invasive cancer spots, visual score range was from 2 to 7 (mean=4.4): 2(n=4), 3(n=7), 4(n=5), 5(n=8), 6(n=4), 7(n=4). 11 of them were weak positive, 12 were in moderate positive and 9 were in strong positive group. List of KCNQ1OT1 summary is shown below in Table 9. RNAscope® SpotStudio™ assigned lower scores than eye (mean=0.97): the minimum was 0.06 and the maximum was 4.95. Figure 17 shows different KCNQ1OT1 expression levels in terms of eye based scoring criteria and their corresponding RNAscope® SpotStudio™ results

Table 9. Summary of KCNQ1OT1 stain

Tissue	**n**	**mean stain grade**	**Four tiered system**				**Two tiered system**	
			Neg	W	M	S	Low	High
NA	30	2.60	5	20	4	1	25	5
DCIS	25	3.76	1	11	10	3	12	13
IC	32	4.40	0	11	12	9	11	21

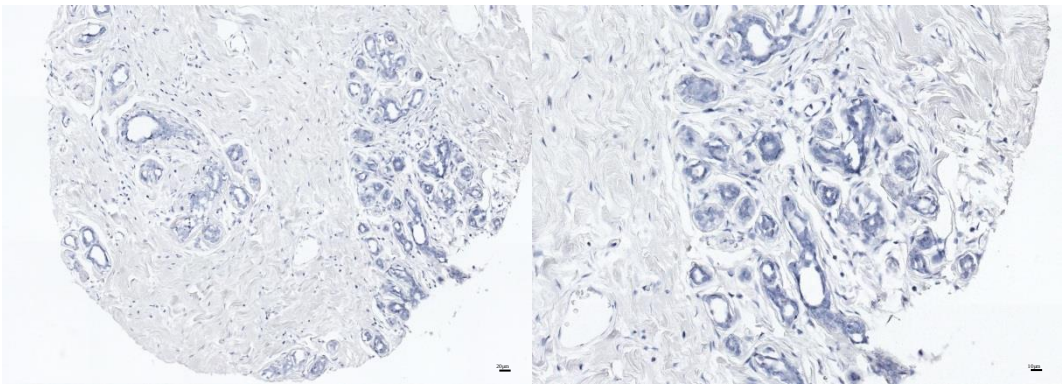
Neg= negative; W=weak positive; M=moderate positive; S=strong positive

A

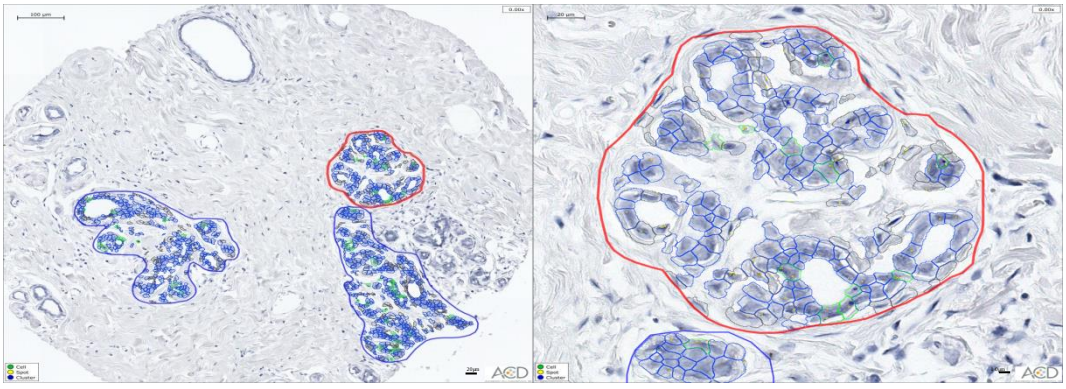


KCNQ1OT1 negative in invasive breast tissue [scale bar: 20 µm]

B

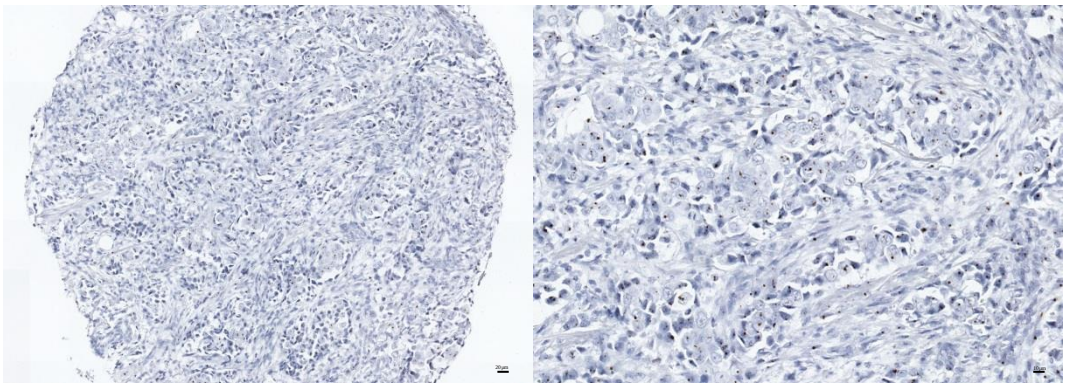


In this NA tissue, KCNQ1OT1 was stained as punctate dots in low to moderate intensity, with a one to two copies per cell in less than 10% epithelia cells. Images were taken in 10X objective lens (left) [scale bar: 20 µm] and 20X lens (right) for same region [scale bar: 10 µm]

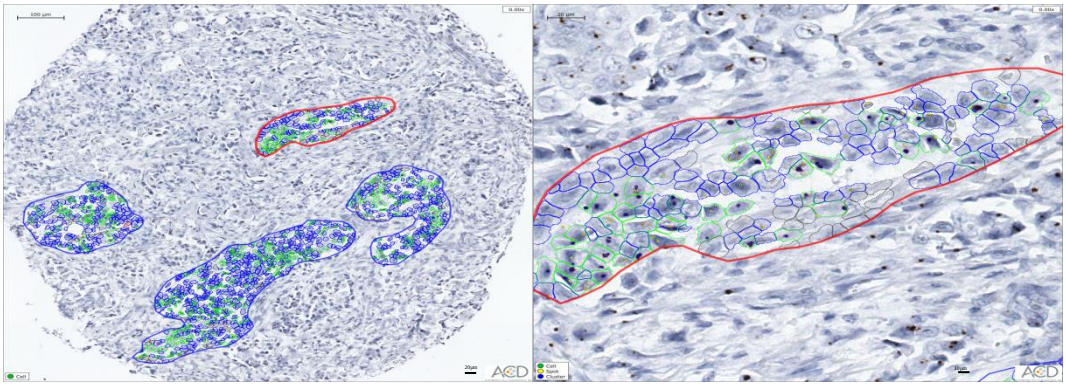


The results generated by RNAscope® SpotStudio™ software package for above case were shown here . Green lined cells represent cells which were estimated by software to have more than two copies of KCNQ1OT1 RNA per cell, while blue circled cells displays cells which have less than two copies per cell. However, black circled cells were unwanted cells which have been eliminated manually. Yellow dot represents single copy of KCNQ1OT1, and blue shows KCNQ1OT1 clusters. In this case, 753 epithelia cells were selected, which turned out to have an overall estimated 0.35 copies of KCNQ1OT1 per cell.

C



In this invasive breast cancer case, KCNQ1OT1 was stained as punctate singals in moderate to high intensity, with more than two copies per cell in less about 50-75% epithelia cells. Images were taken in 10X objective lens (left) [scale bar: 20 µm] and 20X lens (right) for same region [scale bar: 10 µm]



Here, In this case, RNAscope® SpotStudio™ software was used to analyze 1165 selected epithelia cells to have an overall estimated 1.83 copies of KCNQ1OT1 per cell. Images were taken in 10X objective lens (left) [scale bar: 20 µm] and 20X lens (right) for same region [scale bar: 10 µm]

Figure 17. CISH stain of KCNQ1OT1

- MEG3

Similar to staining pattern of H19, MEG3 mainly localized in stromal cells around epithelia ducts as nuclear punctuate stain. 4/91(4%) cases with minor staining in epithelia cells were excluded in our study. Within TMAs probed with MEG3, 34 NA tissue spots, 23 DCIS tissue spots and 34 invasive cancer spots were scored by eye by pathologist (DL). In MEG3, there were 14 patients having all three tissue types, 19 patients having NA and DICS tissues, 27 patients having NA and invasive cancer tissues and 18 cases keeping DCIS and invasive cancer tissues. The mean value of MEG3 score in NA tissue spots is 1.76 and the range was from 0 to 6: 0(n=21), 2(n=2), 3(n=3),

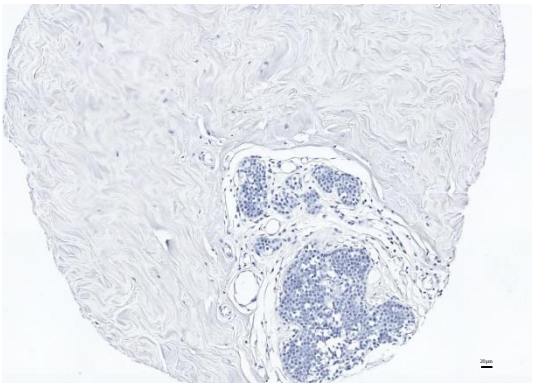
4(n=3), 5(n=1), 6(n=5). 21 cases with a score of zero were regarded as negative and from weak positive to strong positive, we had 2, 6 and 6 cases, respectively. In DCIS, we had fewer cases in each score (mean=1.3) and pattern: 0(n=14), 2(n=3), 3(n=2), 4(n=3), 6(n=1); negative (n=14), weak positive (n=3), moderate positive (n=5), strong positive (n=1). Among 34 cases of invasive cancer spots, 10 were scored 0 and regarded as negative; 2 were scored 2 and grouped as weak positive; 17 were scored 3 (n=4) or 4 (n=14) and grouped as moderate positive and last 2 cases were strong positive with a score of 5. The mean score of MEG3 in invasive spots was 1.7. Table 10 shows MEG3 stain summary and figure 18 shows different MEG3 expression levels in terms of different scoring criteria.

Table 10. Summary of MEG3 stain

Tissue	**N**	**mean stain grade**	**Four tiered system**				**Two tiered system**	
			Neg	W	M	S	Low	High
NA	34	1.76	21	2	6	6	21	14
DCIS	23	1.30	14	3	5	1	14	9
IC	34	1.7	10	2	17	2	10	21

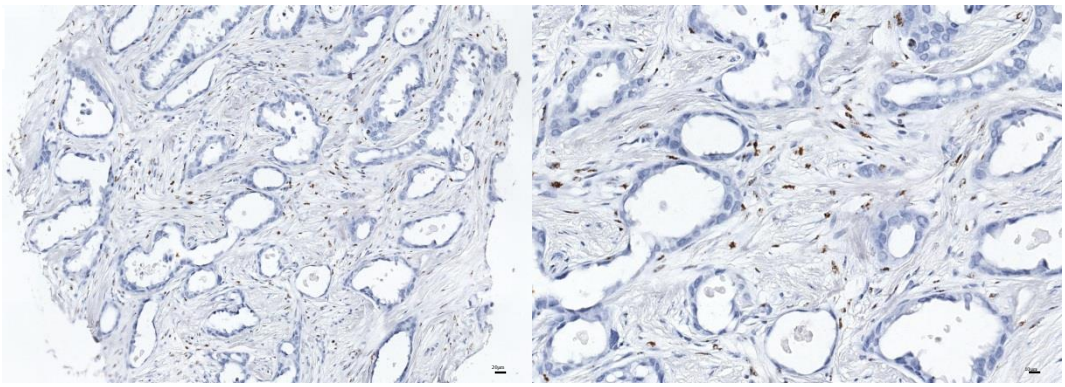
Neg= negative; W=weak positive; M=moderate positive; S=strong positive

A



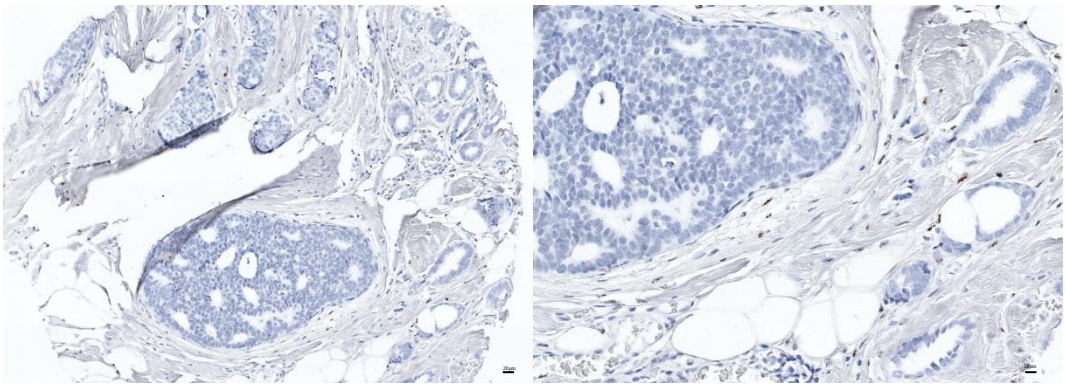
Negative stain of MEG3 on one NA breast case, image was taken with 10X objective lens [scale bar: 20 µm]

B



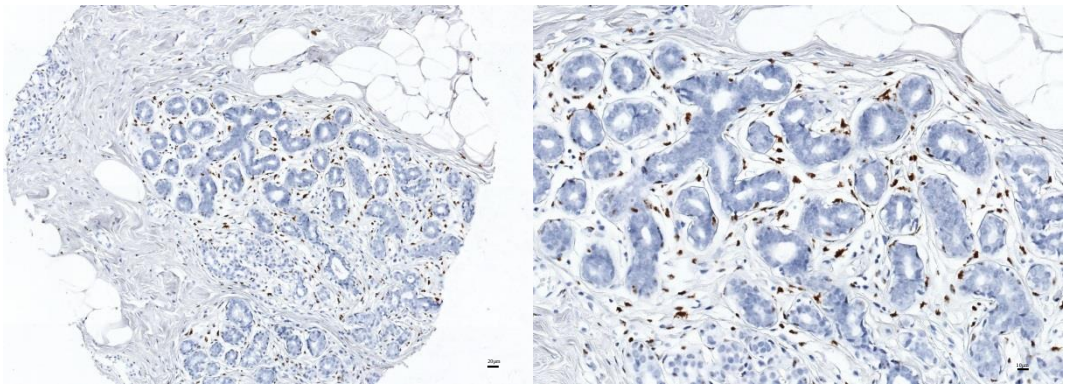
Here, MEG3 was stained in moderate intensity in more than 75% stromal cells in invasive cancer spot. Images were taken in 10X objective lens (left) [scale bar: 20 µm] and 20X lens (right) for same region [scale bar: 10 µm]

C



(C) shows that MEG3 was detected in 50-75% stromal cells, mainly in low stain intensity in DCIS. Images were taken in 10X objective lens (left) [scale bar: 20 µm] and 20X lens (right) for same region [scale bar: 10 µm]

D

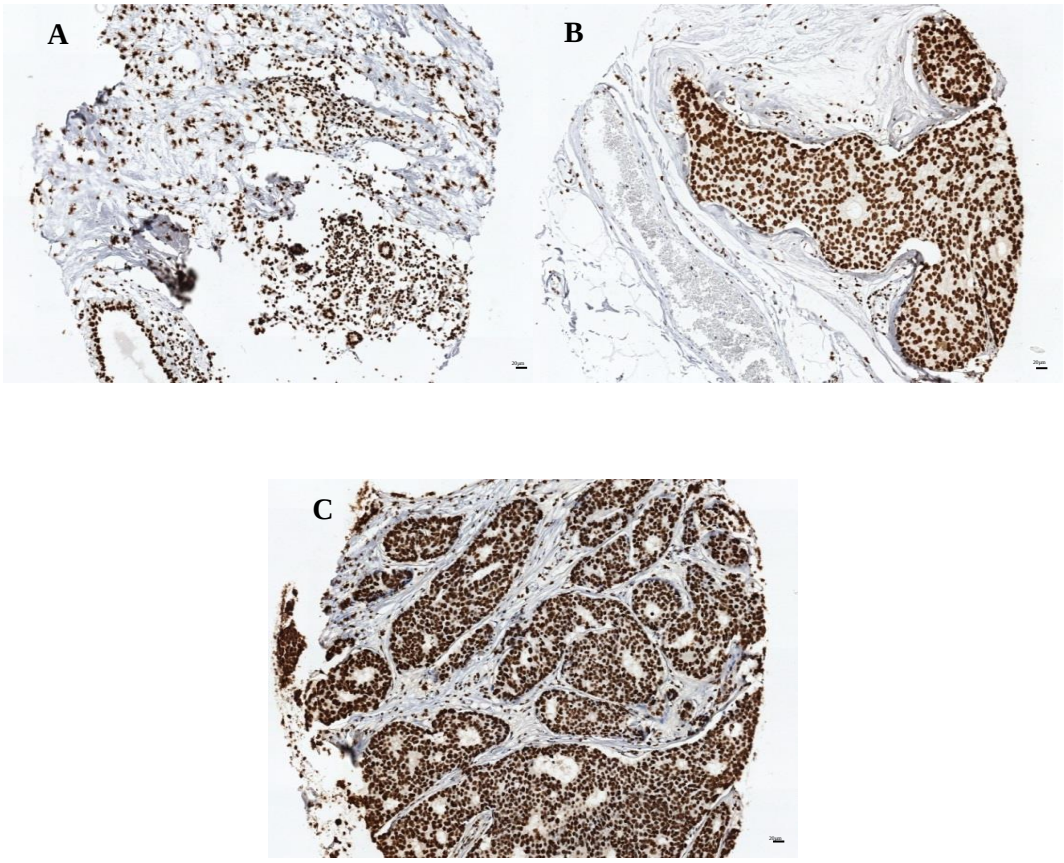


Images represent strong stain of MEG3 in more than 75% stromal cells in benign breast tissue. Images were taken in 10X objective lens (left) [scale bar: 20 µm] and 20X lens (right) for same region [scale bar: 10 µm]

Figure 18. Stain illustrations of MEG3

- MALAT1

MALAT1 was extensively and intensively stained in each cell across three different tissues. We were unable to show any stain difference in MALAT1 and thus we were unable to test any statistical significance between groups. Figure 19 illustrates the generic staining pattern of MALAT1.

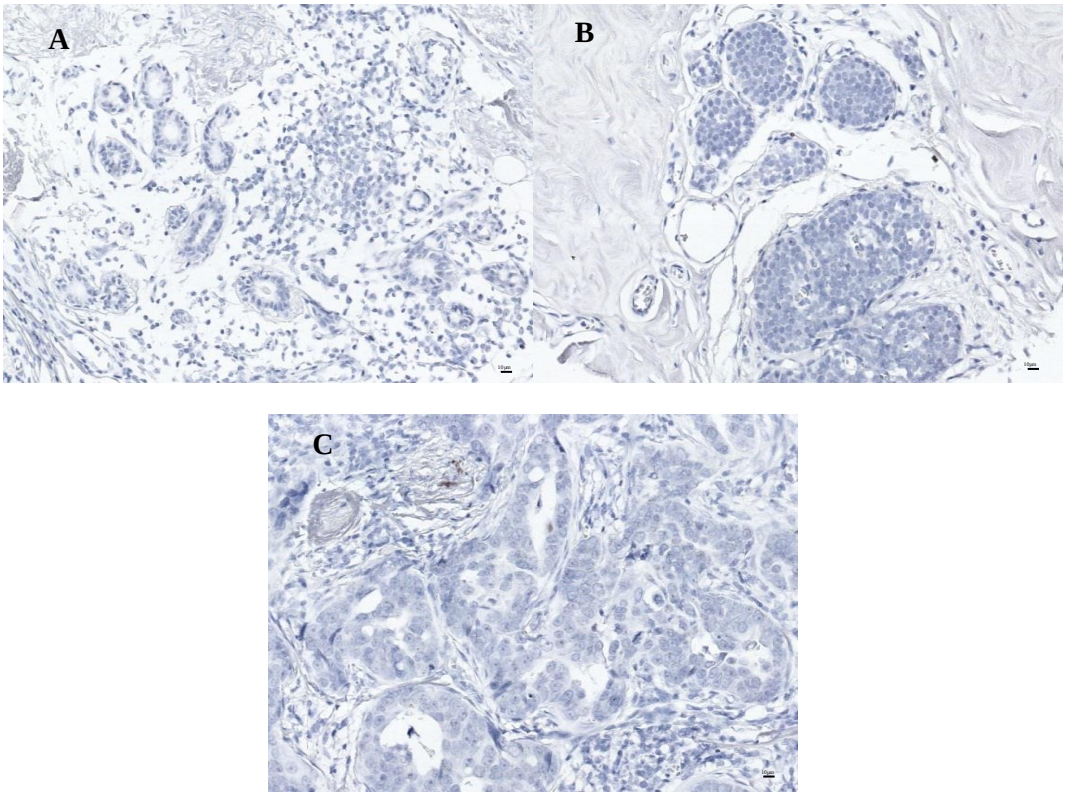


MALAT1 was universally strongly stained on NA tissue (A), DCIS (B) and invasive cancer tissue (C). All images were photographed with 10X objective lens. [scale bar: 20 μm]

Figure 19. Strong positive stain of MALAT1

- Zfas1

Zfas1 stained like HOTAIR and KCNQ1OT1, which were punctate dots in epithelia cell nucleus. Although we saw some minor difference in stain proportion, the overall staining intensity was every low and overall score of cases were very close. So, we were unable to perform statistical analysis to demonstrate any significance in Zfas1. Figure 20 illustrates the generic staining pattern of Zfas1.



Zfas1 was hardly seen in NA tissue (A); DCIS (B) and invasive cancer tissue(C). All images were taken with 20X objective lens. [scale bar: 10 μm]

Figure 20. Faint stain of Zfas1 across TMAs

Long non-coding RNA associated protein staining results

- EZH2

Staining in epithelia cell nucleus, EZH2 was more prevalent in cancer cells rather than normal adjacent epithelia cells. Across our TMAs, we gave EZH2 score for 28 NA spots, 21 DCIS spots and 31 invasive spots. Among all patients, 10 individuals had all three tissue types available; 16 individuals had both NA and DCIS spots; 16 individuals kept both NA and invasive cancer spots; 15 individuals had DCIS and invasive cancer spots. In NA spots, most of cases (23/28) were scored as 0 and two cases were 1 and one case was 2, 3, and 4, respectively. Thus, we had negative pattern (n=23), weak positive (n=2) and moderate positive (n=2). In DCIS regions, 9 cases were scored of zero and we had more score distributions: 1(n=2), 2(n=7), 4(n=1), 6(n=2). In regard to pattern, nine cases were grouped as negative, another nine cases were in weak positive, one case was in moderate positive and rest two cases were in strong positive. However, in invasive cancer spots, we had 15 cases as EZH2 negative while other scores include 1(n=4), 2(n=4), 3(n=1), 4(n=3), 5(n=2), 6(n=2). Eight of them were in weak positive, four in moderate positive and three in strong positive. A summary of EZH2 stain is listed in Table 11. Figure 21 shows different EZH2 expression levels according to different scoring criteria.

Table 11. Summary of EZH2

Tissue	**N**	**mean stain grade**	**Four tiered system**				**Dichotomous scoring system**	
			Neg	W	M	S	Low	High
NA	28	0.39	23	2	3	0	25	3
DCIS	21	1.57	9	9	1	2	11	10
IC	31	1.58	15	8	4	3	19	12

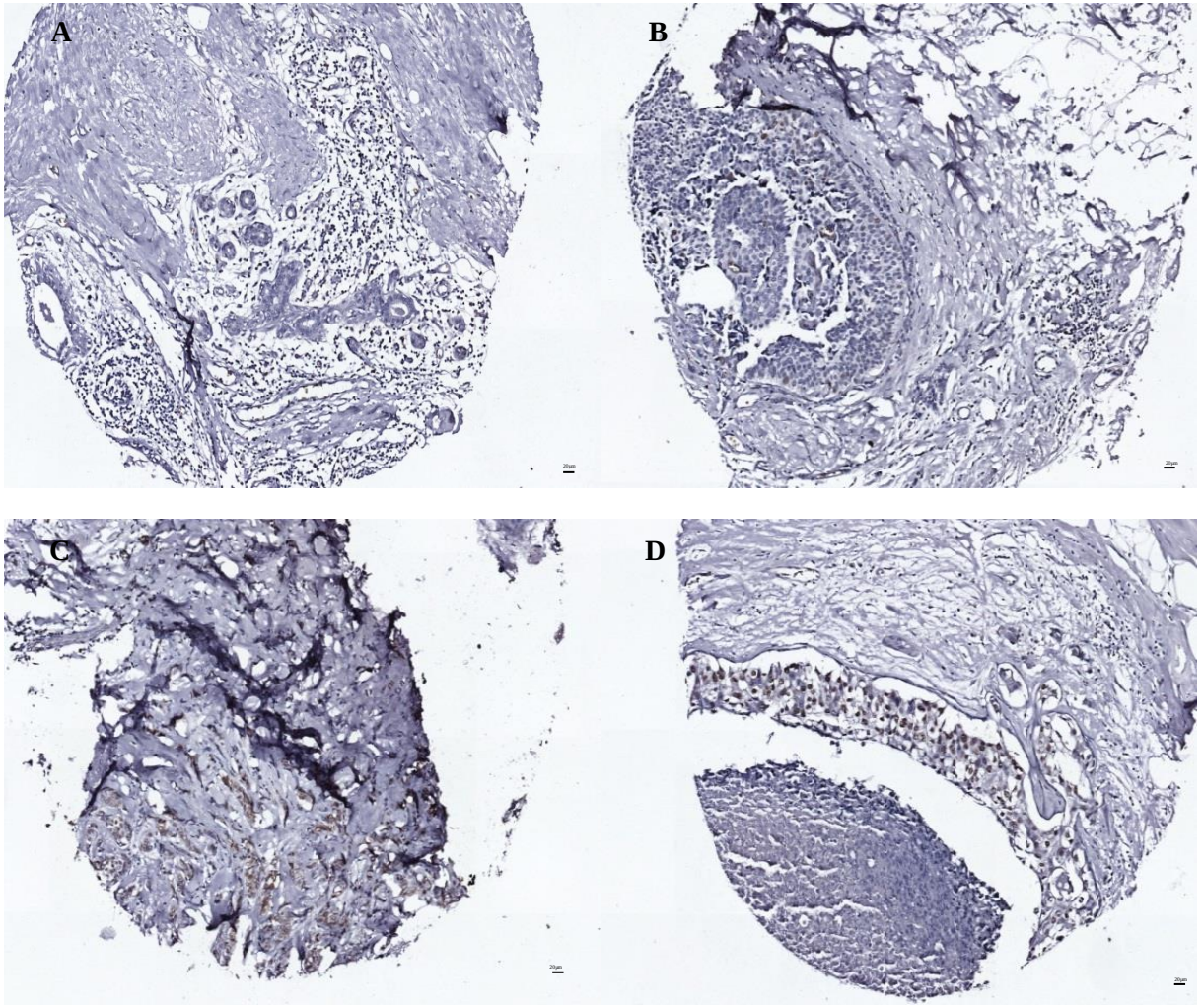


Figure 21. Different expressions of EZH2 across samples (A) Negative case of EZH2 in normal adjacent breast tissue; (B) EZH2 was found less than 10% epithelia cells with low stain intensity in benign breast tissue ;(C) EZH2 appeared in 50-75% epithelia cells in moderate intensity in invasive cancer; (D) Image shows EZH2 strong stain in more than 75% epithelia cells in DCIS (all images were taken with 10X lens) [scale bar: 20 µm]

- CDKN1C/p57

CDKN1C was widely (>50%) and strongly stained in nucleus including epithelia, glandular and adipocyte cell nucleus in any NA spots and cancer spots, like what we found from MALAT1. Therefore, we were unable to show any statistical difference between each tissue type. Here, Figure 22 illustrates the generic staining format of CDKN1C.

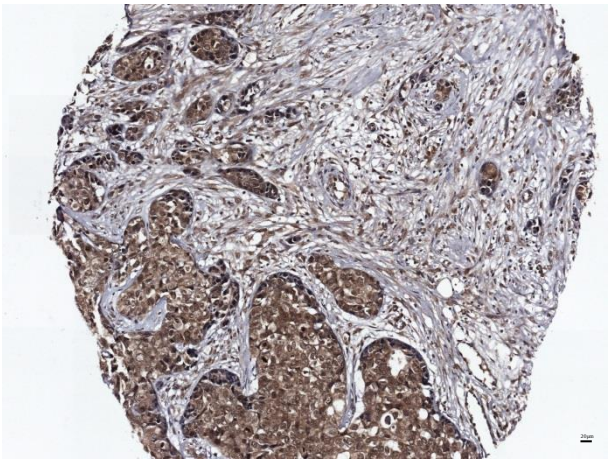


Figure 22. CDKN1C generic stain

CDKN1C extensively expressed in nucleus with high intensity. Image was photographed with 10X lens [scale bar: 20 µm]

- P53

Unlike CDKN1C, p53 was hardly stained in our TMAs. No p53 was found across all NA tissue spots. In our cancer spots, we found 10% (2/20) positive cases in DCIS and 6.3% (2/32) positive cases in invasive cancer. All positive cases had nuclear stain of p53, however, only one case

showing cytoplasmic staining was excluded in our study. Both p53-positive and p53-negative cases are displayed in Figure 23.

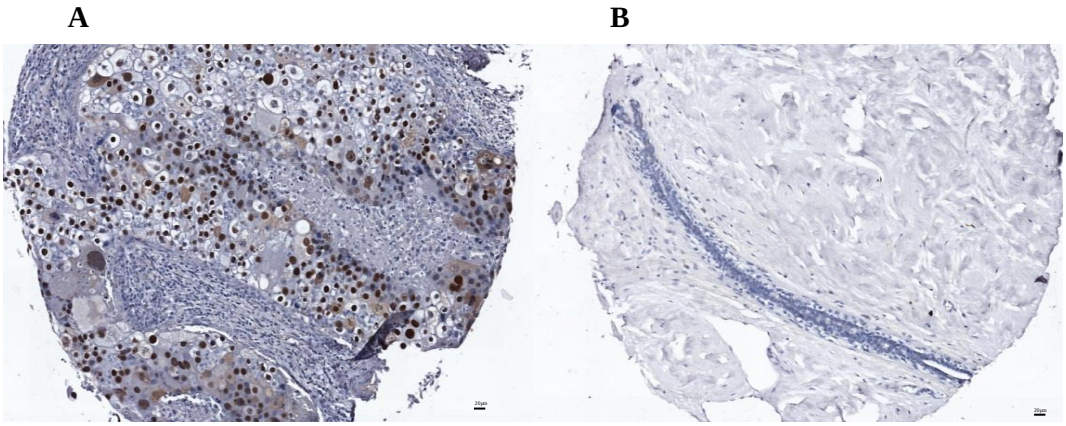


Figure 23. Positive and negative illustrations of p53

(A) p53 positive in invasive cancer (B) p53 negative in NA breast, all images are in 10X objective lens [scale bar: 20 μm]

Clinicopathological factors staining results

Basic clinicpathological data was obtained from patient records. In our study, we confirmed ER, PR and Her2 status, explored Ki67 (MIB1) and E-cadherin status by doing immunohistochemistry on our TMAs.

- Estrogen Receptor

We dichotomized ER status in DCIS and invasive cancer spots. In DCIS regions, ER positive were found in 26/30 cases and the rest was all ER negative. In invasive cancer tissue, we had 40/44 cases were ER positive and 4/44 were ER negative, which is in concord with the original patients information. ER staining illustrations are shown in Figure 24.

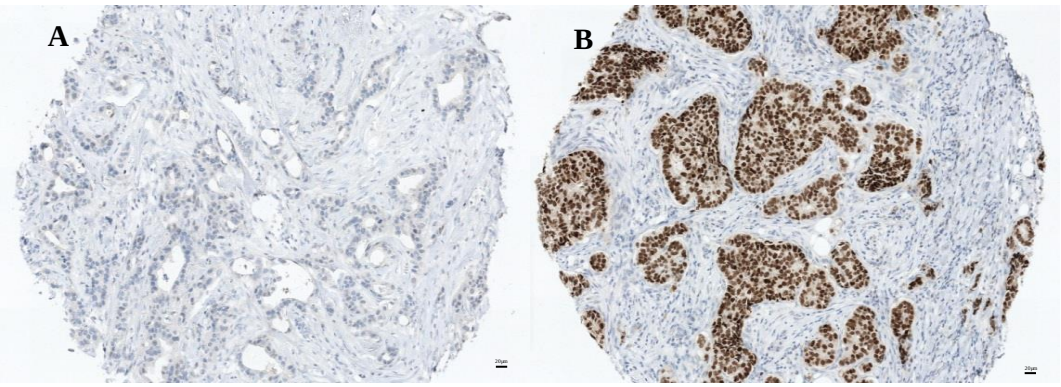


Figure 24. Positive and negative illustrations of ER

(A) ER negative (B) ER positive; Images were taken in 10X objective lens [scale bar: 20 µm]

- Progesterone receptor(PR)

PR status was tested in DCIS and invasive cancer tissues. 24 cases were found PR positive and 6 cases were found PR negative in DCIS spots, while 36 spots were PR positive and 8 spots were PR negative in invasive cancer. Data collected from TMA staining perfectly matched previous clinical information. PR staining examples are shown in Figure 25. Hormone receptor statues were 26/30 positive in DCIS and 40/44 positive in invasive cancer.

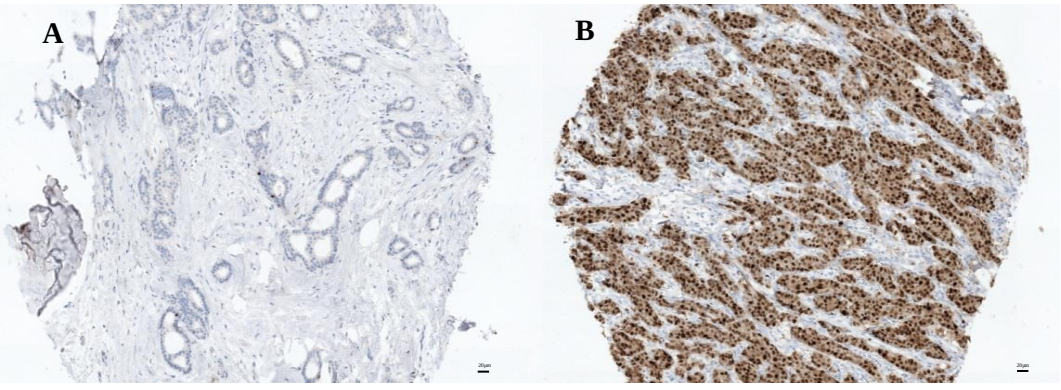


Figure 25. Positive and negative illustrations of PR

(A) PR negative (B) PR positive; Images were taken in 10X objective lens [scale bar: 20 µm]

- Her2

Her2 protein stained the membrane by IHC. In DCIS, there were no equivocal cases among all 16 spots that was scored: 0(n=8), 1(n=5), 3(n=3), so only 3 cases were recorded as Her2 positive. In invasive spots, we scored 0 for 22 cases, 1 for 6 cases, 2 for 2 cases and 2 for 3 cases. After reconfirmation by dual *in situ* hybridization (one centromerical probe for chromosome 17 and one Her2 probe, performed in the FAHC lab), those 2 equivocal cases were grouped as Her2 positive. In conclusion, 28 cases were Her2 negative and 4 were Her2 positive. Due to TMA construction issue, we missed several spots in our TMAs, resulting in fewer Her2 statues than original patient data, however, since all current TMA Her2 statues perfectly concords with original Her2 statues, we made up those missing data from original Her2 statue, resulting in an overall Her2 positive of 5 cases and Her2 negative of 38 cases. Figure 26 shows Her2 staining examples. (All images were photographed with 10x objective lens except dual CISH)

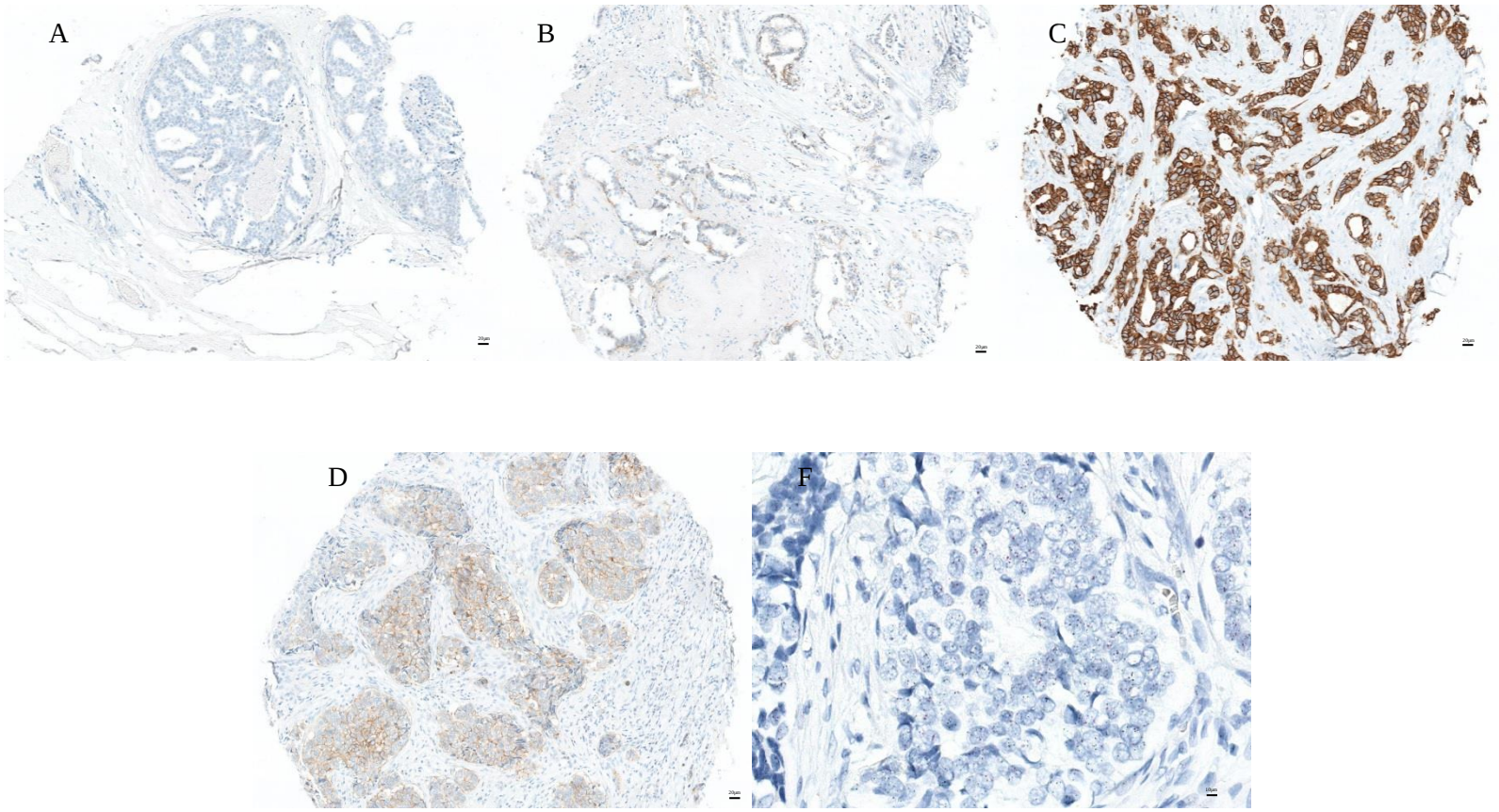


Figure 26. Examples of Her2 stain

(A) Image shows Her2 negative case with a score of zero (B) image represents Her2 negative case with a score of one (C) Her2 positive with a score of three (D) Her2 equivocal case with a score of two by immunohistochemistry (E) Dual color CISH confirmed overexpression of Her2 in equivocal case. Image was taken with 10x objective lens in (A-D) [scale bar: 20 µm] and 20x objective lens (F) [scale bar: 20 µm]

- Ki67 (MIB1)

Ki67 protein was stained mostly as single or multiple dots, sometimes in clusters in epithelia cells. In DCIS spots, Ki67 protein was scored as 1(n=3), 2(n=5), 3(n=4), 4(n=6) based on percentage of cells that were stained. So, only those six cases with a score of 4 were recorded as Ki67 positive. In invasive cases, we found more Ki67 positive case with a score of 4 (13/35). Among the other 22 cases which are Ki67 negative, 8 were scored 1, 10 were 2 and 4 were 3. Staining illustrations are shown in Figure 27.

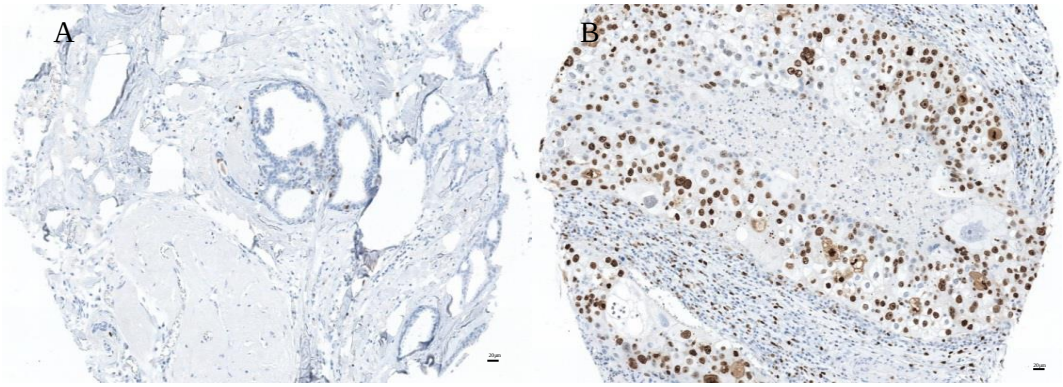


Figure 27. Illustrations of ki67 status

(A) shows Ki67 negative because it only expressed in 5% cells (less than 15% threshold) (B) represents Ki67 positive case [scale bar: 20 μm]

- E-Cadherin

E-cadherin was only stained in ductal carcinoma *in situ* or invasive ductal carcinoma, but negative in lobular carcinoma, shown in Figure.28.

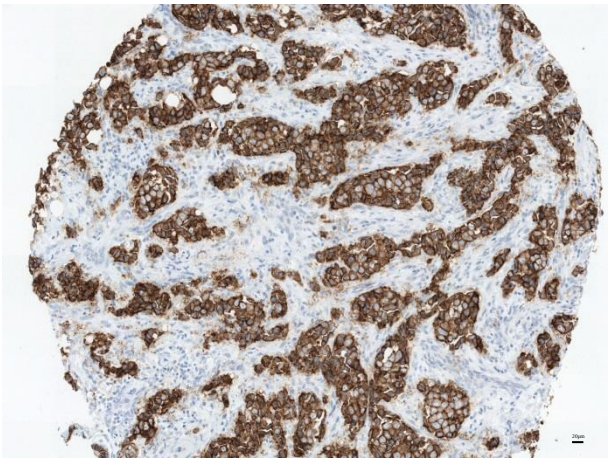


Figure 28. E-cadherin positive in invasive ductal carcinoma case [scale bar: 20 μm]

Statistical analysis

- Correlations between two independent scoring systems

In order to demonstrate the data consistency between eye-based scoring system and RNAscope® SpotStudio™ scoring software we used in for two lncRNA makers, HOTAIR and KCNQ1OT1, we used non-parametric Spearman rank correlation test to check the correlations (Table.12a and 12b). In HOTAIR, significant correlations between two systems were found in DCIS and invasive cancer subjects, while correlations were shown in all three tissue types in KCNQ1OT1.

Table 12. Spearman correlation study between two scoring systems

a. Spearman correlation test for HOTAIR in all three tissue types

Tissue type	Sample size (n)	Spearman's rho	Probabilities
NA	26	0.339	0.067
DCIS	25	0.497	<0.001
IC	32	0.716	<0.0001

b. Spearman correlation test for KCNQ1OT1 in all three tissue types

Tissue type	Sample size (n)	Spearman's rho	Probabilities
NA	30	0.749	<0.0001
DCIS	25	0.681	<0.0001
IC	32	0.769	<0.0001

- Comparison of lncRNA expression across tissues

To test our hypothesis that lncRNAs may express differently across tissue types, we applied non-parametric Friedman two-way analysis of variance test on subjects containing all three tissue types. From the perspective of eye scoring system, HOTAIR, H19 and KCNQ1OT1 expressions are significantly different across tissues. KCNQ1OT1 also has expression difference across our samples, while HOTAIR seems to have equal expression in different tissues in terms of score that produced by RNAscope® software (Table.13). Further paired comparison by same Friedman tests were applied only after overall tests was significant, resulting in more details upon different variables: the eye-determined expression levels of HOTAIR and H19 differs significantly in each paired comparison (Table.14a, Table.14b), however, both eye-determined and software determined KCNQ1OT1 score varies between normal adjacent and cancer, but not non-invasive and invasive (Table.14c).

Table 13. Non-parametric comparison for lncRNAs expression across tissues

lncRNA variable	NA tissue rank sum	DCIS tissue rank sum	Invasive cancer rank sum	Sample size (n)	Friedman Test Statistic	P value
HOTAIR *score*	13.5	23.5	29.0	11	15.9	<0.001
HOTAIR *pattern*	13.5	23.5	29.0	11	15.9	<0.001
HOTAIR *software*	17.5	21.0	27.5	11	4.9	0.086
H19 *score*	27.0	34.5	52.5	19	25.0	<0.001
H19 *pattern*	27.5	34.0	52.5	19	25.4	<0.001
KCNQ1OT1 *score*	11.5	22.0	26.5	10	13.5	0.001
KCNQ1OT1 *pattern*	13.0	22.5	24.5	10	10.4	0.005
KCNQ1OT1 *software*	12.0	23.0	25.0	10	9.8	0.007
MEG3 *score*	24.5	28.0	31.5	14	4.5	0.108
MEG3 *pattern*	25.0	28.0	31.0	14	3.4	0.180

Score represents the score was given by standard visual evaluation; *pattern* means the score was given by four tiered scoring system; *software* indicates the score was given by RNAscope® sotware

Table 14. Non-parametric paired comparisons for lncRNAs

a.Non-parametric paired comparison for HOTAIR

lncRNA variable	NA tissue rank sum	DCIS tissue rank sum	Invasive cancer rank sum	Sample size (n)	Friedman Test Statistics	P value
HOTAIR *score*	30.5	38.5	/	23	5.3	0.021
	34.5	/	64.5	33	30.0	<0.001
	/	23.0	37.0	20	12.3	<0.001
HOTAIR *pattern*	30.5	38.5	/	23	5.3	0.021
	35.0	/	64.0	33	29.0	<0.001
	/	23.0	37.0	20	14.0	<0.001

b. Non-parametric paired comparison for H19

lncRNA variable	NA tissue rank sum	DCIS tissue rank sum	Invasive cancer rank sum	Sample size (n)	Friedman Test Statistics	Probabilities
H19 *score*	17.5	27.5	/	15	10.0	0.002
	22.5	/	37.5	20	15.0	<0.001
	/	22.0	29.0	17	7.0	0.008
H19 *pattern*	17.5	27.5	/	15	10.0	0.002
	22.5	/	37.5	20	15.0	<0.001
	/	22.0	29.0	17	7	0.008

c. Non-parametric paired comparison for KCNQ1OT1

lncRNA variable	NA tissue rank sum	DCIS tissue rank sum	Invasive cancer rank sum	Sample size (n)	Friedman Test Statistics	Probabilities
KCNQ1OT1 *score*	21.0	30.0	/	17	6.2	0.013
	22.5	/	40.5	21	16.2	<0.001
	/	21.0	27.0	16	3.6	0.058
KCNQ1OT1 *pattern*	21.5	29.5	/	17	6.4	0.011
	23.5	/	39.5	21	16.0	<0.001
	/	21.5	26.5	16	2.8	0.096
KCNQ1OT1 *software*	19.0	32.0	/	17	9.9	0.002
	24.0	/	39.0	21	10.7	0.001
	/	24.0	24.0	16	0	1

- Association between lncRNA expressions and DCIS clinical markers

We also tested the hypothesis that lncRNA expression may associate with DCIS clinical markers. Pearson correlation test was firstly conducted to screen potential correlations (Table 15). Here, we kept DCIS nuclear grade as initial report, which had five groups from 1 to 3. All potential correlations were further tested in non-parametric Kruskal-Wallis one-way ANOVA or Mann-Whitney U test (Table 16a, 16b, 16c), suggesting multiple potential associations in HOTAIR and H19 were still valid, but most associations determined by Pearson test in KCNQ1OT1 were incongruent with current methodology. (In Table 16a, 16b, 16c, KW represents Kruskal-Wallis test that was used for groups $\geq$ 3, MW indicates Mann-Whitney U test that was used for groups of 2)

Table 15. Pearson correlation matrix between lncRNA and clinicopathological markers in DCIS

	DCIS nuclear grade	ER status	PR Status	Hormone status	Her2 Status	Ki67 status	P53 Status
HOTAIR score	/	/	/	/	0.66(0.005)	0.50(0.036)	/
HOTAIR pattern	/	/	/	/	0.54(0.032)	/	/
HOTAIR software	/	-0.83(<0.001)	-0.49(0.014)	-0.83(<0.001)	/	0.58(0.011)	/
H19 score	/	/	/	/	/	0.49(0.045)	/
H19 pattern	/	/	/	/	/	0.53(0.03)	/
KCNQ1OT1 score	0.44(0.026)	/	/	/	/	0.51(0.029)	/
KCNQ1OT1 pattern	0.43(0.033)	/	/	/	/	0.51(0.029)	/
KCNQ1OT1 software	/	0.56(0.004)	/	0.56(0.004)	/	0.48(0.044)	0.54(0.014)
MEG3 score	/	/	/	/	/	/	/
MEG3 pattern	/	/	/	/	/	/	/

Data are given as Pearson correlation coefficient (p value) of each cell. "/" indicates there is no significant association between two variables

Table 16. Non-parametric Kruskal-Wallis test/ Mann-Whitney U test of lncRNAs in DCIS

a. Non-parametric Kruskal-Wallis test/ Mann-Whitney U test of HOTAIR

	sample size(n)	KW/MW statistics	Probabilities
Score/Her2	16	1.5	0.013
Score/Ki67	18	15.0	0.044
Pattern/Her2	16	4.5	0.03
Software/ER	25	46.0	0.021
Software/PR	25	47.0	0.711
Software/Hormone	25	46.0	0.021
Software/Ki67	18	7.0	0.007

b. Non-parametric Kruskal-Wallis test/ Mann-Whitney U test for H19

	Sample Size(n)	KW/MW statistics	Probabilities
Score/Ki67	17	14.0	0.045
Pattern/Ki67	17	14.5	0.047

c. Non-parametric Kruskal-Wallis test/ Mann-Whitney U test for KCNQ1OT1

	Sample size (n)	KW/MW statistics	Probabilities
Score/DCIS grade	25	7.0	0.072
Score/Ki67	18	15.5	0.043
Pattern/DCIS grade	25	7.4	0.06
Pattern/Ki67	18	18.0	0.052
Software/ER	25	39.0	0.109
Software/Hormone	25	39.0	0.109
Software/Ki67	18	17.0	0.075
Software/p53	20	11.0	0.38

Besides non-parametric Kruskal-Wallis one-way ANOVA or Mann-Whitney U test that shows some associations between lncRNA expression level and DCIS clinical markers according to statistical difference in mean, we also applied Fisher's exact test to demonstrate associations between lncRNA expression and clinical markers in terms of different frequency. Here, we performed all tests on dichotomized number for both lncRNA expression level and clinical markers in order to provide us with more significant correlations and more understandable interpretations based on our sample size. All the fisher's exact tests were done by 2x2 contingency table. (Table 17)

Table 17. Association study between dichotomous lncRNA level and clinical markers in DCIS by Fisher's exact test

Clinical item	lncRNA	Sample size	Fisher's exact test value	Odds Ratio
DCIS grade	HOTAIR	25	0.294	∞
	H19	25	0.661	0.476
	KCNQ1OT1	25	0.294	∞
	MEG3	23	0.657	0.514
ER	HOTAIR	25	1.00	0
	H19	25	0.507	∞
	KCNQ1OT1	25	1.00	0
	MEG3	23	0.142	0
PR	HOTAIR	25	0.527	2.00
	H19	25	0.140	∞
	KCNQ1OT1	25	1.00	0
	MEG3	23	1.00	0.583
Her2	HOTAIR	16	1.00	∞
	H19	15	1.00	1.50
	KCNQ1OT1	16	1.00	∞
	MEG3	16	1.00	2.333
Ki67	HOTAIR	18	0.515	∞
	H19	17	0.280	4.50
	KCNQ1OT1	18	0.515	∞
	MEG3	18	0.620	2.80
P53	HOTAIR	20	1.00	∞
	H19	19	1.00	2.40
	KCNQ1OT1	20	1.00	∞
	MEG3	20	0.189	∞

- Association between lncRNA expressions and invasive clinical markers

Subsequently, we applied same working procedures to determine the hypothesis that lncRNA expression may associate with invasive clinical markers. Pearson correlation test was firstly used as screen tool (Table 18) and non-parametric Kruskal-Wallis one-way ANOVA or Mann-Whitney U test were applied to validate candidate associations (Table 19a, 19b, 19c). In this test, Nottingham score remained from 4 to 9 and tumor size were trichotomized. (In table 10a, 10b, 10c, KW represents Kruskal-Wallis test that was used for groups $\geq$ 3, MW indicates Mann-Whitney U test that was used for groups of 2)

Table 18. Pearson correlation matrix in invasive cancer

	Nottingham grade	Tumor size	LVI status	ER status	PR status	Hormone status	Triple negative	Her2 status	Ki67 Status	P53 status
HOTAIR score	0.35(0.049)	0.40(0.023)	/	/	/	/	/	/	0.4(0.04)	/
HOTAIR pattern	0.37(0.037)	0.40(0.023)	/	/	/	/	/	0.37(0.039)	0.37(0.041)	/
HOTAIR software	0.35(0.050)	0.37(0.036)	/	-0.53(0.002)	/	-0.53(0.002)	0.53(0.002)	0.07(0.001)	/	/
								/	/	/
H19 score	/	/	/	/	/	/	/	/	/	/
H19 pattern	/	/	/	/	/	/	/	/	/	/
								/	/	/
KCNQ1OT1 score	0.41(0.02)	/	/	/	/	/	/	/	/	/
KCNQ1OT1 pattern	0.51(0.003)	/	/	/	-0.36(0.042)	/	/	0.39(0.027)	/	/
KCNQ1OT1 software	0.54(0.008)	/	/	-0.81(<0.001)	-0.70(<0.001)	-0.81(<0.001)	/	/	/	/
								/	/	/
MEG3 score	/	/	/	/	/	/	/	/	/	/
MEG3 pattern	/	/	/	/	/	/	/	0.38(0.027)	/	/

Data are given as Pearson correlation coefficient (p value) of each cell. "/" indicates there is no significant association between two variables

Table 19. Non-parametric Kruskal-Wallis test/ Mann-Whitney U test of lncRNAs in invasive cancer

a. Non-parametric Kruskal-Wallis test/ Mann-Whitney U test of HOTAIR

	sample size(n)	KW/MW statistics	Probabilities
Score/Nottingham	32	6.9	0.231
Score/tumor size	32	5.1	0.077
Score/Ki67 status	31	58.5	0.019
Pattern/Nottingham	32	7.1	0.215
Pattern/tumor size	32	5.1	0.078
Pattern/Her2	32	30.0	0.032
Pattern/Ki67	31	67.5	0.039
Software/Nottingham	32	4.9	0.431
Software/tumor size	32	4.4	0.112
Software/ER	32	31.0	0.938
Software/hormone	32	60.0	0.019
Software/triple N	32	29.0	0.938
Software/Her2	32	13.0	0.005

b. non-parametric Kruskal-Wallis test/ Mann-Whitney U test for KCNQ1OT1

	sample size(n)	KW/MW statistics	Probabilities
Score/Nottingham	32	6.4	0.267
Pattern/Nottingham	32	10.2	0.071
Pattern/PR	32	103.0	0.05
Pattern/Her2	32	20.5	0.032
Software/Nottingham	32	5.0	0.415
Software/ER	32	60.0	0.019
Software/PR	32	125.5	0.003
Software/hormone	32	60.0	0.019

c. non-parametric Kruskal-Wallis test/ Mann-Whitney U test for MEG3

	sample size (n)	KW/MW statistics	P
Pattern/Her2	34	30.5	0.026

Similarly, we test whether lncRNAs are associated with clinical makers regarding frequency by using Fisher's exact test. We also performed all tests on dichotomized number for both lncRNA expression level and clinical markers except invasive histological grade, which was analyzed non-parametric spearman rank correlation test. All the fisher's exact tests were done by Pearson Chi-Square 2x2 contingency table (Table 20), and spearman rank study was tested in 2x3 table (Table 21).

Table 20. Association study between dichotomized lncRNA level and clinical markers in invasive cancer by Fisher's exact test

clinical items	lncRNA	sample size	Fisher's exact test value	Odd ratio
Tumor size	HOTAIR	32	0.069	∞
	H19	35	0.297	∞
	KCNQ1OT1	32	0.550	∞
	MEG3	34	0.683	2.00
Lymph node	HOTAIR	28	0.128	∞
	H19	31	0.147	∞
	KCNQ1OT1	28	0.295	∞
	MEG3	29	0.107	6.55
LVI	HOTAIR	32	0.025	∞
	H19	35	0.630	2.667
	KCNQ1OT1	32	0.272	∞
	MEG3	34	0.251	3.385
ER	HOTAIR	32	1.000	0
	H19	35	0.477	2.250
	KCNQ1OT1	32	1.000	0
	MEG3	34	1.000	0
PR	HOTAIR	32	1.000	.667
	H19	35	0.568	2.667
	KCNQ1OT1	32	1.000	0
	MEG3	34	0.644	0.422
Triple negative	HOTAIR	32	1.000	∞
	H19	35	0.477	0.444
	KCNQ1OT1	32	1.000	∞
	MEG3	34	1.000	∞
Her2	HOTAIR	32	0.560	∞
	H19	35	1.000	∞
	KCNQ1OT1	32	1.000	∞
	MEG3	34	0.291	∞
Ki67	HOTAIR	31	0.201	5.077
	H19	34	1.000	2.000
	KCNQ1OT1	31	0.601	0.500
	MEG3	32	1.000	1.286
P53	HOTAIR	31	0.406	0.261
	H19	32	1.000	∞
	KCNQ1OT1	30	1.000	∞
	MEG3	32	1.000	∞

Fisher's exact statistics are two-tailed

Table 21. Association study between dichotomized lncRNA level and invasive histological grade in invasive cancer by Spearman rank correlation test

Clinical item	lncRNA	sample size	Spearman Rho	Probabilities
Histological grade	HOTAIR	32	0.375	0.026
	H19	35	0.090	0.583
	KCNQ1OT1	32	0.071	0.719
	MEG3	34	0.168	0.271

Probabilities are two-tailed

- Logistic regression analysis for lncRNA to predict cancer aggressiveness

As we already noticed some associations between lncRNA expression level and clinical markers by either Kruskal-Wallis one-way ANOVA test or Fisher's exact test, we subsequently tested the hypothesis that lncRNAs combined clinical items status can better predict DCIS or invasive cancer grade. First of all, we used Fisher's exact test on dichotomized dataset to look for associations between DCIS nuclear grade and all other markers (Table 22a), and we also applied Spearman rank correlation study to check any correlation between trichotomized invasive histological grade and the other markers (Table 22b). From our analysis, we did not find any significant correlation between DCIS nuclear grade and lncRNAs, but ER ($p=0.012$). However, since there was only one variable suggesting association, we did not perform multivariate test. In invasive cancer subjects, multiple variables have been found correlation with histological grade, including HOTAIR ($p=0.026$), ER ($p<0.001$, negative relationship), triple negative ($p<0.001$) and Ki67 ($p<0.001$). Then we provided multivariate logistic regression to test the predictability of cancer grade by a set of predictor variables.

Table 22. Correlation test of cancer grade and all markers

a. Correlation test for DCIS nuclear grade by Fisher's exact test

Variable 1	Variable 2	Sample size	Fisher's exact test value
DCIS nuclear grade	HOTAIR	25	0.294
	H19	25	0.661
	KCNQ1OT1	25	0.294
	MEG3	23	0.657
	ER	30	0.012
	PR	30	0.156
	Her2	16	0.518
	Ki67	18	0.107
	P53	20	0.521

b. Correlation test for invasive histological grade by Spearman rank correlation test

Variable 1	Variable 2	Sample size	Probabilities
Invasive histological grade	HOTAIR	32	0.026
	H19	35	0.582
	KCNQ1OT1	32	0.719
	MEG3	34	0.271
	Tumor size	44	0.234
	Lymph node	38	0.517
	LVI	44	0.739
	ER	44	<0.001(negative)
	PR	44	0.082
	Triple negative	44	<0.001
	Her2	43	0.155
	Ki67	35	<0.001
	P53	33	0.138

Binary logit analysis was chosen because we dichotomized histological grades as our limited sample size. Firstly, we categorized invasive histological grade into low grade (n=8) coded with 0 and non-low grade (n=19) coded with 1. We took the first step to consider bivariate logistical regression, using dichotomized histological grade as criterion variable and lncRNA expression level as dichotomous predictor variable. Our regression model will be predicting the logit, that is, the natural log of the odds of having one or the other histological grade (model is shown below),

where Y was the probability of the event which is coded with 1, 1-Y was the predicted probability of the event coded with 0, and X was the predictor variable, lncRNA dichotomized score.

$$\ln ODDS = \ln \frac{Y}{1-Y} = a + bX$$

Our data suggests only HOTAIR (p=0.011) was a positive predictor among all lncRNAs and odds ratio was 14.167. Our regression equation is $\ln ODDS = -0.916 + 2.651X$.

Then, we applied same binary logit analysis on all dichotomized clinical items. However, it turned out that no clinical items was a good predictor. Lastly, we performed a multivariate logistical regression to take both dichotomous lncRNA expression and clinical items into account. We set up another model to predict odds as shown below, where Y was the probability of the event which is coded with 1, 1-Y was the predicted probability of the event coded with 0, and Xn was a single predictor variable, lncRNA or clinical item

$$\ln ODDS = \ln \frac{Y}{1-Y} = a + bX1 + cX2 + \cdots nXn$$

Only lymphovascular invasion (LVI) status (p=0.08) and HOTAIR (p=0.043) showed a joint ability to predict cancer grade with an odds ratio 16.5. Our equation is $\ln ODDS = -0.405 - 2.175X1 + 2.803X2$, where X1 represents lymphovascular invasion status and X2 represents HOTAIR.

Similarly, we categorized invasive histological grade into high grade (n=6) coded with1 and non-high grade (n=21) coded with 0. However, we could only find Ki67 status is a good predictor (p=0.013) from binary logit analysis.

- Correlation between protein markers and lncRNA expression

Functionally, one mechanism of lncRNAs influencing gene expression is to recruit epigenetic protein factors to regulate chromatin states of target gene in cis or trans. Polycomb repressive complex 2, (PRC2), as a chromatin-modifying complex, interacts with a large group of lncRNA including HOTAIR, H19 and KCNQ1OT1. From our study, we found EZH2 significantly expressed higher in cancer than in NA breast tissue ($p<0.001$) Here, we tested the hypothesis that the expression of Enhancer of zeste homolog 2 (EZH2), as a key protein member of PRC2 complex as gene silencer, increases as lncRNAs express, by Pearson correlation study (Table 23). Our data suggests only one correlation between KCNQ1OT1 expression and EZH2 expression in invasive cancer ($p=0.006$).

Table 23. Pearson correlation study between protein markers and lncRNA expression

Tissue	Protein marker	lncRNA	Sample size	Pearson correlation	Probability
NA	EZH2	HOTAIR	23	0.034	0.877
		H19	26	0.272	0.179
		KCNQ1OT1	25	-0.007	0.972
DCIS	EZH2	HOTAIR	21	0.294	0.197
		H19	21	0.371	0.098
		KCNQ1OT1	21	0.065	0.781
IC	EZH2	HOTAIR	29	-0.085	0.662
		H19	31	0.016	0.993
		KCNQ1OT1	29	0.495	0.006

Discussion

RNAscope® *in situ* hybridization assay

Regarding the work platform, RNAscope *in situ* hybridization technology has now been employed elsewhere to detect lncRNA (Bao, Wu et al. 2013 [male mouse germline]; Prensner, Iyer et al. 2013 [prostate cancer]; Warrick, Tomlins et al. 2014 [prostate cancer]), providing valuable information on clinical relevant information upon cellular and tissue context that is unable to visualize by routine microarray and quantitative PCR. The present study demonstrates the quality and specificity of RNAscope technology to detect lncRNA on FFPE tissues.

Sample size

In current study, tissue microarray was manufactured to provide a standard and high throughput assays on same tissue source for both RNA candidate markers and immunohistochemical markers with comparatively low cost. Although routine H&E slides were made from selected FFPE blocks to help pathologists to mark concurrent developed DCIS, invasive cancer and adjacent normal tissue regions that would be included in TMAs, mixed tissue areas were accidentally found on our TMA slides after we rescreened TMA slides with H&E slides for confirmation. In that case, we only reviewed tissue piece that was intended to be there. As breast tissue contains most abundant adipocytes, 8 of 44 (18%) normal adjacent spots turned out to only have adipose tissue so that they were excluded for further analysis. We had the least number of DCIS spots (n=34) across our TMAs because of two reasons: 1) several patients only had NA and invasive cancer tissues 2) several spots contained very little DCIS part that was easily removed during manufacturing process. Comparing with NA and DCIS tissue, we kept the most number (n=44) of invasive cancer tissue source. However, we ended up having fewer tissue spots on our TMAs to score and analyze because some more spots were washed away during multiple pretreatment and amplification steps, even with more cautions, resulting in a limited sample size.

HOTAIR

Our results suggest lncRNAs have different staining patterns, mostly in concord with previous publications. The HOTAIR RNA probe stained predominately as single or multiple dots that was widely present in all three tissue sources (Chisholm, Wan et al. 2012 [digoxigenin-labeled riboprobe ISH]). However, most signals were found in nucleus with some scattered dots in cytoplasm, contrary to what Chisholm, et al. found in their paper, which suggests HOTAIR was more prevalent in cytoplasm (Chisholm, Wan et al. 2012). As their primary focus on metastatic breast carcinoma while we were looking at primary tumors, it is possible that HOTAIR escapes from nucleus to cytoplasm as primary tumor metastases. Our data also provides us some evidence that HOTAIR expresses at different levels across different types of tissue. Generally HOTAIR signal enhances as cancer aggressiveness increases. Friedman two-way analysis of variance test shows HOTAIR has significantly higher expression in cancer cells than normal adjacent epithelia cells ($p<0.001$). Even within cancer spots, invasive cancer is more aggressive and has more intense HOTAIR score than that of DCIS ($p=0.021$).

KCNQ1OT1

Prior studies have used fluorescent *in situ* hybridization (FISH) to visualize KCNQ1OT1 signal in nucleus of both human and mouse cells (Mohammad, Pandey et al. 2008; Korostowski, Sedlak et al. 2012). This is the first study to localize KCNQ1OT1 as punctuate dots in cell nucleus on FFPE tissue samples by CISH. Comparing with HOTAIR, KCNQ1OT1 is not that popular in terms of staining grade. Across our tissue microarray, we have 5 negative cases without any stain of KCNQ1OT1, most of which (4/5) are normal adjacent tissue spots, suggesting KCNQ1OT1 may express more frequently in cancer spots as a candidate oncogene. Also, there is an increased KCNQ1OT1 expression in regard to KCNQ1OT score in cancers than in NA ($p=0.013$). However, we did not notice any statistical difference of KCNQ1OT1 expression between non-invasive DCIS tissue and more malignant invasive cancer spots ($p=0.058$), revealing some possibilities

that KCNQ1OT1 is triggered by some pre-cancer events and then keep expressing as cancer develops.

H19

As an important oncogene for tumor growth, H19 was first detected in hepatocellular carcinoma by both radioactive and non-radioactive riboprobe *in situ* hybridization (Ariel, Miao et al. 1998). The present study provides some insights to localize H19 in breast carcinoma by RNAscope CISH. H19 was stained more diffusely and intense in cancer, compared with normal adjacet spots that usually negative or focal expression ($p<0.001$), strikingly, H19 also expressed significantly higher in more malignant invasive cancer than in DCIS ($p=0.008$). In this study, we also found that H19 appeared primarily in stromal cells of breast tissue especially at the boundary of breast cancer ducts as also reported by Ariel and his colleagues in 1998. However, other studies have also found H19 stained in epithelia ovarian cancer cells and bladder cancer cells (Mizrahi, Czerniak et al. 2009; Amit and Hochberg 2010). Together with current research focus on exsomes guided tumorigenesis and the fact lncRNAs has been found in exosomes by deep sequencing (Huang, Yuan et al. 2013), it may give rise to additional research to investigate H19 and tumor microenvironment.

MEG3

Similar to H19, our image depicts that MEG3 stained in stromal cell nucleus around breast ducts, which is the first study to visualize MEG3 RNA molecule on breast tissue. Our results compared MEG expression in each tissue types and found there was no significance between tissue types ($p=0.108$), while previous study showed MEG3 was lost in human breast cancer cell line MCF-7 and thus proposed MEG3 might be a novel tumor suppressor (Zhang, Zhou et al. 2003). This contradiction may be because of much more complexity and chaos in real human tissue than pure cell line, which would turn out to have different conclusions. Other than staining stromal cells, MEG3 has also been found previously to localize in most normal pituitary cells (Gejman, Batista

et al. 2008) and in cytoplasm of non-neoplastic liver cells(Braconi, Kogure et al. 2011), indicating MEG3 might be involved in a spectrum of cell activity with different functionalities.

MALAT1

As a key regulator in lung cancer, MALAT1 was implicated to play an oncogenic role and upregulation of MALAT1 was also observed in several other cancers, including breast cancer (Lin, Maeda et al. 2007). On the contrary, our results exhibited a universal strong nuclear stain of MALAT1 for every tissue spots of our tissue microarray slides. However, the reason that results in contradictory results is largely unknown, leading us to investigate both expressional and functional perspectives of MALAT1 in breast cancer in more details.

Zfas1

One Snord-host long noncoding RNA Zfas1, has been suggested detected by chromogenic in situ hybridization in mice pregnant mammary gland epithelia cells and the paper also evidenced that Zfas1 might be a breast cancer tumor suppressor by showing relatively lower expression level of Zfas1 in human breast cancer than in normal by quantitative PCR (Askarian-Amiri, Crawford et al. 2011). Current study on Zfas1 depicted a similar Zfas1 staining pattern as previous study showed, both of which were stained as low intensity dots in epithelia cell nucleus. Nonetheless, in order to assign them scores, even experienced pathologists had trouble to tell the minor difference between subjects so we did not review Zfas1 as carefully as other lncRNAs in order to prevent scoring bias. In situ hybridization assay, together with quantitative PCR or advanced quantitative image software, would be the best choice to further analyze Zfas1.

LncRNA and EZH2

EZH2, a key subunit of polycomb repressive complex 2 that is usually interacts with lncRNA, has been extensively associated with lncRNA, especially HOTAIR, H19 and KCNQ1OT1(Gupta, Shah et al. 2010; Luo, Li et al. 2013; Zhang, Zeitz et al. 2014). In this study, using immunohistochemistry to assess EZH2 expression level, we found an increase expression of

EZH2 in cancer than NA tissue (p=0.003), however, there was no difference of EZH2 between DCIS tissue and invasive cancer tissue (p>0.05). We did not find any correlation between EZH2 and lncRNAs in either NA or DCIS tissue, however, we noticed only one strong correlation (p=0.006) between EZH2 and KCNQ1OT1 in invasive tissue. Although whether H19 or KCNQ1OT1 induce EZH2 is largely unknown, Chisholm's paper showed HOTAIR and EZH2 are co-expressed in breast cancer. Because of our limited sample size, we might not be able to confirm some correlations, but we still provided some evidence for co-expression of KCNQ1OT1 and EZH2 in invasive breast cancer.

Clinicopathological correlations

In the present study, by using Pearson correlation test and non-parametric ANOVA test, we found HOTAIR positively trends trend with Her2 (p=0.013) and Ki67 (p=0.044), H19 positively correlates with Ki67 (p=0.045) in DCIS. Interestingly, only one true correlation between KCNQ1OT1 and clinicopathological factors, which was Ki67, was identified (p=0.043) although our screening Pearson test proposed multiple associations, indicating Pearson test is not robust when one variable is categorical. With invasive clinicopathological factors, only HOTAIR significantly associates with Ki67 (p=0.019). However, by performing same test by using three-tiered score of lncRNAs as a filter, we could find more robust associations, for example, in DCIS, HOTAIR only significantly correlated with Her2 (p=0. 03) and H19 positively associates with Ki67 (p=0.047); in invasive cancer, HOTAIR highly associates with Her2 (p=0.032) and Ki67 (p=0.039), KCNQ1OT1 positively correlates with Her2 (p=0.032) and negatively marginal correlates with PR positivity (p=0.05) , providing more important insights for clinicopathological correlations. By using univariate analysis on our dichotomous data, we were able to find a significant positive association between HOTAIR and histological grade (p=0.011) with odds ratio of 14.167. Our regression equation is $\ln ODDS = -0.916 + 2.651\mathrm{X}$, indicating the odds is 5.668 if this subject has high expressed HOTAIR and odds is 0.4 if this case has low HOTAIR

expression level. We subsequently converted odds to probabilities, resulting in a probability of 0.85 with predictor variable coded with 1, but a probability of 0.29 with predictor variable coded with 0, indicating our model can predict 85% HOTAIR high expression (coded with 1) will have non-low grade invasive breast cancer (coded with 1), while this model can predict only 29% cases with higher HOTAIR expression will have non-low grade cancer, also suggesting HOTAIR is an informative independent prognostic factor for invasive cancer aggressiveness. Previous studies have already reported HOTAIR highly associated with primary tumor grade in colorectal cancer, epithelia ovarian cancer and gastric cancer, and we confirm it in primary breast tumor (Kogo, Shimamura et al. 2011; Emadi-Andani, Nikpour et al. 2014; Qiu, Lin et al. 2014). Subsequent multivariate analysis suggested the combination of lymphovascular invasion and HOTAIR expression was able to result in better prediction for histological grade with an odds ratio of 16.5. Our equation is $\ln ODDS = -0.405 - 2.175\text{X1} + 2.803\text{X2}$, where X1 represents lymphovascular invasion status and X2 represents HOTAIR. As we can notice that HOTAIR positively relates with odds while LVI status had a negative relationship with odds in our model. The odds is 11 when the case presents LVI negative while has highly expressed HOTAIR, resulting in a probability of 0.92, which indicates our model predicts that 92% of cases with LVI negative and high HOTAIR status will have a more aggressive cancer. However, the odds is 0.08 for cases that have both positive LVI and low HOTAIR expression. The probability is 0.07, indicating that our model predicts only 7% of patients who have been found with low HOTAIR and positive LVI, will eventually have a cancer beyond low histological grade. Although currently standard H&E staining is an easy and informative approach for grading cancer, HOTAIR also may be a useful adjunct test for the evaluation of equivocal grades. We also found a few cases that deviated from the trend, i.e., low HOTAIR expression in high-grade tumors: this type of dysregulation in invasive cancer might represent a specific subgroup of invasive cancers; further HOTAIR studies are required. As all the comparisons in this study were tested between

DCIS, invasive tissue and normal adjacent tissue, further studies are also required comparing lncRNA expression in tumor tissues to lncRNA expression in normal tissues from healthy women.

RNAscope SpotStudio™ software

Manual scoring is time consuming and prone to subjectivity and poor reproducibility, however, RNAscope SpotStudio™ software is an intuitive automated system that is designed specifically for RNAscope technology for quantification. Generally, it quantifies signal intensity and area in single molecule sensitivity. This present study also utilized this software package for HOTAIR and KCNQ1OT1 as an adjunct quantification because their signal are mostly dots rather than whole nucleus stain of MEG3 or "streak" stain around ducts boundary of H19. Significant correlations between eye-scored lncRNA expression level and software-scored lncRNA expression level were both found in HOTAIR and KCNQ1OT1 across different tissue types, with an exception of HOTAIR in NA tissue. However, we found some inconsistent results when we used software generated results for tissue comparison study and clinicopathological correlations study, largely because there were a few extreme observations identified by SpotStudio™ that increase too much variability. Although image analysis tool prone to provide standard quantification, it also requires operator to manually select regions of interest, optimize nucleus diameter and hematoxylin stain parameter that also might result in subjectivity. Moreover, strong hematoxylin stain is very likely to perplex this image analysis system in recognition of true signals. And more than that, too much variation has been observed in our limited samples so that deviation increases remarkably in our statistical analysis, resulting in inconsistent consequences.

Conclusion:

The primary goal of this research was to detect lncRNA molecules in FFPE specimens and test for associations between lncRNA expression and clinicopathological factors. From our results, we conclude that the RNAscope® CISH assay is a highly effective tool for visualizing lncRNA expression. With this novel technology, we can not only quantify RNA expression to single copy sensitivity, but also visualize RNA molecules and interpret their expression within the cellular context, which is not possible by qPCR. Regarding the expression level of lncRNA across tissues, we can conclude that oncogenic activity of HOTAIR, H19 and KCNQ1OT1 is demonstrated in breast cancer by showing that all of them have significantly higher expression levels in cancer tissue than normal adjacent. Also, both HOTAIR and H19 express significantly higher in invasive cancer than non-invasive ductal carcinoma *in situ*. Moreover, some associations between lncRNA and clinical items are suggested by the current investigation. HOTAIR was found to associate with Her2 status in both DCIS and invasive cancer. HOTAIR was also concluded to be a critical independent positive predictor for invasive cancer histological grade. The other two important lncRNAs, KCNQ1OT1 and H19, were also found to correlate with some important clinical factors. Although this study did not directly address the causal mechanisms by which lncRNA aberrant expression relates to clinical status, it can be hypothesized on the basis of other functional studies that lncRNAs participate in a spectrum of cell activities and that lncRNA dysregulation promotes or attenuates tumor development. We also concluded that the use of image analysis software facilitates standardized RNAscope data interpretation.

Project Summary and Future Direction

Both projects examined novel approaches for assessing biomarkers of breast carcinoma. The perspectives are different: in the first study, we focused on tumor cell DNA content, which has

previously been shown to have an association with breast cancer clinical and pathological characteristics. Here, we developed an alternative technique for DNA content analysis by flow cytometric analysis and show the utility of our novel tissue core sampling approach for measuring DNA content in FFPE tissue specimens. Strikingly, we also found this alternative tissue sampling method is a good approach for the intratumoral heterogeneity assay of solid breast tumors. Future investigations would apply multiparametric approaches that combined DNA content, protein markers labeling, and chromosomal changes to better identify distinct groups in breast carcinoma. Moreover, future studies would be also able to utilize the proposed technique and single cell sequencing to obtain a better understanding of the composition of tumors, and the significance for prognosis and even personalized therapeutics.

In the second study, lncRNAs that may be biomarkers for both invasive breast cancer and DCIS were examined by using RNAscope® CISH. We visualized and quantified lncRNA stains by bright field microscopy. Several associations between lncRNAs and clinicopathlogical factors have been found, suggesting potential utility of lncRNA CISH in breast carcinoma diagnostics. Future studies require an increased sample size. A small sample size might cause too much deviation that result in spurious correlations. With the help of the RNAscope CISH assay in combination with TMAs, more and more lncRNAs probes can be designed and tested on either preserved tissue of fresh tissue. Automated RNAscope CISH is also now possible supporting high-throughput standardized research and clinical assays that can be combined with the advanced semi-quantitative software supporting data analyses uncompromised by human subjectivity. As we point out some associations already, functional studies would be another direction to promote research of this discipline.

Comprehensive Bibliography

Aguilo, F., M. M. Zhou, et al. (2011). "Long noncoding RNA, polycomb, and the ghosts haunting INK4b-ARF-INK4a expression." Cancer Res **71**(16): 5365-5369.

Allred, D. C. (2010). "Ductal carcinoma in situ: terminology, classification, and natural history." J Natl Cancer Inst Monogr **2010**(41): 134-138.

Ariel, I., H. Q. Miao, et al. (1998). "Imprinted H19 oncofetal RNA is a candidate tumour marker for hepatocellular carcinoma." Mol Pathol **51**(1): 21-25.

Arnerlov, C., S. O. Emdin, et al. (2001). "Intratumoral variations in DNA ploidy and s-phase fraction in human breast cancer." Anal Cell Pathol **23**(1): 21-28.

Amit, D. and A. Hochberg (2010). "Development of targeted therapy for bladder cancer mediated by a double promoter plasmid expressing diphtheria toxin under the control of H19 and IGF2-P4 regulatory sequences." J Transl Med **8**: 134.

Artandi, S. E. and R. A. DePinho (2000). "A critical role for telomeres in suppressing and facilitating carcinogenesis." Curr Opin Genet Dev **10**(1): 39-46.

Askarian-Amiri, M. E., J. Crawford, et al. (2011). "SNORD-host RNA Zfas1 is a regulator of mammary development and a potential marker for breast cancer." RNA **17**(5): 878-891.

Atzori, F., T. A. Traina, et al. (2009). "Targeting insulin-like growth factor type 1 receptor in cancer therapy." Target Oncol **4**(4): 255-266.

Bao, J., J. Wu, et al. (2013). "Expression profiling reveals developmentally regulated lncRNA repertoire in the mouse male germline." Biol Reprod **89**(5): 107.

Bast, R. C., Jr., P. Ravdin, et al. (2001). "2000 update of recommendations for the use of tumor markers in breast and colorectal cancer: clinical practice guidelines of the American Society of Clinical Oncology." J Clin Oncol **19**(6): 1865-1878.

Baxter, N. N., B. A. Virnig, et al. (2004). "Trends in the treatment of ductal carcinoma in situ of the breast." J Natl Cancer Inst **96**(6): 443-448.

Bekpen, C., T. Marques-Bonet, et al. (2009). "Death and resurrection of the human IRGM gene." PLoS Genet **5**(3): e1000403.

Bergamaschi, A., Y. H. Kim, et al. (2006). "Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer." Genes Chromosomes Cancer **45**(11): 1033-1040.

Bergers, E., P. J. van Diest, et al. (1996). "Tumour heterogeneity of DNA cell cycle variables in breast cancer measured by flow cytometry." J Clin Pathol **49**(11): 931-937.

Berteaux, N., S. Lottin, et al. (2005). "H19 mRNA-like noncoding RNA promotes breast cancer cell proliferation through positive control by E2F1." J Biol Chem **280**(33): 29625-29636.

Boardman, L. A., S. N. Thibodeau, et al. (1998). "Increased risk for cancer in patients with the Peutz-Jeghers syndrome." Ann Intern Med **128**(11): 896-899.

Bonneterre, J., J. P. Peyrat, et al. (1990). "Prognostic significance of insulin-like growth factor 1 receptors in human breast cancer." Cancer Res **50**(21): 6931-6935.

Bonsing, B. A., H. Beerman, et al. (1993). "High levels of DNA index heterogeneity in advanced breast carcinomas. Evidence for DNA ploidy differences between lymphatic and hematogenous metastases." cancer **71**(2): 382-391.

Bosch, A., P. Eroles, et al. (2010). "Triple-negative breast cancer: molecular features, pathogenesis, treatment and current lines of research." Cancer Treat Rev **36**(3): 206-215.

Braconi, C., T. Kogure, et al. (2011). "microRNA-29 can regulate expression of the long non-coding RNA gene MEG3 in hepatocellular cancer." Oncogene **30**(47): 4750-4756.

Bray, F., P. McCarron, et al. (2004). "The changing global patterns of female breast cancer incidence and mortality." Breast Cancer Res **6**(6): 229-239.

Bremmer, J. F., R. H. Brakenhoff, et al. (2011). "Prognostic value of DNA ploidy status in patients with oral leukoplakia." Oral Oncol **47**(10): 956-960.

Britten, R. J. and E. H. Davidson (1971). "Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty." Q Rev Biol **46**(2): 111-138.

Buerger, H., F. Otterbach, et al. (1999). "Comparative genomic hybridization of ductal carcinoma in situ of the breast-evidence of multiple genetic pathways." J Pathol **187**(4): 396-402.
Buerger, H., F. Otterbach, et al. (1999). "Different genetic pathways in the evolution of invasive breast cancer are associated with distinct morphological subtypes." J Pathol **189**(4): 521-526.
Bussemakers, M. J., A. van Bokhoven, et al. (1999). "DD3: a new prostate-specific gene, highly overexpressed in prostate cancer." Cancer Res **59**(23): 5975-5979.
Cabili, M. N., C. Trapnell, et al. (2011). "Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses." Genes Dev **25**(18): 1915-1927.
Campbell, P. J., E. D. Pleasance, et al. (2008). "Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing." Proc Natl Acad Sci U S A **105**(35): 13081-13086.
Campbell, P. J., S. Yachida, et al. (2010). "The patterns and dynamics of genomic instability in metastatic pancreatic cancer." Nature **467**(7319): 1109-1113.
Carninci, P., T. Kasukawa, et al. (2005). "The transcriptional landscape of the mammalian genome." Science **309**(5740): 1559-1563.
Carrieri, C., L. Cimatti, et al. (2012). "Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat." Nature **491**(7424): 454-457.
Castellani, R., D. W. Visscher, et al. (1994). "Interaction of transforming growth factor-alpha and epidermal growth factor receptor in breast carcinoma. An immunohistologic study." cancer **73**(2): 344-349.
Chen, F. J., M. Sun, et al. (2013). "Upregulation of the long non-coding RNA HOTAIR promotes esophageal squamous cell carcinoma metastasis and poor prognosis." Mol Carcinog **52**(11): 908-915.
Chen, J. J., D. Silver, et al. (1999). "BRCA1, BRCA2, and Rad51 operate in a common DNA damage response pathway." Cancer Res **59**(7 Suppl): 1752s-1756s.
Chen, L. L. and G. G. Carmichael (2009). "Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA." Mol Cell **35**(4): 467-478.
Chen, W., W. Bocker, et al. (1997). "Expression of neural BC200 RNA in human tumours." J Pathol **183**(3): 345-351.
Chisholm, K. M., Y. Wan, et al. (2012). "Detection of long non-coding RNA in archival tissue: correlation with polycomb protein expression in primary and metastatic breast carcinoma." PLoS One **7**(10): e47998.
Chung, S., H. Nakagawa, et al. (2011). "Association of a novel long non-coding RNA in 8q24 with prostate cancer susceptibility." Cancer Sci **102**(1): 245-252.
Clemson, C. M., J. N. Hutchinson, et al. (2009). "An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles." Mol Cell **33**(6): 717-726.
Comings, D. E. (1972). "The structure and function of chromatin." Adv Hum Genet **3**: 237-431.
Corver, W. E. and N. T. ter Haar (2011). "High-resolution multiparameter DNA flow cytometry for the detection and sorting of tumor and stromal subpopulations from paraffin-embedded tissues." Curr Protoc Cytom **Chapter 7**: Unit 7 37.
Corver, W. E., N. T. Ter Haar, et al. (2011). "Cervical carcinoma-associated fibroblasts are DNA diploid and do not show evidence for somatic genetic alterations." Cell Oncol (Dordr) **34**(6): 553-563.
Couch, F. J., M. L. DeShano, et al. (1997). "BRCA1 mutations in women attending clinics that evaluate the risk of breast cancer." N Engl J Med **336**(20): 1409-1415.
Creighton, C. J., X. Fu, et al. (2010). "Proteomic and transcriptomic profiling reveals a link between the PI3K pathway and lower estrogen-receptor (ER) levels and activity in ER+ breast cancer." Breast Cancer Res **12**(3): R40.
Danesi, D. T., M. Spano, et al. (1997). "Flow cytometric and immunohistochemical correlations in high incidence human solid tumors." Tumori **83**(3): 689-697.
Day, J. R., M. Jost, et al. (2011). "PCA3: from basic molecular science to the clinical lab." Cancer Lett **301**(1): 1-6.
Dayal, J. H., M. J. Sales, et al. (2013). "Multiparameter DNA content analysis identifies distinct groups in primary breast cancer." Br J Cancer **108**(4): 873-880.

Dhanasekaran, S. M., T. R. Barrette, et al. (2001). "Delineation of prognostic biomarkers in prostate cancer." Nature **412**(6849): 822-826.
Djebali, S., C. A. Davis, et al. (2012). "Landscape of transcription in human cells." Nature **489**(7414): 101-108.
Duffy, S. W., O. Agbaje, et al. (2005). "Overdiagnosis and overtreatment of breast cancer: estimates of overdiagnosis from two trials of mammographic screening for breast cancer." Breast Cancer Res **7**(6): 258-265.
Dumont, N. and C. L. Arteaga (2000). "Transforming growth factor-beta and breast cancer: Tumor promoting effects of transforming growth factor-beta." Breast Cancer Res **2**(2): 125-132.
Easton, D. F. (1999). "How many more breast cancer predisposition genes are there?" Breast Cancer Res **1**(1): 14-17.
Easton, D. F., S. A. Narod, et al. (1994). "The genetic epidemiology of BRCA1. Breast Cancer Linkage Consortium." Lancet **344**(8924): 761.
Eheman, C. R., K. M. Shaw, et al. (2009). "The changing incidence of in situ and invasive ductal and lobular breast carcinomas: United States, 1999-2004." Cancer Epidemiol Biomarkers Prev **18**(6): 1763-1769.
Elston, E. W. and I. O. Ellis (1993). "Method for grading breast cancer." J Clin Pathol **46**(2): 189-190.
Emadi-Andani, E., P. Nikpour, et al. (2014). "Association of HOTAIR expression in gastric carcinoma with invasion and distant metastasis." Adv Biomed Res **3**: 135.
Ernster, V. L. and J. Barclay (1997). "Increases in ductal carcinoma in situ (DCIS) of the breast in relation to mammography: a dilemma." J Natl Cancer Inst Monogr(22): 151-156.
Esteva, F. J., C. D. Cheli, et al. (2005). "Clinical utility of serum HER2/neu in monitoring and prediction of progression-free survival in metastatic breast cancer patients treated with trastuzumab-based therapies." Breast Cancer Res **7**(4): R436-443.
Eusebi, V., E. Feudale, et al. (1994). "Long-term follow-up of in situ carcinoma of the breast." Semin Diagn Pathol **11**(3): 223-235.
Eusebi, V., M. P. Foschini, et al. (1989). "Long-term follow-up of in situ carcinoma of the breast with special emphasis on clinging carcinoma." Semin Diagn Pathol **6**(2): 165-173.
Faghihi, M. A., F. Modarresi, et al. (2008). "Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase." Nat Med **14**(7): 723-730.
Fellenberg, J., L. Bernd, et al. (2007). "Prognostic significance of drug-regulated genes in high-grade osteosarcoma." Mod Pathol **20**(10): 1085-1094.
Fellig, Y., I. Ariel, et al. (2005). "H19 expression in hepatic metastases from a range of human carcinomas." J Clin Pathol **58**(10): 1064-1068.
Fridlyand, J., A. M. Snijders, et al. (2006). "Breast tumor copy number aberration phenotypes and genomic instability." BMC Cancer **6**: 96.
Galea, M. H., R. W. Blamey, et al. (1992). "The Nottingham Prognostic Index in primary breast cancer." Breast Cancer Res Treat **22**(3): 207-219.
Genestie, C., B. Zafrani, et al. (1998). "Comparison of the prognostic value of Scarff-Bloom-Richardson and Nottingham histological grades in a series of 825 cases of breast cancer: major importance of the mitotic count as a component of both grading systems." Anticancer Res **18**(1B): 571-576.
Gejman, R., D. L. Batista, et al. (2008). "Selective loss of MEG3 expression and intergenic differentially methylated region hypermethylation in the MEG3/DLK1 locus in human clinically nonfunctioning pituitary adenomas." J Clin Endocrinol Metab **93**(10): 4119-4125.
Giaretti, W., S. Monteghirfo, et al. (2013). "Chromosomal instability, DNA Index, dysplasia and sub-site in oral pre-malignancy as intermediate end-points of risk of cancer." Cancer Epidemiol Biomarkers Prev.
Gong, C. and L. E. Maquat (2011). "lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements." Nature **470**(7333): 284-288.
Gordon, D. J., B. Resio, et al. (2012). "Causes and consequences of aneuploidy in cancer." Nat Rev Genet **13**(3): 189-203.
Gowen, L. C., A. V. Avrutskaya, et al. (1998). "BRCA1 required for transcription-coupled repair of oxidative DNA damage." Science **281**(5379): 1009-1012.

Gray, J. W., C. Collins, et al. (1994). "Molecular cytogenetics of human breast cancer." Cold Spring Harb Symp Quant Biol **59**: 645-652.
Guil, S., M. Soler, et al. (2012). "Intronic RNAs mediate EZH2 regulation of epigenetic targets." Nat Struct Mol Biol **19**(7): 664-670.
Gupta, R. A., N. Shah, et al. (2010). "Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis." Nature **464**(7291): 1071-1076.
Hammond, M. E., P. L. Fitzgibbons, et al. (2000). "College of American Pathologists Conference XXXV: solid tumor prognostic factors-which, how and so what? Summary document and recommendations for implementation. Cancer Committee and Conference Participants." Arch Pathol Lab Med **124**(7): 958-965.
Han, Y., Y. Liu, et al. (2013). "Hsa-miR-125b suppresses bladder cancer development by down-regulating oncogene SIRT7 and oncogenic long non-coding RNA MALAT1." FEBS Lett **587**(23): 3875-3882.
Hannemann, J., A. Velds, et al. (2006). "Classification of ductal carcinoma in situ by gene expression profiling." Breast Cancer Res **8**(5): R61.
Hayes, D. F. (1996). "Serum (circulating) tumor markers for breast cancer." Recent Results Cancer Res **140**: 101-113.
He, X., W. Bao, et al. (2014). "The long non-coding RNA HOTAIR is upregulated in endometrial carcinoma and correlates with poor prognosis." Int J Mol Med **33**(2): 325-332.
He, Y., B. Vogelstein, et al. (2008). "The antisense transcriptomes of human cells." Science **322**(5909): 1855-1857.
Hedley, D. W., M. L. Friedlander, et al. (1983). "Method for analysis of cellular DNA content of paraffin-embedded pathological material using flow cytometry." J Histochem Cytochem **31**(11): 1333-1335.
Helleday, T., E. Petermann, et al. (2008). "DNA repair pathways as targets for cancer therapy." Nat Rev Cancer **8**(3): 193-204.
Hessels, D., J. M. Klein Gunnewiek, et al. (2003). "DD3(PCA3)-based molecular urine analysis for the diagnosis of prostate cancer." Eur Urol **44**(1): 8-15; discussion 15-16.
Hibi, K., H. Nakamura, et al. (1996). "Loss of H19 imprinting in esophageal cancer." Cancer Res **56**(3): 480-482.
Hilborne, L. H., L. Cheng, et al. (1986). "Evaluation of an antibody to human milk fat globule antigen in the detection of metastatic carcinoma in pleural, pericardial and peritoneal fluids." Acta Cytol **30**(3): 245-250.
Hillig, T., J. Thode, et al. (2012). "Assessing HER2 amplification by IHC, FISH, and real-time polymerase chain reaction analysis (real-time PCR) following LCM in formalin-fixed paraffin embedded tissue from 40 women with ovarian cancer." APMIS **120**(12): 1000-1007.
Hortobagyi, G. N. (2005). "Trastuzumab in the treatment of breast cancer." N Engl J Med **353**(16): 1734-1736.
Huang, X., T. Yuan, et al. (2013). "Characterization of human plasma-derived exosomal RNAs by deep sequencing." BMC Genomics **14**: 319.
Hung, T., Y. Wang, et al. (2011). "Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters." Nat Genet **43**(7): 621-629.
Iacoangeli, A., Y. Lin, et al. (2004). "BC200 RNA in invasive and preinvasive breast cancer." Carcinogenesis **25**(11): 2125-2133.
National Cancer Institute (2005). Paget's Disease of the Nipple: Questions and Answers.
Ji, P., S. Diederichs, et al. (2003). "MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer." Oncogene **22**(39): 8031-8041.
Jia, L. F., S. B. Wei, et al. (2013). "Expression, Regulation and Roles of MiR-26a and MEG3 in Tongue Squamous Cell Carcinoma." Int J Cancer.
John, B. and G. L. Miklos (1979). "Functional aspects of satellite DNA and heterochromatin." Int Rev Cytol **58**: 1-114.
Jordan, V. C. and S. Koerner (1975). "Tamoxifen (ICI 46,474) and the human carcinoma 8S oestrogen receptor." Eur J Cancer **11**(3): 205-206.

Kallioniemi, O. P. (1988). "Comparison of fresh and paraffin-embedded tissue as starting material for DNA flow cytometry and evaluation of intratumor heterogeneity." Cytometry **9**(2): 164-169.
Kapranov, P., J. Cheng, et al. (2007). "RNA maps reveal new RNA classes and a possible function for pervasive transcription." Science **316**(5830): 1484-1488.
Khalil, A. M., M. Guttman, et al. (2009). "Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression." Proc Natl Acad Sci U S A **106**(28): 11667-11672.
King, E. B., K. L. Chew, et al. (1983). "Nipple aspirate cytology for the study of breast cancer precursors." J Natl Cancer Inst **71**(6): 1115-1121.
Kogo, R., T. Shimamura, et al. (2011). "Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers." Cancer Res **71**(20): 6320-6326.
Korostowski, L., N. Sedlak, et al. (2012). "The Kcnq1ot1 long non-coding RNA affects chromatin conformation and expression of Kcnq1, but does not regulate its imprinting in the developing heart." PLoS Genet **8**(9): e1002956.
Kosters, J. P. and P. C. Gotzsche (2003). "Regular self-examination or clinical examination for early detection of breast cancer." Cochrane Database Syst Rev(2): CD003373.
Krepischi, A. C., M. I. Achatz, et al. (2012). "Germline DNA copy number variation in familial and early-onset breast cancer." Breast Cancer Res **14**(1): R24.
Kuerer, H. M., C. T. Albarracin, et al. (2009). "Ductal carcinoma in situ: state of the science and roadmap to advance the field." J Clin Oncol **27**(2): 279-288.
Lacey, J. V., Jr., S. S. Devesa, et al. (2002). "Recent trends in breast cancer incidence and mortality." Environ Mol Mutagen **39**(2-3): 82-88.
Lee, A. K., R. A. DeLellis, et al. (1984). "Alpha-lactalbumin as an immunohistochemical marker for metastatic breast carcinomas." Am J Surg Pathol **8**(2): 93-100.
Lee, E. H., S. K. Park, et al. (2010). "Effect of BRCA1/2 mutation on short-term and long-term breast cancer survival: a systematic review and meta-analysis." Breast Cancer Res Treat **122**(1): 11-25.
Lengauer, C., K. W. Kinzler, et al. (1997). "Genetic instability in colorectal cancers." Nature **386**(6625): 623-627.
Lengauer, C., K. W. Kinzler, et al. (1998). "Genetic instabilities in human cancers." Nature **396**(6712): 643-649.
Lewin, R. (1982). "Repeated DNA still in search of a function." Science **217**(4560): 621-623.
Liaw, D., D. J. Marsh, et al. (1997). "Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome." Nat Genet **16**(1): 64-67.
Lin, R., S. Maeda, et al. (2007). "A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas." Oncogene **26**(6): 851-858.
Linderholm, B. K., T. Lindahl, et al. (2001). "The expression of vascular endothelial growth factor correlates with mutant p53 and poor prognosis in human breast cancer." Cancer Res **61**(5): 2256-2260.
Livasy, C. A., G. Karaca, et al. (2006). "Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma." Mod Pathol **19**(2): 264-271.
Loo, L. W., D. I. Grove, et al. (2004). "Array comparative genomic hybridization analysis of genomic alterations in breast cancer subtypes." Cancer Res **64**(23): 8541-8549.
Lu, K. H., W. Li, et al. (2013). "Long non-coding RNA MEG3 inhibits NSCLC cells proliferation and induces apoptosis by affecting p53 expression." BMC Cancer **13**: 461.
Lu, X., Y. Fang, et al. (2013). "Downregulation of gas5 increases pancreatic cancer cell proliferation by regulating CDK6." Cell Tissue Res **354**(3): 891-896.
Luo, J. H., B. Ren, et al. (2006). "Transcriptomic and genomic analysis of human hepatocellular carcinomas and hepatoblastomas." Hepatology **44**(4): 1012-1024.
Luo, M., Z. Li, et al. (2013). "Long non-coding RNA H19 increases bladder cancer metastasis by associating with EZH2 and inhibiting E-cadherin expression." Cancer Lett **333**(2): 213-221.
Ma, X. J., R. Salunga, et al. (2003). "Gene expression profiles of human breast cancer progression." Proc Natl Acad Sci U S A **100**(10): 5974-5979.

Maass, N., K. Nagasaki, et al. (2002). "Expression and regulation of tumor suppressor gene maspin in breast cancer." Clin Breast Cancer **3**(4): 281-287.
Malkin, D., F. P. Li, et al. (1990). "Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms." Science **250**(4985): 1233-1238.
Mandelblatt, J. S., K. A. Cronin, et al. (2009). "Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms." Ann Intern Med **151**(10): 738-747.
Martianov, I., A. Ramadass, et al. (2007). "Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript." Nature **445**(7128): 666-670.
Matouk, I. J., I. Abbasi, et al. (2009). "Highly upregulated in liver cancer noncoding RNA is overexpressed in hepatic colorectal metastasis." Eur J Gastroenterol Hepatol **21**(6): 688-692.
Mercer, T. R., D. J. Gerhardt, et al. (2012). "Targeted RNA sequencing reveals the deep complexity of the human transcriptome." Nat Biotechnol **30**(1): 99-104.
Meyer, J. S. and J. L. Wittliff (1991). "Regional heterogeneity in breast carcinoma: thymidine labelling index, steroid hormone receptors, DNA ploidy." Int J Cancer **47**(2): 213-220.
Michor, F. and K. Polyak (2010). "The origins and implications of intratumor heterogeneity." Cancer Prev Res (Phila) **3**(11): 1361-1364.
Miles D, C. A., Romieu G, et al (2008). "Randomized, double-blind, placebocontrolled, phase III study of bevacizumab with docetaxel or docetaxel with placebo as first-line therapy for patients with locally recurrent or metastatic breast cancer (mBC): AVADO." J Clin Oncol.
Miller, K., M. Wang, et al. (2007). "Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer." N Engl J Med **357**(26): 2666-2676.
Mizrahi, A., A. Czerniak, et al. (2009). "Development of targeted therapy for ovarian cancer mediated by a plasmid expressing diphtheria toxin under the control of H19 regulatory sequences." J Transl Med **7**: 69.
Mohammad, F., T. Mondal, et al. (2009). "Epigenetics of imprinted long noncoding RNAs." Epigenetics **4**(5): 277-286.
Mohammad, F., R. R. Pandey, et al. (2008). "Kcnq1ot1/Lit1 noncoding RNA mediates transcriptional silencing by targeting to the perinucleolar region." Mol Cell Biol **28**(11): 3713-3728.
Mourtada-Maarabouni, M., M. R. Pickard, et al. (2009). "GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer." Oncogene **28**(2): 195-208.
Nagano, T., J. A. Mitchell, et al. (2008). "The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin." Science **322**(5908): 1717-1720.
Nelson, H. D., L. H. Huffman, et al. (2005). "Genetic risk assessment and BRCA mutation testing for breast and ovarian cancer susceptibility: systematic evidence review for the U.S. Preventive Services Task Force." Ann Intern Med **143**(5): 362-379.
O'Brien, N., T. M. Maguire, et al. (2002). "Mammaglobin a: a promising marker for breast cancer." Clin Chem **48**(8): 1362-1364.
O'Shaughnessy, J. (2007). "A decade of letrozole: FACE." Breast Cancer Res Treat **105 Suppl 1**: 67-74.
Page, D. L., W. D. Dupont, et al. (1982). "Intraductal carcinoma of the breast: follow-up after biopsy only." cancer **49**(4): 751-758.
Panzitt, K., M. M. Tschernatsch, et al. (2007). "Characterization of HULC, a novel gene with striking up-regulation in hepatocellular carcinoma, as noncoding RNA." Gastroenterology **132**(1): 330-342.
Perou, C. M., T. Sorlie, et al. (2000). "Molecular portraits of human breast tumours." Nature **406**(6797): 747-752.
Pinto, A. E., P. Monteiro, et al. (2005). "Prognostic biomarkers in renal cell carcinoma: relevance of DNA ploidy in predicting disease-related survival." Int J Biol Markers **20**(4): 249-256.
Poliseno, L., L. Salmena, et al. (2010). "A coding-independent function of gene and pseudogene mRNAs regulates tumour biology." Nature **465**(7301): 1033-1038.
Ponting, C. P., P. L. Oliver, et al. (2009). "Evolution and functions of long noncoding RNAs." Cell **136**(4): 629-641.
Pradhan, M., V. M. Abeler, et al. (2012). "Prognostic importance of DNA ploidy and DNA index in stage I and II endometrioid adenocarcinoma of the endometrium." Ann Oncol **23**(5): 1178-1184.

Prensner, J. R., M. K. Iyer, et al. (2011). "Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression." Nat Biotechnol **29**(8): 742-749.
Prensner, J. R., M. K. Iyer, et al. (2013). "The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex." Nat Genet **45**(11): 1392-1398.
Prey, M. U., J. S. Meyer, et al. (1985). "Heterogeneity of breast carcinomas determined by flow cytometric analysis." J Surg Oncol **29**(1): 35-39.
Qiu, J. J., Y. Y. Lin, et al. (2014). "Overexpression of long non-coding RNA HOTAIR predicts poor patient prognosis and promotes tumor metastasis in epithelial ovarian cancer." Gynecol Oncol **134**(1): 121-128.
Rinn, J. L., M. Kertesz, et al. (2007). "Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs." Cell **129**(7): 1311-1323.
Rodriguez, B. A., Y. I. Weng, et al. (2011). "Estrogen-mediated epigenetic repression of the imprinted gene cyclin-dependent kinase inhibitor 1C in breast cancer cells." Carcinogenesis **32**(6): 812-821.
Romond, E. H., E. A. Perez, et al. (2005). "Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer." N Engl J Med **353**(16): 1673-1684.
Ross, J. S. and J. A. Fletcher (1999). "HER-2/neu (c-erb-B2) gene and protein in breast cancer." Am J Clin Pathol **112**(1 Suppl 1): S53-67.
Roylance, R., P. Gorman, et al. (1999). "Comparative genomic hybridization of breast tumors stratified by histological grade reveals new insights into the biological progression of breast cancer." Cancer Res **59**(7): 1433-1436.
Russo, J. and I. H. Russo (1992). "The pathology of breast cancer: staging and prognostic indicators." J Am Med Womens Assoc **47**(5): 181-187.
Sauter, E. R., W. Zhu, et al. (2002). "Proteomic analysis of nipple aspirate fluid to detect biologic markers of breast cancer." Br J Cancer **86**(9): 1440-1443.
Schor, I. E., D. Lleres, et al. (2012). "Perturbation of chromatin structure globally affects localization and recruitment of splicing factors." PLoS One **7**(11): e48084.
Schvimer, M., R. H. Lash, et al. (1995). "Intratumoral heterogeneity of DNA ploidy in breast carcinomas: a flow cytometric assessment of sampling techniques." Cytometry **22**(4): 292-296.
Scully, R., J. Chen, et al. (1997). "Association of BRCA1 with Rad51 in mitotic and meiotic cells." Cell **88**(2): 265-275.
Scully, R. and D. M. Livingston (2000). "In search of the tumour-suppressor functions of BRCA1 and BRCA2." Nature **408**(6811): 429-432.
Shao, Z. M., M. Nguyen, et al. (2000). "Human breast carcinoma desmoplasia is PDGF initiated." Oncogene **19**(38): 4337-4345.
Siddiq, A., F. J. Couch, et al. (2012). "A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11." Hum Mol Genet **21**(24): 5373-5384.
Siegel, R., D. Naishadham, et al. (2013). "Cancer statistics, 2013." CA Cancer J Clin **63**(1): 11-30.
Simpson, J. F., R. Gray, et al. (2000). "Prognostic value of histologic grade and proliferative activity in axillary node-positive breast cancer: results from the Eastern Cooperative Oncology Group Companion Study, EST 4189." J Clin Oncol **18**(10): 2059-2069.
Smaldone, M. C. and B. J. Davies (2010). "BC-819, a plasmid comprising the H19 gene regulatory sequences and diphtheria toxin A, for the potential targeted therapy of cancers." Curr Opin Mol Ther **12**(5): 607-616.
Smith, I. E. and M. Dowsett (2003). "Aromatase inhibitors in breast cancer." N Engl J Med **348**(24): 2431-2442.
Sonkoly, E., Z. Bata-Csorgo, et al. (2005). "Identification and characterization of a novel, psoriasis susceptibility-related noncoding RNA gene, PRINS." J Biol Chem **280**(25): 24159-24167.
Sone, M., T. Hayashi, et al. (2007). "The mRNA-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of neurons." J Cell Sci **120**(Pt 15): 2498-2506.
Sorin, V., P. Ohana, et al. (2012). "H19-promoter-targeted therapy combined with gemcitabine in the treatment of pancreatic cancer." ISRN Oncol **2012**: 351750.
Staaf, J., G. Jonsson, et al. (2011). "Landscape of somatic allelic imbalances and copy number alterations in HER2-amplified breast cancer." Breast Cancer Res **13**(6): R129.

Stephens, P. J., P. S. Tarpey, et al. (2012). "The landscape of cancer genes and mutational processes in breast cancer." Nature **486**(7403): 400-404.
Struewing, J. P., R. E. Tarone, et al. (1996). "BRCA1 mutations in young women with breast cancer." Lancet **347**(9013): 1493.
Sun, M., R. Xia, et al. (2013). "Downregulated long noncoding RNA MEG3 is associated with poor prognosis and promotes cell proliferation in gastric cancer." Tumour Biol.
Szabo, C., A. Masiello, et al. (2000). "The breast cancer information core: database design, structure, and scope." Hum Mutat **16**(2): 123-131.
Tanaka, K., G. Shiota, et al. (2001). "Loss of imprinting of long QT intronic transcript 1 in colorectal cancer." Oncology **60**(3): 268-273.
Tavassoli, F. A., P. Devilee, et al. (2003). Pathology and genetics of tumours of the breast and female genital organs. LyonOxford, International Agency for Research on Cancer ;Oxford University Press (distributor).
Therapy, M. M. o. D. a. (2003). Breast Disorders: Breast Cancer.
Thomas, C. A., Jr. (1971). "The genetic organization of chromosomes." Annu Rev Genet **5**: 237-256.
Tripathi, V., J. D. Ellis, et al. (2010). "The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation." Mol Cell **39**(6): 925-938.
Tripathi, V., Z. Shen, et al. (2013). "Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB." PLoS Genet **9**(3): e1003368.
Tsai, M. C., O. Manor, et al. (2010). "Long noncoding RNA as modular scaffold of histone modification complexes." Science **329**(5992): 689-693.
Tsang, W. P., E. K. Ng, et al. (2010). "Oncofetal H19-derived miR-675 regulates tumor suppressor RB in human colorectal cancer." Carcinogenesis **31**(3): 350-358.
Tsuiji, H., R. Yoshimoto, et al. (2011). "Competition between a noncoding exon and introns: Gomafu contains tandem UACUAAC repeats and associates with splicing factor-1." Genes Cells **16**(5): 479-490.
Veronesi, U., G. Paganelli, et al. (1997). "Sentinel-node biopsy to avoid axillary dissection in breast cancer with clinically negative lymph-nodes." Lancet **349**(9069): 1864-1867.
Virnig, B. A., T. M. Tuttle, et al. (2010). "Ductal carcinoma in situ of the breast: a systematic review of incidence, treatment, and outcomes." J Natl Cancer Inst **102**(3): 170-178.
Wang, J., J. Zhang, et al. (2004). "Mouse transcriptome: neutral evolution of 'non-coding' complementary DNAs." Nature **431**(7010): 1 p following 757; discussion following 757.
Wang, X., S. Arai, et al. (2008). "Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription." Nature **454**(7200): 126-130.
Warrick, J. I., S. A. Tomlins, et al. (2014). "Evaluation of tissue PCA3 expression in prostate cancer by RNA in situ hybridization--a correlative study with urine PCA3 and TMPRSS2-ERG." Mod Pathol **27**(4): 609-620.
Willingham, A. T., A. P. Orth, et al. (2005). "A strategy for probing the function of noncoding RNAs finds a repressor of NFAT." Science **309**(5740): 1570-1573.
Wooster, R., J. Mangion, et al. (1992). "A germline mutation in the androgen receptor gene in two brothers with breast cancer and Reifenstein syndrome." Nat Genet **2**(2): 132-134.
Yamada, K., J. Kano, et al. (2006). "Phenotypic characterization of endometrial stromal sarcoma of the uterus." Cancer Sci **97**(2): 106-112.
Yang, Z., L. Zhou, et al. (2011). "Overexpression of long non-coding RNA HOTAIR predicts tumor recurrence in hepatocellular carcinoma patients following liver transplantation." Ann Surg Oncol **18**(5): 1243-1250.
Yap, K. L., S. Li, et al. (2010). "Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a." Mol Cell **38**(5): 662-674.
Yiangou, C., J. J. Gomm, et al. (1997). "Fibroblast growth factor 2 in breast cancer: occurrence and prognostic significance." Br J Cancer **75**(1): 28-33.
Yoon, J. H., K. Abdelmohsen, et al. (2012). "LincRNA-p21 suppresses target mRNA translation." Mol Cell **47**(4): 648-655.
Yu, W., D. Gius, et al. (2008). "Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA." Nature **451**(7175): 202-206.

Zhang, H., M. J. Zeitz, et al. (2014). "Long noncoding RNA-mediated intrachromosomal interactions promote imprinting at the Kcnq1 locus." J Cell Biol **204**(1): 61-75.
Zhang, J. X., L. Han, et al. (2013). "HOTAIR, a cell cycle-associated long noncoding RNA and a strong predictor of survival, is preferentially expressed in classical and mesenchymal glioma." Neuro Oncol **15**(12): 1595-1603.
Zhang, X., R. Gejman, et al. (2010). "Maternally expressed gene 3, an imprinted noncoding RNA gene, is associated with meningioma pathogenesis and progression." Cancer Res **70**(6): 2350-2358.
Zhang, X., Y. Zhou, et al. (2003). "A pituitary-derived MEG3 isoform functions as a growth suppressor in tumor cells." J Clin Endocrinol Metab **88**(11): 5119-5126.
Zhao, J., B. K. Sun, et al. (2008). "Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome." Science **322**(5902): 750-756.
Zhao, Y., Q. Guo, et al. (2014). "Role of long non-coding RNA HULC in cell proliferation, apoptosis and tumor metastasis of gastric cancer: A clinical and in vitro investigation." Oncol Rep **31**(1): 358-364.
http://www.cdc.gov/cancer/breast/statistics/.