COMMITTEE CERTIFICATION OF APPROVED VERSION

The committee for Yafei Huang certifies that this is the approved version of the following dissertation:

EARLY MARKERS OF BREAST CANCER IN NIPPLE ASPIRATE FLUID

Committee:

Lee-Jane W Lu, Supervisor

Jonathan Ward

Alexander Kurosky

Karl E Anderson

Anthony M Haag

Suzanne AW Fuqua

Dean, Graduate School

EARLY MARKERS OF BREAST CANCER IN NIPPLE ASPIRATE FLUID

by
Yafei Huang

Dissertation
Presented to the Faculty of The University of Texas Graduate School of Biomedical Sciences at Galveston
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

Approved by the Supervisory Committee

Lee-Jane W Lu
Jonathan Ward
Karl E Anderson
Alexander Kurosky
Anthony M Haag
Suzanne AW Fuqua

July, 2007
Galveston, Texas

Key words: breast secretion, proteomics, biomarker

To my dear parents, Xidong Huang and Jinrong Kong,
who always motivate me to move forward;

To my wonderful husband, Ping Liu,
who has provided unselfish support for my work throughout the years;

To my two lovely angels, Marissa and Serena,
who bring peace, joy as well as chaos into our lives.

ACKNOWLEDGEMENTS

First and foremost I give my heartfelt thanks to my mentor, Dr. Lee-Jane Lu, for her exhaustive dedication and unselfish support throughout my graduate education here at UTMB. Despite her busy work schedule, she is always available whenever I have questions, concerns or need help. I still remember that the day before my first public presentation she rehearsed the talk with me for several hours, giving advice on everything from English, style of talking to background choice of the power point slides. She makes every effort to ensure that I receive proper training in scientific writing, research design, and new technologies used in our field. Her enthusiasm in research, her meticulous attitude in treating scientific findings, her patience in mentoring trainees, and her kindness towards anyone who is in need have been a true inspiration for my career and my life. All past and present members of Dr. Lu's lab have also provided tremendous help for my research, without which I could not have gotten this far today. In particular, I would like to thank Neelam Bhopale for her unconditional support in both my research and my personal life throughout the years.

My appreciation also goes to Dr. Alexander Kurosky and Dr. Anthony Haag for being the "technical support" and mentors of mine on all the proteomic techniques, Dr. Karl E. Anderson for advising on patient recruitment and participation, Dr. Jonathan Ward for providing academic and personal guidance for me at difficult times, Dr. James J Grady for supervising statistical data analysis, Dr. Suzanne Fuqua for being on my dissertation committee and giving valuable input on this project, Dr. Marinel Ammenheuser for editing all of my manuscripts and tutoring me on English writing through the years, Dr. Astrid Inniss for analyzing data for food records, and the staff in Breast Imaging and the GCRC for helping with patient recruitment and sample collections.

Grants and fellowships supported this project included U.S. Army MRMC under W81XWH-04-1-0345, U.S. Army MRMC under DAMD17-01-1-0417, NIH NCRR GCRC M01 RR00073, NIH R01 CA95545, NIH 2 P30 ES06676, 1 R24 CA88317, AICR grant 01B110, and USPHS CA65628

EARLY MARKERS OF BREAST CANCER IN NIPPLE ASPIRATE FLUID

Publication No._____

Yafei Huang, Ph.D.
The University of Texas Graduate School of Biomedical Sciences at Galveston, 2007

Supervisor: Lee-Jane W. Lu, Ph.D.

Nipple aspirate fluid (NAF) refers to the small amount of secretion that is found in breast ducts/lobules of most non-lactating women. This fluid can be collected repeatedly and non-invasively via the nipple using a modified breast pump, and therefore, it is considered to be a potential source for identifying markers of breast cancer. The purpose of this study was to understand factors associated with the ability to secrete fluid and factors associated with the major protein profiles in NAF; and to identify protein profiles of NAF in a group of healthy non-lactating women who were 30-40 years old, not pregnant, not breastfeeding, and not taking contraceptive medications.

Among 238 women studied, 66% were secretors of NAF. Using multivariate logistic regression models, higher dietary intake of lactose [Odds Ratio (OR)=2.7; 95% Confidence Interval (CI): 1.5-4.8], earlier menarche (OR=0.8, CI: 0.7-1.0), being parous (OR=2.3, CI: 1.0-5.6), and older at first childbirth (OR=1.5, CI: 1.0-2.1) were found to be independent and positive predictors for being a secretor of NAF. These findings suggest that dietary intake of lactose, a modifiable factor, may be used to change the NAF secretor status of women.

NAF were analyzed for major proteins. Two major types of protein profiles, type I and type II, were identified. Type I NAF contains proteins found in cystic disease fluid of the breast, whereas type II NAF is enriched in milk-associated proteins. Using multiple logistic regression, type I NAF was predicted independently ($P<0.05$) by higher body fat mass (Odds Ratio=3.0; CI: 1.5-6.1), more years since last childbirth (OR=2.6; 95% CI: 1.3-5.2) and a higher percentage of calories from saturated fat (OR=4.1; 95% CI: 1.1-14.6). These results suggest that protein profiles of NAF might be influenced by amounts and/or types of dietary and body fat.

Two different analytical strategies, 2D gel analysis coupled with MALDI-TOF/TOF, and 1D gel coupled with LC-MS/MS, were used to characterize protein profiles of type I and II NAF. Using these two strategies, a total of 99 proteins were identified: 13 unique to type I NAF, 57 unique to type II NAF, and 29 common to both types. These strategies will be used to characterize proteins in NAF of breast cancer cases.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

2D gel	two-dimensional gel electrophoresis
2D-LC	two-dimensional liquid chromatography
ApoD	apolipoprotein D
bFGF	basic fibroblast growth factor
BMI	body mass index (kg/m^2)
CEA	carcinoembryonic antigen
CHAPS	3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate
DCIS	ductal carcinoma in situ
DEXA	dual energy X-ray absorptiometry
DTT	dithiothreitol
GCDFP-15	gross cystic disease fluid protein 15 (prolactin-inducible protein)
HDL	high density lipoprotein
IEF	isoelectric focusing
LDL	low density lipoprotein
MALDI-TOF	matrix-assisted laser desorption/ionization time-of-flight
MS	mass spectrometry
MS/MS	tandem mass spectrometry
NAF	nipple aspirate fluid
PSA	prostate specific antigen
SDS PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SELDI-TOF	surface enhanced laser desorption/ionization time-of-flight
SERMS	selective estrogen receptor modulators
T3	triiodothyronine
T4	tetraiodothyronine
TCA	trichloroacetic acid
TSH	thyroid stimulating hormone
vEGF	vascular endothelial growth factor
VLDL	very low density lipoprotein
ZAG	zinc α2-glycoprotein

CHAPTER 1: GENERAL INTRODUCTION

"The breast is a gland which throughout life is exhibiting secretory activity, the difference between a lactating breast being one partly of degree and partly of the chemical composition"
------Geoffrey Keynes, 1923

1. 1 Breast anatomy

The human breast is mainly composed of glandular tissue and the supporting connective tissues including adipose and fibrous/dense connective tissue. The breast glands consist of approximately 10-15 lobular-ductal units that extend from the nipple, course through the mammary connective tissues, and terminate in grape-like clusters of alveoli (Figure 1.1). Each duct serves a specific lobule and opens to the nipple, which is surrounded by an area of pigmented skin, the areola. Each mammary duct expands slightly to form a sinus beneath the areola. These sinuses function as reservoirs for breast secretions. The areola also serves as the termination point for the fourth intercostal nerve which carries signals of suckling to the spinal cord and brain. In lactating women, this mechanism is extremely important in regulating the secretion of oxytocin and prolactin from the pituitary.

In addition to the supporting connective tissues, the breast epithelia are also surrounded by blood vessels, nerves and lymphatics which may be involved in the nourishment of the breast glands as well as the production and clearance of breast secretions. Another structure that is important for the secretory activities of the breast is the myoepithelial cells, which circle around the breast ducts and induce the ejection of milk when contracted after stimulation. These structures are probably involved in producing breast secretions in non-lactating women as well.

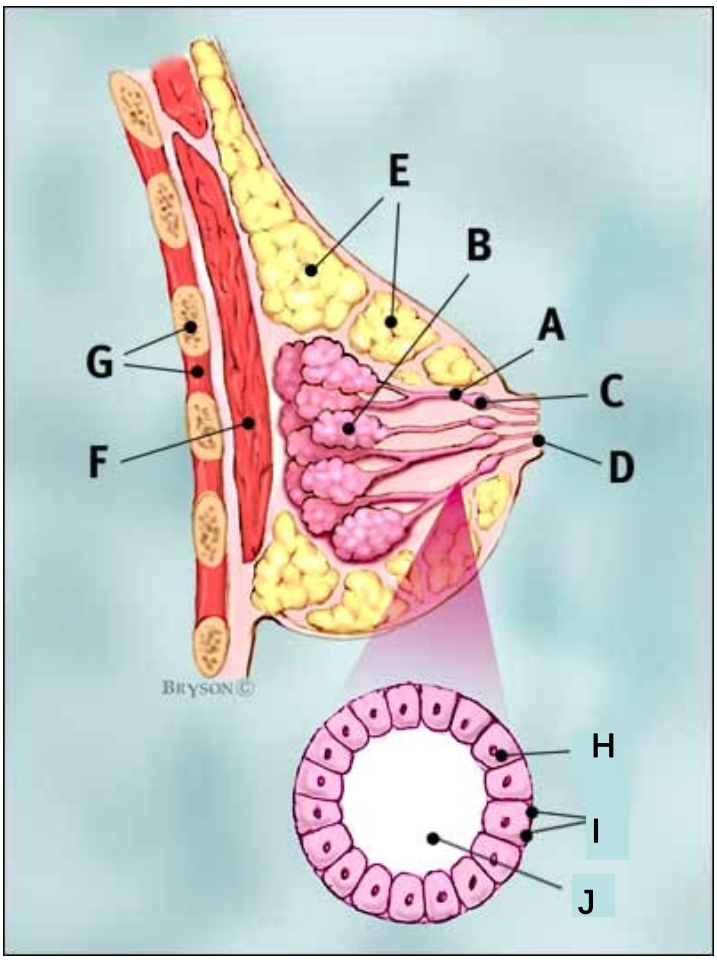


Figure 1.1. Medial lateral view of the female breast: A, breast ducts; B, lobules; C, dilated ducts as reservoirs of milk; D, nipple; E, adipose tissue; F, pectoralis major muscle; G, chest wall; H, ductal epithelial cells; I, basement membrane; J, lumen (Figure adapted from www.breastcancer.org).

1.2 The epidemiology of breast cancer

Breast cancer is the most frequently diagnosed cancer and the second leading cause of cancer death in the US women. It is estimated that in 2007, 178,480 women will be diagnosed with breast cancer and 40,460 women will die from this disease in the United States. The lifetime risk of developing breast cancer for women in the US is one out of eight.

Breast cancer is a heterogeneous disease that may encompass a wide range of histological and pathological presentations, such as invasive ductal carcinoma, lobular carcinoma, medullary carcinoma, mucinous carcinoma, tubular carcinoma, and apocrine carcinoma e*tc*. However, most breast cancer develops in glandular tissue and is generally classified as adenocarcinoma. The earliest form of this disease, ductal carcinoma in situ (DCIS), arises solely in these milk ducts. Invasive ductal carcinoma, which is the most common type and accounts for 70-80% of all breast cancers, develops from DCIS, spreads through the duct walls, and invades other types of breast tissues [1,2].

Cancer in general is a multi-step process which involves initiation, promotion, and progression. Initiation of cancer is usually induced by exposure to a genotoxic carcinogen and/or mutations in a critical gene (for example, *p53*) in a single cell. Promotion of cancer is often stimulated by a promoting agent, for example, estrogen in the case of breast cancer, through the proliferation of the mutated cell. Promotion of cancer results in selective cloning/focal proliferation of the initiated cell. Progression of cancer may involve both mutagenic and non-mutagenic effects that lead to morphological changes and malignant features of the affected tissues/organs, such as invasion and metastasis. The cause of breast cancer is not entirely clear. Familial breast cancers account for about 10% of all cases [3]. The etiology of the remaining breast cancer cases is largely unknown. The promotion phase of breast cancer is most often affected by the presence of excess estrogen and progesterone, which stimulates the growth of breast tumors. Risk

factors for breast cancers include: age (over 50), first completed pregnancy after age 30, long-term hormone replacement therapy (HRT) and obesity in postmenopausal women, menarche before age 12, menopause after age 50, and nulliparity.

Depending on cancer stage, treatment options for breast cancer include surgery, chemotherapy, radiation therapy, hormone therapy, and more recently, immune therapy for HER2+ breast cancers. The most important factor that determines the prognosis of this disease after treatment is the stage of breast cancer. In general, the earlier the stage of detection, the better the prognosis. Therefore, early detection is considered key to decreasing the mortality associated with this disease. However, the most widely used diagnostic methods, including breast imaging and palpation, are useful only after a lump of cancer cells has formed in the breast, which is a late stage in mammary carcinogenesis. Currently, there is no specific blood test for breast cancer. New diagnostic tools that can detect earlier lesions are needed.

1.3 Production of breast secretions

In lactating women, the breast epithelium is the site where most constituents of the milk are formed and assembled. However, other types of cells in the stroma of the breast, including fibroblasts, adipocytes, and cells in the blood stream, also contribute to the production of components of milk.

As shown in Figure 1.2, five distinct mechanisms, which operate in parallel, have been recognized for the formation of different constituents in milk during lactation [4]. First, the exocytosis of secretory vesicles, which are Golgi-derived envelopes, provides certain proteins and lactose. Second, lipid droplets derived from smooth endoplasmic reticulum are secreted by apocrine mechanisms. Third, interstitial molecules derived from the blood stream and other cells in stroma are secreted by transcytosis. Biochemicals transported via endocrine hormones, such as insulin, prolactin, and IGF-1; cytokines; and lipoprotein

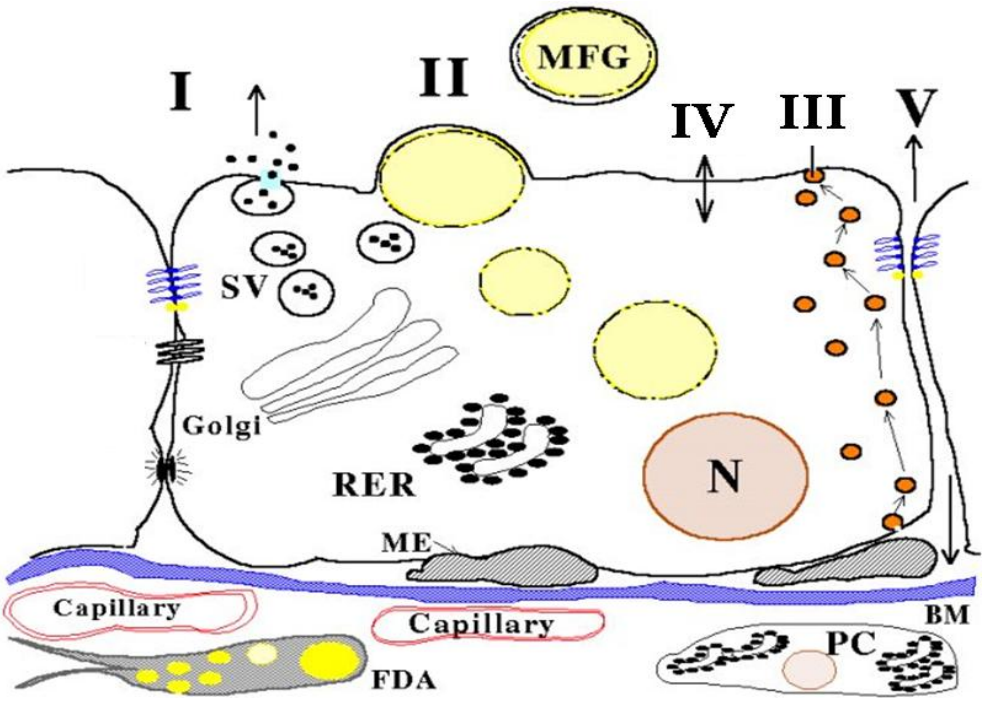


Figure 1.2. Mechanisms for milk secretion: I, exocytosis; II, apocrine secretion; III transcytosis; IV, passive transport; and V, paracellular pathway. N, nucleus; SV, secretory vesicle; FDA, fat-depleted adipocyte; PC, Plasma Cell; BM, basement membrane; ME, myoepithelial cell; RER, rough endoplasmic reticulum; MFG, milk fat globule. (Figure adapted from Biology of the Mammary Gland, mammary.nih.gov).

lipase. Four, small molecules, such as water, iron, glucose, amino acids, and drugs, are transported across the basal and apical plasma membranes by passive diffusion. Fifth, paracellular transport of constituents between the interstitial space and the milk duct lumen bypasses epithelial cells, and also contributes to milk components. This pathway is active during pregnancy and involution or in the presence of mastitis.

In most non-lactating women, a small amount of fluid (1-200 μl) is continuously produced by the breast lobular and ductal systems, and this fluid can be collected non-invasively and repeatedly from the nipples by gentle aspiration [5,6]. The fluid that is collected in this manner has been called nipple aspirate fluid (NAF). Like breast milk, NAF is rich in proteins, hormones, lipids, and carbohydrates [7].

The mechanisms for the secretion of NAF in non-lactating women have not been fully explored. It is not clear which of the above five pathways might be involved. It is very possible that the synthesis of NAF is an activity in which various mammary cells and tissues contribute to its contents in a coordinated fashion. We and others have speculated that the composition of NAF is potentially an important index of the metabolic and physiological activities of the breast. NAF is, therefore, a useful window for probing the physiological and pathological conditions of the breast. This study focuses on NAF proteins that may be useful as biomarkers of normal and abnormal breast processes, including cancer development.

1.4 Procedures for obtaining NAF

In general, NAF is obtained by gentle suction after cleansing, which removes plugs in the ductiles where they exit the nipples. However, the specific procedures for obtaining NAF may vary from laboratory to laboratory. In our laboratory, the standard procedure consists of the following three steps:

Step 1: Warm up the breast with a heating device.
The subject is asked to lie down in a supine position. A heating pillow is warmed up in a microwave to about 50-60°C. It is wrapped around the breast for about three minutes. This step helps to relax and dilate the breast glands and eases the flow of the secretion from the breast.

Step 2: Massage the breast.
The subject, in a seated position, is instructed to massage her breast for two minutes in the direction of fluid secretion. This is useful in facilitating fluid collection. This step may stimulate secretion from the lobules, relax the myoepithelia, and help fluid to flow down the ducts.

Step 3: Collect secretion.
After steps 1 and 2, gentle suction is applied to the breast for 15 seconds, using a modified breast pump developed in our laboratory. Drops of fluid usually appear on the nipple after this step. NAF is then collected, using a capillary collection tube, and the NAF is stored in a vial at -80°C.

1.5 History of research on NAF secretion

It has been recognized for decades that the non-lactating breast in adults is capable of secreting fluid into the breast ducts. Because this fluid comes directly from the breast, the target tissue of breast cancer, it is potentially a good source for discovering breast cancer biomarkers. Researchers in the past have found that about 50-70% of women could secrete NAF, while the remaining 30-50% could not [8-10]. In addition, the volume of fluid collected varies from 1 to 200 μl, and the typical yield is 5-10 μl (our observation). For NAF to be a useful source for detecting breast cancer biomarkers, it is important, first, to understand why NAF can not be obtained from all women, to identify the factors associated with the ability to secrete NAF, and to develop possible strategies to induce

secretion in non-secretors. Second, it is necessary to understand whether the components of NAF exhibit inter-individual variability, so that we will have a better understanding of which components are associated with breast cancer.

There has been enormous interest in the search for factors that are associated with the secretion of NAF and the health relevance of NAF secretion. Petrakis *et al.* [6] suggested that the secretion of NAF might be associated with increased risk of breast cancer. Because mammary carcinogens have been detected in NAF, they speculated that prolonged soaking of breast epithelia with these carcinogens could initiate breast cancer among secretors. However, current or past history of benign breast diseases have not been found to be associated with ability to secrete NAF [11-14]. Table 1.1 summarizes studies that investigated the associations of demographic factors, hormones, nutrients, socioeconomic status, and breast disease with the ability to secrete NAF. Family history of breast cancer, type of menopause, marital status, smoking, phase of menstrual cycle, and serum estrogen level were not associated with secretor status of NAF. Age of 30-55 years, early age of menarche, past history of parity and/or lactation, non-Asian ethnicity, wet-type cerumen, high fat intake, soy consumption, certain endocrine disorders, and oxytocin usage have been associated with being a secretor of NAF. These results indicate that the secretion of NAF may be under the regulation of multiple factors, and that these factors may also interact with each other. Very few studies, however, have evaluated the independent effects of individual factors in multiple-adjusted models. The first specific aim of this study is, therefore, to evaluate factors that can independently predict the secretor status of nipple aspirate fluid, using multivariate logistic regression models.

Another focus in NAF research has been analyses of the composition of this fluid as related to health, as detailed in Table 1.2. Components that have been detected in NAF include various types of cells, proteins, lipids, carbohydrates, DNA, hormones and micronutrients. The detection of malignant epithelial cells in NAF provides a very specific diagnosis of breast cancer. In addition, the presence of atypical/hyperplasic

Table 1.1. Literature review of studies on factors associated with secretor status of NAF

Factors		Significance	Literature
Demographics/ Reproductives/ Genetics	Age 30-55 years	positively associated with secretion of NAF	[13,15]
	Early age of menarche	positively associated with secretion of NAF	[13,15]
	History of parity	positively associated with secretion of NAF	[12,16]
	History of lactation	positively associated with secretion of NAF	[12,13]
	None Asian ethnicity	positively associated with secretion of NAF	[16,17]
	Family history of breast cancer	not associated with secretion of NAF	[17]
	Menopause status	premenopause positively associated with secretion of NAF	[12,13]
	Type of menopause	not associated with secretion of NAF	[18]
	Marital status	not consistent	[19]
	Wet-type cerumen	positively associated with secretion of NAF	[19]
	Age at first birth	not associated with secretion of NAF	[15]
Diet	High fat consumption	positively associated with secretion of NAF	[20]
	Soy consumption	positively associated with secretion of NAF	[21]
	Cigarette smoking	not associated with secretion of NAF	[14,22]
Endocrine	Endocrine disorders	positively associated with secretion of NAF	[19]
	Bilateral salpingo-oophorectomy	negatively associated with secretion of NAF	[10]
	Oxytocin usage	positively associated with secretion of NAF	[23]
	Serum estrogen levels	not associated with secretion of NAF	[18]
	Oral contraceptive use	not consistent	[17]
	Phase of menstrual cycle	not associated	[16]
	Tranquilizer use	not consistent	[19,22]
	Menopausal estrogen use	not associated with secretion of NAF	[12]
Breast disease	History/current benign breast diseases	not associated with secretion of NAF	[12-15]
Education		not consistent	[15,18]

epithelial cells in NAF has generally been found to be associated with increased risk of breast cancer [27, 28]. However, the sensitivity of cytology for the detection of breast cancer is low, due to the extremely low prevalence of exfoliated cells in NAF. Ductal lavage has been attempted to increase the yields of cells. For the same reason, the small amount of DNA available in NAF may also limit its application for cancer marker screening. The few studies that attempted to investigate the potential of DNA as biomarkers of breast cancer in NAF have yielded inconclusive results on DNA alterations due to limited number of cases studied and available cells in NAF.

In contrast, many biochemical substances, including lipids, hormones, and proteins, are generally found to be more abundant in NAF, with concentrations much higher than in plasma. For example, the average level of proteins in NAF is 71-170 mg/ml, which is much higher than the 6-8 mg/ml reported for plasma [23,24]. This facilitates analyses of these chemicals and their association with disease status. Previous studies demonstrated that high levels of Tf and Tn antigens, basic fibroblast growth factor (bFGF) and cancer embryonic antigen (CEA), and lower levels of prostate specific antigen (PSA) and kallikrein proteins in NAF, have been associated with breast cancer, using antibody-based technologies [32,37-40]. Recent development in proteomic technology, which enabled protein profiling with a very small amount of material, has further advanced research on protein biomarkers. For example, using as little as a few μl of NAF, a global protein profile has been accomplished by the combined use of two-dimensional gel electrophoresis and mass spectrometry, aiming at identifying breast-cancer associated biomarkers [25]. Because proteins play important roles in cellular function, and the amount of material in NAF is limited, we focused our studies on the characterization of proteins as an approach to discover potential biomarkers for breast cancer.

A few prior studies investigated the major protein profiles in NAF, and found two typical patterns of profiles in healthy women (Figure 1.3). The type I profile is enriched in proteins that are also found in gross cystic disease fluid, including zinc-α2-glycoprotein,

Table 1.2. Literature review of contents analyzed in NAF

	Contents in NAF	Significance	Literature
Cells	foam cells	not associated with breast cancer	[6]
	epithelial cells	increased risk for breast cancer	[27]
	atypia/hyperplasia	increased risk for breast cancer	[28]
Lipids	cholesterol	associated with alcohol and tobacco usage and weaning.	[25]
	cholesterol epoxides	not determined	[29]
	8-isoprostane	not determined	[25]
	lipid peroxides	not determined	[25]
Micronutrients	carotenoids	associated with lactation	[25]
Hormones	estradiol	lower in OC users	[30]
	progesterone	lower in OC users and postmenopausal women	[30]
	prolactin	marginally higher in cystic breast disease patients	[31]
	bFGF	high levels associated with breast cancer	[32]
	EGF	high in benign breast disease	[30,31,33]
	TGF-alpha	high in benign breast disease	[31]
	estrone sulfate	lower in OC users	[30]
	IL-6	higher in postmenopausal women	[30]
	leptin	associated with BMI, not associated with breast cancer	[34]
	prostaglandin E2	not responding to COX-2 treatment	[34]
	DHEA	not determined	[35]
	cathepsin-D	not determined	[30,35]
	testosterone	not associated with breast cancer	[36]
Carbohydrates	lactose	associated with coloration of NAF and age	[13]
	TF and Tn antigens	diagnostic of breast cancer	[37]
Proteins	PSA	Not consistent	[38-40]
	CEA	elevation associated with cancer	[40]
	CRP	associated with pregnancy, breast feeding, anthropometrics	[41]
	vitD binding protein	increased in cancer cases	[2]
	kallikrein (2,3,6,10)	inversely associated with breast cancer	[42]
	AAG	increased in cancer samples	[26]
	uPA, uPAR, and PAI-1	levels increased in advanced breast cancer	[43]
	osteopontin	not determined	[44]
	IGFBP3	high levels associated with breast cancer	[45]
DNA	mitochondria DNA alteration	mitochondria DNA mutation found in some cancer samples	[46,47]
	LOH	found in some cancer samples	[46]
	promoter hypermethylation	found in cancer samples not controls	[48]
	microsatellite instability	Associated with disease progression	[49]

apolipoprotein D, and prolactin-inducible protein, while the type II profile is abundant in milk-associated proteins, including lactoferrin, lactalbumin and lysozyme. Despite the similarities in their major proteins, significant differences were found between type II NAF and milk samples in terms of other biochemical contents, as listed in Table 1.3 [49].

It was suggested that the prevalence of these two types of protein profiles differed between controls versus breast cancer patients. However, the mechanisms underlying the formation of these two completely different patterns of protein profiles have not been explored carefully. To shed some light on the possible association of these two patterns with disease risk, it is important first to understand the distribution of these two patterns in a healthy population. Therefore, the second specific aim of this study was to explore demographic, hormonal, anthropometric, reproductive and nutritional factors that are associated with the major protein patterns in women who can secrete NAF but have no mammography-detectable cancer. In addition to differences in the pattern of the major protein profiles, it is highly likely that other proteins of low to medium abundance may also differ between these two profiles. Therefore, the third specific aim of this study was to map the global protein profiles in both types of NAF, and identify proteins that may potentially be used as biomarkers for breast cancer. It was also planned to explore predictors for breast cancer risk by comparing the NAF proteomes in breast cancer cases and controls, but due to a low response rate of recruitment for breast cancer cases, this aim was temporarily postponed.

In summary, the three specific aims were: first, to investigate factors that are associated with secretor status of NAF, and explore possible strategies to increase the production of NAF; second, to evaluate factors that affect the major protein patterns of NAF among secretors; and third, to globally characterize protein profiles of two distinct types of NAF. We had intended to include breast cancer cases, but due to a low response rate in recruiting breast cancer patients, this preliminary investigation has focused on healthy women.

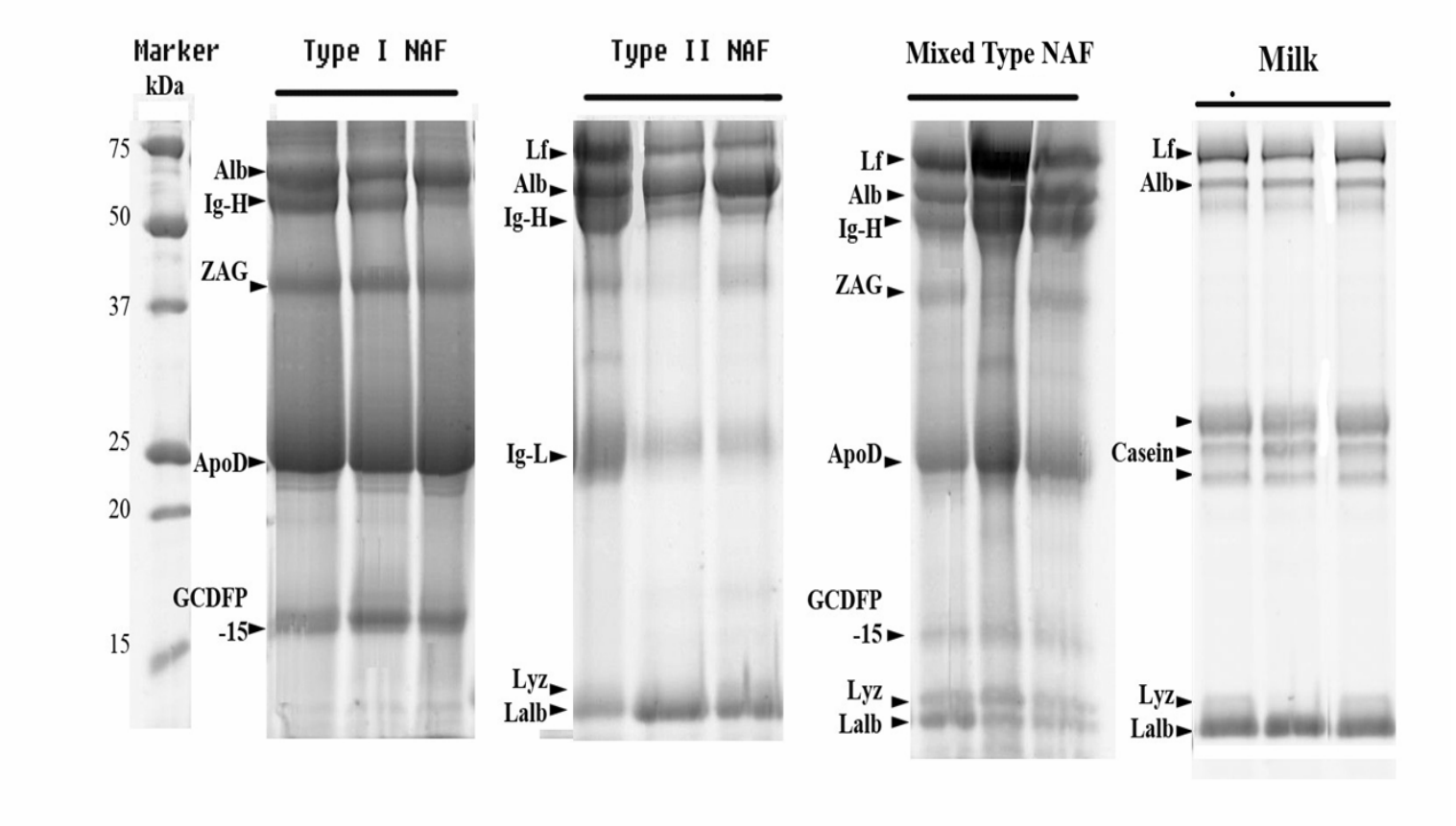


Figure 1.3. One-dimensional (1D) gel (12.5% polyacrylamide) images of proteomes of different types of NAF and breast milk. Lf, lactoferrin; Alb, albumin; Ig-H, immunoglobulin heavy chain; Ig-L, immunoglobulin light chain; Lyz, lysozyme C; Lalb, α-lactalbumin.

Table 1.3. Biochemical profiles of NAF and milk samples

Analyte	**Type II NAF (X ± SE)**	**Milk (X ± SE)**
Na (mmol/l)	119.64 ± 8.82	83.89 ± 11.4
K (mmol/l)	9.53 ± 3.87	37.17 ± 4.77*
Total protein (g/l)	37.88 ± 6.54	12.23 ± 5.93*
Albumin (g/l)	4.55 ± 1.33	0.52 ± 0.11*
Immunoglobulins (mg/l)	8.11 ± 2.24	24.95 ± 6.82*
Lysozyme (g/l)	1.92 ± 0.36	0.49 ± 0.05*
Casein (g/l)	0.12 ± 0.02	3.76 ± 0.42*
a-Lactalbumin (g/l)	1.94 ± 0.07	5.65 ± 0.86*
Prolactin (mg/l)	136.24 ± 9.55	63.94 ± 7.38*
Alkaline phosphate (IU/l)	166.37 ± 31.43	95.45 ± 6.83*
LDH (IU/l)	972.14 ± 59.65	155.17 ± 12.49*
PSA (mg/l)	48.75 ± 9.23	3.23 ± 0.68*
CEA (mg/l)	310.49 ± 19.45	23.48 ± 3.97*
Total cholesterol (g/l)	8.49 ± 3.65	0.24 ± 0.08*
Lactose (g/l)	11.43 ± 4.28	71.44 ± 7.27*

*t-test with $P < 0.05$ showing significant difference.

Table adapted from Malatesta *et al.*, 2000, *Journal of Clinical Laboratory Analysis*

CHAPTER 2: FACTORS AFFECTING THE SECRETOR STATUS OF NIPPLE ASPIRATE FLUID

2.1 Introduction

The success rate for obtaining NAF ranges from 50% to 70%, as reported in the literature [6, 8, 13-15], with variation possibly due to the differences in populations studied. To better interpret risk factors for breast cancer and enable a wider application of NAF analysis for breast cancer detection and risk assessment, we investigated factors that are associated with the ability of women to secrete NAF.

Several large cross-sectional studies have investigated the effects of demographic and genetic features, reproductive history, and hormonal status of women on their ability to secrete NAF. Wrensch *et al.* [17] found a positive association between the secretion of NAF and an earlier age of menarche, age range of 35 to 50 years, non-Asian ethnicity, and prior history of lactation. Miller *et al.* [8] reported that current oral contraceptive users were significantly less likely to yield NAF than those who had never taken those medications. Petrakis *et al.* [28] found no significant differences in serum concentrations of 17β-estradiol and estrone between secretors and non-secretors of NAF. Higgins *et al.* [9] found that a reduced volume of NAF was associated with postmenopausal status, BRCA germline mutations and risk reduction therapies such as salpingo-oophorectomy and use of selective estrogen receptor modulators (SERMS). Lee *et al.* [19] reported a positive association between higher dietary fat consumption and ability to secrete NAF. Petrakis *et al.* [20] found that soy feeding increased the success rate for obtaining NAF and the volume of NAF collected. These data suggested that NAF production may be under the influence of multiple factors. However, few of these studies investigated whether these factors were independent predictors in multivariate-adjusted models. In this study, we examined the nutritional, demographic, reproductive, anthropometric, and

hormonal factors that may be independent predictors of secretor status in a well-defined study population.

2.2 Materials and Methods

2.2.1 Study design

This cross-sectional study included a well-defined population of healthy premenopausal women of all races. The women were 30-40 years old, not users of contraceptive medications (pills, injections, or depots) for at least 6 months, not vegetarians, not pregnant, and not nursing. Women with irregular menstrual cycles or first degree relatives with breast cancer were excluded from the study. All women had normal mammograms at entry to the study. All participating subjects (N=238) were recruited from communities within a 50 mile radius of Galveston, TX by posted advertisements and postal mailings. Written informed consent was obtained from all subjects. A total of six study visits were scheduled during the luteal phases of two menstrual cycles, usually on the days between cycle day 20 and 24, with three visits during each menstrual cycle. The study protocol was approved by the Institutional Review Board of the University of Texas Medical Branch (UTMB) and the Human Subject Research Review Board of the US Army Medical Research and Materiel Command. These women volunteered for a dietary intervention study. Their baseline data obtained prior to the onset of the dietary intervention were the source for this analysis.

2.2.2. Procedures for obtaining NAF

Attempts to acquire NAF were made on three separate study visits (see Chapter 1.4 for detailed procedures for nipple aspiration). A subject was classified as a "non-secretor" if no fluid was obtained from either breast on any of three visits. If fluids were successfully collected from at least one breast on at least one of three study visits, the patient was classified as a NAF secretor. For secretors, the volume of NAF obtained was recorded, and categorized as low (<10 μl), medium (10-30 μl), or high (>30 μl).

2.2.3 Measurements of nutrients, reproductive variables, anthropometrics, hormones, and lipids

Three dietary recall records for the 24 hours preceding three scheduled study visits were obtained and later analyzed using the Nutrition Data System for Research software (developed by the Nutrition Coordinating Center, the University of Minnesota, MN). The average nutrient intake of three 24-hour food records was used for statistical analyses of nutritional influences on secretor status. A food frequency questionnaire (Harvard School of Public Health and the Brigham and Women's semi-quantitative food frequency questionnaires, 96/97 GP) was also administered to each subject at the entry to the study, to assess dietary habits during the 12 months preceding enrollment. Questionnaires were analyzed by the Nutrition Questionnaire Service Center at Harvard School of Public Health (Boston, MA).

Reproductive history was obtained using a self-administered standard clinic questionnaire originally designed for gynecologic consultations. Age of menarche and parity (yes or no) were recorded for all subjects. For parous women, history of lactation (yes or no), cumulative length (months) of breastfeeding, age at first and last childbirth, and number of childbirths were also recorded.

During each study visit, body weight, and height were measured. Total body mass, body lean mass, and body fat mass were measured using dual energy X-ray absorptiometry (DEXA) (Model QDR4500A, Hologic, Waltham, MA). The subjects were examined in a supine position in duplicate, with interval repositioning on the examination table, during one of the study visits. Average values for the duplicate measurements were used for statistical analyses.

Fasting blood samples were obtained during all study visits, but only the first three samples, all from one luteal phase, were used for measuring steroid hormones. All samples collected were immediately stored at -80°C until analysis. Plasma samples were

analyzed for levels of progesterone, testosterone, and 17β-estradiol using commercial immunoassay kits (Diagnostics Labs, Webster, TX) according to the manufacturer's instructions. A radioactive immunoassay kit was used to measure progesterone levels (sensitivity, 0.1 ng/ml). Enzymatic immunoassay kits were used for measuring testosterone levels (sensitivity, 0.04 ng/ml) and 17β-estradiol levels (sensitivity, 7 pg/ml). For all the assays, the average intra- and inter-assay coefficients of variations (CV) were less than 15%. Each sample was assayed at least twice (and was repeated if the intra-assay CV was >15%), and results were averaged. Hormone levels from three different cycle days of the same luteal phase were averaged for each subject for statistical analyses. Triiodothyronine (T3), tetraiodothyronine (T4), thyroid stimulating hormone (TSH), high density lipoprotein (HDL), low density lipoprotein (LDL), very low density lipoprotein (VLDL), total cholesterol, and triglycerides were measured in fasting blood samples from two different study visits by the certified UTMB hospital clinical laboratory. The average of the two measurements was used for statistical analyses.

2.2.5 Statistical analyses

Means and standard deviations were computed for continuous data, and percentages for categorical data. For univariate analyses, the Student's *t*-test was performed on continuous variables for comparisons between secretors and non-secretors, and the χ^2 test used for categorical variables to estimate the distribution of secretors in each category. Pearson's coefficients (r) were computed for correlations among variables of interest. Factors that showed association with NAF secretor status in univariate analysis ($P<0.05$) were further entered into multivariate logistic regression models to determine independent effects of each factor. Odds ratios (OR) and 95% confidence intervals (CI) were estimated using multiple logistic regression for factors that independently predict secretor status of NAF. General Linear Model procedures were used to estimate P trend between lactose/milk intake and the volume of NAF among secretors. All statistical analyses were performed using SAS® (Version 9.1, SAS Institute Inc., Cary, NC).

2.3 Results

2.3.1 Prevalence and characteristics of secretors and nonsecretors in the study population

NAF was successfully obtained from 156 (66%) of 238 women participating in the study (Table 2.1). Our success rate for obtaining NAF is consistent with reports by others [8].

Table 2.1. The frequency and percentage-distribution of secretors

Secretor status	N	Percentage
Secretors	156	66%
Non-secretors	82	34%
Total	238	100%

Table 2.2 describes the nutritional, demographic, anthropometric, reproductive, and hormonal characteristics of our study subjects. Data for nutrient intake was estimated from both food records and food frequency questionnaires. Data from food records were the main source for statistical analyses, unless otherwise specified. Compared to non-secretors, secretors consumed significantly more lactose (by 34%, $P<0.001$, Table 2.2) when analyzed using food records. Since the primary dietary source of lactose is milk, milk intake was compared between secretors and nonsecretors using data from food records. Secretors also had significantly higher consumption of milk compared to nonsecretors (by 37.8 g, $P=0.01$, Table 2.2). Consistent with the above data obtained from food records, secretors on average had significantly higher consumption of lactose (11.1 ± 10.7 g) in the year preceding enrollment in the study compared to non-secretors (8.5 ± 8.0 g, $P=0.05$, not shown in Table 2.2) using data from food frequency questionnaires. In addition, secretors were slightly older in age (by 0.7 years, $P=0.07$), younger at menarche (by 0.5 years, $P=0.03$), and older at first childbirth (by 2.1 years, $P<0.001$) as well as last childbirth (by 1.6 years, $P=0.005$), and had lower plasma 17β-

estradiol levels (by 10.3 pg/ml, P=0.04) compared to nonsecretors. Table 2.2 also shows that, compared to nulliparous women, parous women were more likely to secrete NAF (by 20%, P=0.04), and that parous women who had breastfed their babies for longer than 3 months were more likely to be secretors compared to those who had never breastfed or breastfed for 3 months or less (by about 20%, P=0.02).

There were no significant differences between the two groups in the distribution of ethnicity, anthropometric measurements, plasma levels of testosterone or progesterone, serum levels of lipids, T3, T4 or TSH, or dietary intake of calories, proteins, fats, or carbohydrates. Among parous women, no difference was found between secretors and non-secretors in the number of childbirths.

Many of the variables examined correlated with one another in our study. For example, lactose intake correlated significantly with age at first childbirth (r=0.22, P=0.001), age at last childbirth (r=0.16, P=0.02), and length of breastfeeding (r=0.27, P<0.01). Age at last childbirth also correlated significantly with age (r=0.22, P=0.01) and age at first childbirth (r=0.52, P<0.001).

2.3.2 Logistic regression models to predict secretor status

Logistic regression models were constructed to predict NAF secretor status, as shown in Table 2.3. Model 1 included lactose intake only, which was a significant factor for being a secretor (OR=2.29, CI: 1.35-3.90). In model 2, the effect of lactose intake was adjusted for intake of total calories, proteins, fats, and carbohydrates. High lactose intake remained an independent predictor for being a secretor of NAF (OR=2.59, CI: 1.44-4.65). Intake of total calories, proteins, fats, and carbohydrates were not predictors. In model 3A, the effect of lactose was adjusted for demographic and reproductive variables that have previously been reported to be associated with secretor status [17], including age, ethnicity, age of menarche, and parity. Higher lactose intake was still a strong predictor for being a secretor (OR=2.70, CI: 1.52-4.80), independent of age (OR=1.12, CI: 1.00-

Table 2.2. Characteristics of the study population, by secretor status

A. Continuous variables	All subjects (N=238)	Non-secretors (N=82)	Secretors (N=156)	P[a]
Nutrients and food groups[b]				
Lactose (g)	7.6 ± 6.7	5.7 ± 4.6	8.6 ± 7.4	<0.001
Total caloric intake (kcal)	1751.4 ± 520.7	1736.1 ± 616.2	1759.4 ± 464.7	0.74
Total protein intake (g)	67.7 ± 21.5	67.8 ± 23.3	67.6 ± 20.6	0.94
Total fat intake (g)	75.3 ± 26.7	75.4 ± 29.2	75.3 ± 25.4	0.99
Total carbohydrate intake (g)	200.9 ± 72.5	196.9 ± 85.9	203.0 ± 64.6	0.54
Milk (g)	71.0 ± 122.3	48.9 ± 82.3	86.7 ± 142.9	0.01
Age and reproductive history				
Age (yr)	36.1 ± 2.7	35.7 ± 2.7	36.4 ± 2.6	0.07
Age of menarche (yr)	12.6 ± 1.6	13.0 ± 1.6	12.5 ± 1.5	0.03
Number of childbirths[c]	2.6 ± 1.1	2.5 ± 1.1	2.6 ± 1.2	0.92
Age at first childbirth[c]	23.0 ± 5.0	21.6 ± 4.0	23.7 ± 5.3	<0.001
Age at last childbirth[c]	29.4 ± 4.5	28.3 ± 4.8	29.9 ± 4.3	0.02
Hormones				
Testosterone (ng/ml)	0.8 ± 0.9	0.8 ± 0.6	0.8 ± 1.0	0.90
17β-Estradiol (pg/ml)[d]	78.2 ± 34.2	85.0 ± 39.7	74.7 ± 30.7	0.04
Progesterone (ng/ml)	10.8 ± 5.5	10.5 ± 5.6	10.9 ± 5.5	0.56
T3 (ng/dl)	124.8 ± 36.3	126.8 ± 35.3	123.7 ± 37.0	0.55
T4 (μg/dl)	8.3 ± 1.4	8.2 ± 1.5	8.3 ± 1.4	0.68
TSH (μIU/ml)	2.4 ± 5.4	2.3 ± 1.8	2.4 ± 6.6	0.82
Anthropometrics				
Height (cm)	161.8 ± 6.8	162.2 ± 6.3	161.6 ± 7.0	0.52
Weight (kg)	74.7 ± 14.8	76.8 ± 16.3	73.6 ± 13.9	0.12
BMI (kg/m^2)	28.6 ± 5.5	29.2 ± 6.1	28.3 ± 5.3	0.21
Body fat mass (kg)	27.9 ± 9.7	29.2 ± 10.5	27.2 ± 9.3	0.16
Body lean mass (kg)	46.5 ± 6.2	46.8 ± 6.2	46.2 ± 6.2	0.50
B. Categorical variables	**All subjects**	**Secretors**	**% of Secretors**	**P[e]**
Race/ethnicity				
Caucasian	122	88	72.1	0.19
Hispanic	72	46	63.9	
Black	35	18	51.4	
Asian	3	2	66.7	
Other	6	3	50	
Parity				
No	27	13	48.2	0.04
Yes	211	143	68.2	
Length of breastfeeding[f]				
Never	46	31	67.4	0.02
3 months or less	35	21	60	
More than 3 months	95	78	82.1	

[a]P value for two group t-test ;
[b]Data obtained from food records;
[c]For parous subjects only, N=211;
[d]N=209, with 139 secretors (67%) and 70 non-secretors (33%);
[e]P value for χ^2 test;
[f]A total of 176 parous women had data for length of breastfeeding; 130 secretors (73.9%) and 46 non-secretors (26.1%).

1.24), earlier age of menarche (OR=0.79, CI: 0.65-0.95) and being parous (OR=2.29, CI: 0.95-5.52). When milk consumption was entered in model 3A to substitute for lactose intake (data not shown in Table 2.3), milk was also a strong predictor for secretor status (OR=1.37, CI: 1.02-1.84, increment of 100 g), that was independent of age (OR=1.12, CI: 0.99-1.22), earlier age of menarche (OR=0.79, CI: 0.65-0.96) and being parous (OR=2.26, CI: 0.94-5.44). Model 3B was restricted to parous women only and included all variables shown in model 3A plus the variable "age at first childbirth". Age at first childbirth (OR=1.48, CI: 1.04-2.10) remained a strong predictor for secretor status in parous women, independent of lactose intake (OR=2.62, CI: 1.40-4.92), age (OR=1.10, CI: 0.97-1.23), and age at menarche (OR=0.79, CI: 0.64-0.97). Age at last childbirth was not entered in the model because of its strong correlation with "age at first childbirth". If age at last childbirth replaced age at first childbirth in the model, age at last childbirth is not an independent predictor (OR=1.28, CI: 0.88-1.86, P=0.18, increment of 5 years), while the effects of lactose (OR=2.69, CI: 1.41-5.15, P=0.003) and age at menarche (OR=0.80, CI: 0.65-0.98, P=0.04) remained strong and independent predictors of the secretor status. After mutual adjustment, an increment of 10 g in lactose intake and being parous both increased the odds of being a secretor by about 2 fold. An increment of 1 year in age of menarche decreased the odds of being a secretor by 0.79 fold. For parous women, in addition to the above factors, an increment of 5 years in age at first childbirth increased the odds of being a secretor by 1.5 fold.

In univariate analysis, secretors also had significantly lower levels of plasma estradiol. Data for estradiol levels were missing from 29 subjects, because the kit used for measuring this hormone was discontinued by the manufacturer during the middle of our analyses. Multi-variate analyses were performed on the subset of subjects with estradiol levels (N=209). When demographic and reproductive variables (age, ethnicity, age of menarche, and parity) were included in the model, the independent effect of estradiol was not significant (OR=0.99, CI: 0.99-1.00, P=0.21), while high lactose intake (OR=3.43, CI: 1.66-7.08, P<0.001), earlier age of menarche (OR=0.80, CI: 0.65-0.99, P=0.03), and

Table 2.3. Multi-level logistic regression models to predict secretor of NAF (N=238)

Variables[1]	Models[2] Odds Ratio (95% Confidence Interval)			
	Model 1	**Model 2**	**Model 3A**	**Model 3B**
Lactose intake	2.3 (1.4, 3.9)	2.6 (1.4, 4.7)	2.7 (1.5, 4.8)	2.6 (1.4, 4.9)
Total caloric intake	NE[3]	1.2 (0.6, 2.2)	NE	NE
Protein intake	NE	0.9 (0.6, 1.2)	NE	NE
Fat intake	NE	0.9 (0.5, 1.6)	NE	NE
Carbohydrate intake	NE	0.9 (0.7, 1.2)	NE	NE
Age	NE	NE	1.1 (1.0, 1.2)	1.1 (1.0, 1.2)
Ethnicity	NE	NE	0.6 (0.3, 1.3)	0.9 (0.4, 1.8)
Menarche	NE	NE	0.8 (0.7, 1.0)	0.8 (0.6, 1.0)
Parity	NE	NE	2.3 (1.0, 5.6)	NE
Age at first childbirth	NE	NE	NE	1.5 (1.0, 2.1)

[1]Lactose, increment of 10 g; total caloric intake, increment of 100 kcal; protein intake, increment of 10 g; fat intake, increment of 10 g; carbohydrate intake, increment of 10 g; age at first childbirth, increment of 5 years; parity, yes vs. no.

[2]Model 1: the effect of lactose was estimated. Model 2: the effect of lactose was adjusted for other nutrients including intake of total calories, proteins, fats and carbohydrates. Model 3A: the effect of lactose was adjusted for demographic and reproductive variables, including age, ethnicity, age of menarche, and parity. Model 3B: for parous women only (N=211), the effect of lactose was adjusted for demographic and reproductive variables, including age, ethnicity, age of menarche, and age at first childbirth.

[3]NE, not entered

being parous (OR=2.54, CI: 0.99-6.52, P=0.05) remained strong and independent predictors for being secretors of NAF.

Data for length of breastfeeding were missing for 35 parous women, but univariate analysis based on the data available (N=176) showed that women who breastfed their babies for longer periods of time (>3 months) were more likely to be secretors. Multivariate analyses were also performed on this subset of subjects. When demographic and reproductive variables (age, ethnicity, age of menarche, and age at first childbirth) were included in the model, the independent effect of breastfeeding was not significant (OR=1.24, CI: 0.47-3.28, P=0.13, breastfed for over 3 months vs. never breastfed), while high lactose intake (OR=2.86, CI: 1.23-6.63, P=0.008), earlier age of menarche (OR=0.72, CI: 0.55-0.95, P=0.02), and older age at first childbirth (OR=2.95, CI: 1.20-7.29, P=0.02, increment of 5 years) remained independent predictors for being secretors of NAF.

The characteristics of the subjects with missing data (N=27 for estradiol levels and N=35 for breastfeeding data) did not differ from the remaining subjects included in the subsets of models by anthropometric, reproductive, nutritional, or hormonal levels (results not shown). Therefore, the outcome of the above analyses is not expected to be significantly affected by the missing data.

2.3.3 Association between volume of NAF and lactose intake

Since lactose has osmotic properties that may affect the volume of breast secretion, and therefore, secretor status, a trend analysis was performed between lactose intake and NAF volumes among secretors only. The fluid volumes among the secretors were categorized as low (<10 μl), medium (10 to 30 μl), and high (>30 μl). As shown in Table 2.4, lactose intake was also marginally associated with volume of NAF in secretors (P trend=0.08)

Table 2.4. Average lactose intake among secretors (mean ± SD), by volume of NAF

Volume of NAF (μl)	Lactose intake (g/day)
Low	7.6 ± 5.9
Medium	8.3 ± 5.6
High	9.9 ± 9.8
P for trend	0.08

2.4 Discussion

In this study, we found that higher dietary intake of lactose and milk, earlier age of menarche, being parous, and older age at first childbirth are all independent and positive predictors for being secretors of NAF. Moreover, among secretors lactose intake is positively associated with fluid volumes recoverable by nipple aspiration. These data suggest that the ability to secrete fluid and the fluid volume are under the regulation of multiple factors. We suggest that reproductive factors and the osmotic effect of lactose intake both play important roles, as discussed below.

The reproductive factors that we found to be associated with secretor status of NAF are consistent with those reported by others. Several studies have attempted to link NAF secretor status with breast cancer risk [9,29] but with mixed results. This is not surprising based on our observations that some of the reproductive factors associated with being secretors increase whereas others decrease the risk for breast cancer. For example, earlier age of menarche and older age at first childbirth increase breast cancer risk and the odds of being a secretor. However, being parous, having breastfed for longer periods of time, and having lower plasma estradiol levels decrease the risk for breast cancer but increase the odds of being a secretor. Therefore, we hypothesize that the secretor status may not be related to breast cancer risk. This hypothesis deserves confirmation by additional studies, because for NAF to be a valid material for early diagnosis of breast cancer, any effort to increase NAF yield should not alter the risk of breast cancer.

Age was included in all multivariate models as a co-variate, and older age was found to be marginally associated with being a secretor (OR=1.1, P=0.05 to 0.13). However, due to the narrow age range of our study population, this observation may have limited inference. Being parous and of older age at first childbirth increase the odds of being a secretor. Women in our study were all between 30 to 40 years old and had confirmed ongoing menses, and therefore, were presumably capable of reproduction. Thus, those who were older rather than younger at first childbirth in our cohort were more likely to have had a recent childbirth at the time of NAF samplings. These data suggest that pregnancy itself and time since last pregnancy are both determinants for remaining a secretor. Consistent with this hypothesis, we observed that secretors were also older at last childbirth in univariate analysis (Table 2.2).

However, in multi-variate analysis, age at last childbirth was not a significant predictor for being a secretor, probably because this variable correlated well with age (P=0.01), and lactose intake (P=0.02). Consistent with the report by Petrakis *et al.* [17], a longer period of lactation was positively associated with being a secretor of NAF in univariate analysis in our study. However, the length of lactation is not an independent predictor after adjusting for other variables, such as lactose intake, probably because women who breastfed for longer periods of time also tended to consume more lactose (P<0.01). Therefore, we suggest that some biochemical changes induced by pregnancy persist even beyond lactation to influence secretor status, as discussed below.

In our study, the effect of lactose on secretor status was independent of reproductive factors and intake of total calories, fats, proteins, and carbohydrates, suggesting a direct effect of lactose on breast secretion in non-lactating women. This observation was found regardless of whether lactose intake was estimated from 3-day food records or from food frequency questionnaires. Additionally, milk, the main dietary source of lactose, was also a strong independent predictor when it replaced lactose as a variable in the multi-variate analysis in Model 3A.

Lactose is a disaccharide and is not absorbed intact by the intestine. Dietary lactose is metabolized in the intestine to galactose and glucose before intestinal absorption. However, lactose can be re-synthesized in the mammary glands from two precursors, glucose and galactose, by lactose synthase. Lactose synthase is a complex of two proteins, protein A (also known as galactosyltransferase) and protein B (known as α-lactalbumin) [50]. Lactose is an important energy component in milk, but it is also an osmotic agent. Its physiological effect is to induce fluid influx/efflux, and therefore, it influences fluid volume. This is an important mechanism by which lactose increases milk volume in the lactating breast. Lactose in the non-lactating breast is expected also to have the same osmotic effect by increasing fluid volume in the breast ducts and, thus, allowing easy recovery of fluids through nipple aspiration.

In the lactating mammary gland, 50 to 80% of galactose is synthesized *de novo* from glucose by a panel of six enzymes that are highly expressed in the mammary gland during lactation [51]. In the non-lactating breast, however, the enzymes for the *de novo* synthesis of galactose from glucose, such as UDP-4-galactose-epimerase, are of much lower activity than in the lactating mammary gland [52,53]. Thus, circulating galactose from the breakdown product of dietary lactose may become a major source for lactose synthesis in the non-lactating mammary glands. This may explain the positive association between dietary intake of lactose and milk and being a secretor. Consistent with the ability of the non-lactating breast to biosynthesize lactose is our other observation that α-lactalbumin (protein B of lactose synthase) (see Chapter 3), was present abundantly in nipple aspirate fluids of 1/3 of non-lactating breasts [10], and also, that a strong correlation was found between lactose intake and volume of NAF among secretors. Therefore, it is tempting to speculate that higher levels of α-lactalbumin might be found in the mammary glands of non-lactating women with a more recent history of pregnancy. Consequently, higher intake of dietary lactose, a precursor for galactose, and a more recent history of pregnancy (with a greater residual amount of α-lactalbumin) are both

important but independent predictors for having a higher volume of fluid in the breast ducts and being a secretor.

Other dietary factors that have been associated with being a secretor of NAF include higher consumption of dietary fats, as reported by Lee *et al.* [19]. However, we could not confirm this association in our study. In addition, Petrakis *et al.* [28] found no differences in serum concentrations of 17β-estradiol between secretors and non-secretors of NAF. In our study cohort, a lower plasma level of estradiol was observed in secretors, although the independent effect of estradiol was not significant after adjusting for other factors. None of the sex steroids (estradiol, progesterone and testosterone) that we examined in this study are independent predictors of being a secretor. Others have reported that prolactin is not a predictor of being a secretor. Therefore, the hormones that influence ability to secrete NAF remain to be determined.

The strength of this study is a well-defined population with a variety of variables, which makes it possible to investigate effects of dietary factors in multivariate-adjusted analyses. Dietary intakes were measured using food records from 24 hr recalls allowing more precise estimate of nutrient intake than with a food frequency questionnaire. Blood was collected only during the luteal phase of a menstrual cycle. This timed blood sampling allowed us to more precisely estimate the influence of two female hormones, estradiol and progesterone, on secretor status. However, a limitation of our study is that we excluded women who had abnormal mammograms, were outside the age range of 30 to 40 years old (including postmenopausal women), or were on contraceptive medications. Therefore, our study findings may have limited inference to these other women.

In summary, our study indicated that dietary intake of lactose, in the presence of residual factor(s) from pregnancy, has a strong influence on secretor status in 30 to 40 year old women. Thus, diet can influence the secretory activity of the breast, and NAF might be a suitable material for probing patho-physiologic changes in the breast. Modulation of

secretor status by diet may have important implications for breast cancer biomarker discoveries.

CHAPTER 3: MAJOR PROTEIN PATTERNS OF NAF AND FACTORS ASSOCIATED WITH THESE PATTERNS

3.1 Introduction

Proteins, lipids, steroid hormones, growth factors, and carbohydrates have been found in NAF at concentrations much higher than in plasma [7]. For example, the average level of proteins in NAF is 71-170 mg/ml, which is much higher than the 6-8 mg/ml reported for plasma [23,24]. Several proteins, including hormone-regulated proteins and cancer markers such as prostate-specific antigen and carcinoembryonic antigen, have been found in NAF using specific antibodies [44,54]. Sanchez *et al.* [55] described two distinct types of NAF proteomes based on the profiles of highly abundant proteins. Type I NAF is enriched in many of the same proteins found in gross cystic disease fluid, including zinc α2-glycoprotein (ZAG), apolipoprotein D (apoD), and gross cystic disease fluid protein-15 (GCDFP-15, also known as prolactin-inducible protein). Type II NAF is characterized by the great abundance of milk-associated proteins, including lactoferrin, α-lactalbumin, and lysozyme C.

Several proteins that are highly abundant in type I NAF have also been found in other human body fluids, such as saliva, sweat, and nasal lavage fluid [56-58]. Higher levels of ZAG, apoD, and GCDFP-15 in serum and other body fluids have been associated with increased apocrine activity [59]. The expression of type I NAF proteins has been studied in cultures of malignant mammary epithelial cells. In these cultures, androgens (dihydrotestosterone or fluoxymesterone) stimulated the secretion of ZAG, apoD, and GCDFP-15 into the culture media [60-63], progesterone and prolactin up-regulated the expression of GCDFP-15 [64,65], and estradiol decreased the synthesis of GCDFP-15 and apoD [61-63]. In contrast, Zhou *et al.* showed that testosterone, progesterone, and tamoxifen decreased apoD mRNA expression in the mammary glands of ovariectomized rhesus monkeys [66].

Regulation of the synthesis of milk proteins by hormones during pregnancy and lactation is complex. Prolactin is believed to play a primary role in stimulating the expression of milk proteins [67]. Several type II NAF proteins, such as lactoferrin, lysozyme and lactalbumin, are inducible by prolactin, and also by insulin and corticosterone, in cultures of breast cells and in the mammary glands of laboratory animals [68]. In mice, estrogen up-regulated the expression of lactoferrin in the uterus, but not in the mammary glands [69].

Limited studies have shown that both types of NAF are found in healthy women and in breast cancer patients, but with different prevalence. Sanchez *et al.* found type II NAF in only 10% of women without breast pathology or with benign breast disease, but in slightly more than half of women with breast cancer [55]. However, more recent studies from the same group showed that cancer patients with type II NAF generally had a better prognosis than those with type I [70]. Therefore, whether these protein profiles are predictors for breast cancer risk and prognosis deserve more study. In addition, little is known, about the mechanisms and factors affecting the secretion of these proteins in women. This study investigated the demographic, reproductive, anthropometric, hormonal, and nutritional factors that are associated with the secretion of these two types of NAF. This preliminary study focused on the prevalence of these two types of NAF in healthy women, with the goal of providing for a better understanding of the association of these proteins with breast cancer.

3.2 Materials and Methods

3.2.1 Study Design

The study population for accomplishing this specific aim has been described in Chapter 2: see Section 2.2.1 for the Study Design.

3.2.2 Measurements of anthropometrics, hormones, reproductive variables, lipids and nutrients

See Section 2.2.3 of Chapter 2 for measurements of these variables.

3.2.3 Protein profiling of NAF

One dimensional (1D) SDS-PAGE (12.5 %, 1.0 mm, 200 Volts, 1 hr) was used to separate abundant NAF proteins (30 μg) for the classification of NAF types. After staining with Coomassie blue (Biosafe Coomassie, Biorad, Hercules, CA), visible bands were excised and trypsin-digested (15 μg/ml of trypsin) at 37°C overnight. Mass spectra of peptide digests were obtained using a matrix-assisted laser desorption/ionization-time of flight mass spectrometer (MALDI-TOF MS) (Applied Biosystems, Foster City, CA). Proteins were identified by peptide mass fingerprinting with a search against the National Center for Biology Information (NCBI) protein database using the Profound algorithm. Positive protein identifications were accepted for those with scores above the 95% confidence level.

Four of sixteen breast milk samples and four of forty-two type II NAF samples from age-matched women were randomly chosen for comparison of protein compositions. Human milk samples (N=16) were obtained and donated by the Mother's Milk Bank at Austin (Austin, TX) and separated on 1D gels. After staining with Coomassie blue, the gel (randomly chosen from breast milk and type II NAF) was scanned using the AlphaImager^{TM} 2200 (Alpha Innotech Co., San Leandro, CA**)**, and density and percentage of total density for each individual band were measured using the Alpha Ease FC software. The major bands were excised, digested, and subjected to protein identification by MALDI-TOF MS. The mean percentage staining density for the major proteins, including lactoferrin, albumin, β-casein, and lactalbumin, were compared between type II NAF and human breast milk samples.

3.2.4 Analyses of secretion types on different menstrual cycle days

To determine whether secretion types of NAF change with the phases of menstrual cycles, NAF and fasting blood samples were obtained from a total of 7 healthy women, who were 20-30 years old and had regular menstrual cycles, every other day for a full menstrual cycle (from menstrual cycle day 3 to the next menstrual cycle day 1). Women were excluded if they were on contraceptive medications during the 6 months preceding enrollment, or if they were vegetarians, pregnant, or nursing. Serum samples were analyzed for levels of progesterone and 17β-estradiol using commercial radioactive immunoassay kits (Diagnostics Labs, Webster, TX), as described in Section 2.2.3 of Chapter 2. NAF samples from three women on three different cycle days, representing early follicular phase with low levels of estradiol and progesterone (generally cycle days 3 to 5), peak levels of estradiol prior to LH surge, and the mid-luteal phase with levels of both estradiol and progesterone (generally around cycle day 22) were selected and analyzed for major protein patterns using 1D SDS-PAGE, as detailed in Section 3.2.3.

3.2.5 Statistical analysis

Means and standard deviations were computed for continuous data, and percentages were computed for categorical data. For univariate analyses, the χ^2 test was performed for categorical variables, and the Student's *t*-test was used for continuous variables for comparisons between groups with type I and type II NAF. Variables associated with NAF protein profiles by univariate analysis (using $P<0.10$ as a cut-off) were further entered into age-adjusted multi-level logistic regression models, allowing the identification of independent predictors. For the multivariate logistic regression models, a new ordinal variable was created to code "years since last childbirth". Specifically, the first quartile (0 to 3.3 yrs) was coded as "0", second quartile (3.3 to 7.2 yrs) as "1", third quartile (7.2 to 10.5 yrs) as "2", the 4th quartile (over 10.5 yrs) as "3", and nulliparous women as "4". The linearity of the effect of the new variable was assessed and confirmed. Stepwise selection was used to determine predictors associated with secretion types of NAF. Variables with a two-tailed P value of <0.05 and two covariates, including age and

ethnicity, were kept in the final models. Correlations among parameters of anthropometrics and among demographic and reproductive variables were tested using Pearson correlation analysis. All statistical analyses were performed using SAS® (Version 9.1, SAS Institute Inc., Cary, NC).

3.3 Results

3.3.1 Identification of NAF proteins and classification of NAF types

The abundant proteins in NAF that were visible in a 1D gel after Coomassie blue staining (Figure 1.2) were subjected to identification by peptide mass fingerprinting [71]. Two protein bands, albumin (abbreviated as Alb in Figure 1.2, 67 kDa) and Ig heavy chain (Ig-H, 59 kDa), were found in all NAF samples. ZAG (41 kDa), apoD (24 kDa), and GCDFP-15 (17 kDa) were typically found and were more abundant in type I NAF, while lactoferrin (Lf, 78 kDa), lysozyme C (Lyz, 15 kDa), and α-lactalbumin (Lalb, 14 kDa) were typically found and were more enriched in type II NAF, as shown in Figure 1.2. β-Casein (25 kDa, multiple bands in Figure 1.2) was also present but with varying abundance in type II NAFs. Based on the gel mobility, the molecular weights (shown in parentheses) of these NAF proteins were consistent with published data. Type I NAFs were found in 58%, and type II in 33% of the women in our study (see Table 3.1 below). A small number of women (9%) had a protein profile that was a mixture of types I and II, showing the presence of lactoferrin, ZAG, apoD, GCDFP-15, and lactalbumin (mixed type in Figure 1.2).

Figure 1.2 also shows the 1D-gel profile of breast milk samples for comparison with type II NAF. Breast milk samples contained lactoferrin, albumin, casein, lysozyme and lactalbumin. Although type II NAF contains mostly milk-associated proteins, their relative proportions in breast milk and type II NAF differed. As shown in Figure 1.2, breast milk generally had a higher content of casein proteins (29% ± 7.0% of the total protein-staining-density for milk vs. 2.3% ± 4.6% for type II NAF, $P<0.05$, N=4 per

group), whereas type II NAF had more albumin (9.3% ± 1.6% of the total density for milk vs. 24.7% ± 10.0% for type II NAF, $P<0.05$, N=4 per group). There were no differences in the mean percentage densities of lactoferrin and lactalbumin.

3.3.2 Prevalence of the two protein profiles in the study population

Nipple aspiration was attempted on three separate study visits from a total of 219 subjects. Women were classified as secretors if fluid was obtained at least once after three attempts. Of the 219 subjects, 148 (68%) were secretors, and 128 of these samples were large enough for protein profiling. Our success rate for obtaining NAF is consistent with the rate reported by others [23,30].

Table 3.1 shows the distribution of NAF protein profiles in our subjects (N=128). Type I NAF was more common than type II (ratio 1.8 to 1). Mixed type NAF (containing both type I and II signature proteins) was found in less than 10% of the subjects. These subjects were not included in *t*-test and multivariate analyses due to the small sample size. Among women who had given birth within the last four years, 63% showed type II NAF (Table 3.1).

Table 3.1. The frequency and %-distribution of different protein profiles in our study population by status of recent childbirth

Secretion type	All subjects	Childbirth within the last 4 years	
		Yes	No
Type I	74 (58%)	9 (24%)	65 (72%)
Type II	42 (33%)	24 (63%)	18 (20%)
Mixed type	12 (9%)	5 (13%)	7 (8%)
Total	128 (100%)	38 (100%)	90 (100%)

3.3.3 Characteristics of the study population by NAF secretion types

Table 3.2 describes demographic, reproductive, nutritional, and hormonal characteristics of our study subjects, as related to secretion types. Ethnicity did not differ significantly

between the type I and type II groups. There were no differences in the distribution of secretion types between those who were past users and those who had never used contraceptive medications. Women with type II NAF had a significantly higher incidence of breastfeeding compared to those with type I NAF (by 29.8%, P=0.002).

Women with type I NAF were older in age compared to women with type II (by 1.2 years, P=0.01), were slightly younger at first childbirth (by 1.8 years, P=0.08), and had a lower number of completed pregnancies (2.1 compared to 2.6, P=0.04), and a longer time since last childbirth (by 3.9 years, P<0.0001). Women with type I NAF were also heavier (by 5.8 kg, P=0.03), had higher BMIs (by 2.1 kg/m^2, P=0.02) , more %-body fat (by 3.1%, P=0.02), more total body fat mass (by 4.9 kg, P=0.008), more total body lean mass (by 2.1 kg, P=0.09), higher blood LDL levels (by 11.4 mg/dl, P=0.05), and higher plasma 17β-estradiol levels (by 10.3 pg/ml, P=0.08), and they consumed more calories from saturated fat (by 0.9%, P=0.09). Characteristics such as age and years since last childbirth for women with a mixed NAF type had values that fell between those of women with type I and type II NAF (i.e. age, age at first childbirth, history of breastfeeding, parity, and years since last childbirth) (Table 3.2). Years since last childbirth was the only significant predictor for secretion types (P=0.0008). Age at first childbirth, history of breastfeeding, and parity were not predictors. Model 3 included variables in model 2 plus dietary fat intake. Percentage of caloric intake from saturated fat (P=0.03) was another significant predictor for secretion types, in addition to years since last childbirth (P=0.008) and total body fat mass (P=0.002). Model 4 included the variables in model 3 plus estradiol and blood LDL levels, neither of which was associated with secretion types. Estradiol level as a continuous variable was not a significant predictor of secretion types in model 4 (P=0.11). However, when estradiol levels were divided into quartiles and coded as a categorical variable, a significantly higher percentage of type I NAF was found among women in the highest quartile (estradiol level >97.5 pg/ml) compared to those in the lowest quartile (<55.7 pg/ml), after adjustment for body fat mass, years since last complete pregnancy, and %-caloric intake from saturated fat (OR=5.9, 95% CI from 1.3 to 27.7, P<0.01). In summary, results shown in Table 3.3 indicate that incremental

Table 3.2. Characteristics (mean ± SD, unless otherwise specified) of our study population by secretion types

Variables	All subjects N=128	Type I N=74	Type II N=42	Mixed type N=12	P[a]
Demographics					
Age	36.5 ± 2.5	36.9 ± 2.4	35.7 ± 2.3	36.0 ± 3.2	0.01
Race/Ethnicity, column percentage					
Caucasian	58.3	57.5	57.1	66.7	0.21
Hispanic	28.4	27.4	31	25	
African-American	10.2	13.7	4.8	8.3	
Asian	1.6	0	4.8	0	
Other	1.6	1.4	2.4	0	
Reproductive history					
Contraceptive usage, column percentage					
Never	27	28.4	24.4	27.3	0.64
Past	73	71.6	75.6	72.7	
Breastfeeding, column percentage					
No (including nulliparous subjects)	28.7	42.3	12.5	0	0.002
Yes	71.3	57.7	87.5	100	
Age of menarche	12.5 ± 1.6	12.5 ± 1.6	12.5 ± 1.5	11.8 ± 1.3	0.88
Parity	2.3 ± 1.3	2.1 ± 1.3	2.6 ± 1.4	2.2 ± 1.1	0.04
Age at first childbirth (parous subjects	24.2 ± 5.2	23.6 ± 4.8	25.4 ± 5.5	23.9 ± 6.0	0.08
Years since last childbirth (parous only)	7.2 ± 4.8	8.8 ± 4.7	4.9 ± 4.1	6.4 ± 4.1	<0.001
Anthropometrics					
Weight (kg)	72.5 ± 14.1	75.6 ± 13.4	69.8 ± 15.0	62.0 ± 7.8	0.03
Height (cm)	161.5 ± 7.2	161.8 ± 7.6	161.6 ± 7.3	159.4 ± 4.7	0.91
BMI (kg/m^2)	27.8 ± 5.3	28.9 ± 5.3	26.8 ± 5.2	24.5 ± 3.4	0.02
%-Body fat	35.5 ± 6.7	37.1 ± 6.2	34.0 ± 6.8	31.1 ± 6.6	0.02
Total fat mass (kg)	26.3 ± 9.4	28.7 ± 9.0	23.8 ± 9.5	19.6 ± 5.8	0.008
Total lean mass (kg)	45.8 ± 6.3	47.0 ± 6.2	44.9 ± 6.8	42.4 ± 3.5	0.09
Nutrient intake					
Total calorie (kcal)	1747± 457	1754 ± 477	1694 ± 403	1879 ± 511	0.5
Total fat (g)	74.9 ± 24.9	76.7 ± 24.8	71.6 ± 26.7	75.3 ± 20.0	0.3
%-Calorie from fat	38.0 ± 6.6	38.8 ± 6.5	37.2 ± 7.0	35.8 ± 5.9	0.24
Saturated fat (g)	24.9 ± 8.3	25.9 ± 8.9	23.4 ± 7.1	24.0 ± 8.0	0.13
%-Calorie from saturated fat	12.7 ± 2.8	13.1 ± 2.9	12.2 ± 2.3	11.3 ± 2.6	0.09
Monounsaturated fat (g)	29.2 ± 10.1	29.8 ± 9.4	28.2 ± 11.7	29.1 ± 7.3	0.41
%-Calorie from monounsaturated fat	14.8 ± 2.9	15.1 ± 2.7	14.5 ± 3.4	14.2 ± 2.1	0.31
Polyunsaturated fat (g)	14.8 ± 6.9	15.2 ± 7.1	14.0 ± 7.3	15.0 ± 4.5	0.36
%-Calorie from polyunsaturated fat	7.4 ± 2.2	7.5 ± 2.3	7.1 ± 2.3	7.3 ± 1.8	0.38
Hormones					
Testosterone (ng/ml)	0.71 ± 0.38	0.73 ± 0.42	0.59 ± 0.35	0.65 ± 0.24	0.41
17β-Estradiol (pg/ml)	75.7 ± 31.8	77.9 ± 32.7	67.6 ± 24.0	92.3 ± 23.3	0.08
Progesterone (ng/ml)	11.0 ± 5.6	10.8 ± 5.5	11.4 ± 4.9	10.6 ± 7.8	0.61
T3 (ng/dl)	121.2 ± 36.1	119.7 ± 32.7	123.7 ± 41.6	101.5 ± 22.2	0.59
T4 (μg/dl)	8.3 ± 1.5	8.2 ± 1.3	8.5 ± 1.9	7.3 ± 0.6	0.57
TSH (μIU/ml)	1.9 ± 1.1	1.9 ± 1.1	2.0 ± 1.3	1.8 ± 1.3	0.15
Lipids					
VLDL (mg/dl)	20.6 ± 12.6	21.2 ± 13.7	20.1 ± 11.6	18.9 ± 7.5	0.67
LDL (mg/dl)	107.1 ± 29.6	111.0 ± 29.2	99.6 ± 32.0	107.3 ± 18.5	0.05
HDL (mg/dl)	53.5 ± 10.9	52.2 ± 9.7	55 ± 12.9	57.4 ± 9.8	0.18
TRIG (mg/dl)	103.1 ± 62.7	105.9 ± 68.3	100.5 ± 58.2	94.6 ± 37.5	0.67
CHOL (mg/dl)	181.1 ± 33.0	184.2 ± 33.7	174.7 ± 34.5	183.6 ± 19.7	0.15

[a]**all comparisons are between women with type I and type II NAF profiles**

Table 3.3. Multi-level logistic regression models for predicting type I secretors of NAF

Models[a]	Variables[b]			
	Age	**Years since last childbirth**	**Total fat mass**	**%-Calorie from saturated fat**
	OR (95% CI)	**OR (95% CI)**	**OR (95% CI)**	**OR (95% CI)**
Model 1 (N=104) Reproductive	1.1 (0.9, 1.4)	2.8 (1.5, 5.1)		
Model 2 (N=99) Anthropometrics	1.0 (0.8, 1.3)	2.9 (1.4, 5.8)	2.9 (1.6, 5.5)	
Model 3 (N=98) Nutrients	1.1 (0.9, 1.4)	2.6 (1.3, 5.1)	3.2 (1.6, 6.3)	4.2 (1.2, 15.1)
Model 4 (N=96, final) Hormones and Lipids	1.1 (0.9, 1.4)	2.6 (1.3, 5.2)	3.0 (1.5, 6.1)	4.1 (1.1, 14.6)

[a]Model 1: reproductive variables entered in the model included parity, breastfeeding, age at first childbirth, and years since last childbirth; Model 2: model 1 plus anthropometric parameters including weight, BMI, %-body fat, total fat mass and total lean mass; Model 3: model 2 plus %-calories from saturated fat; Model 4: model 3 plus estradiol and LDL.

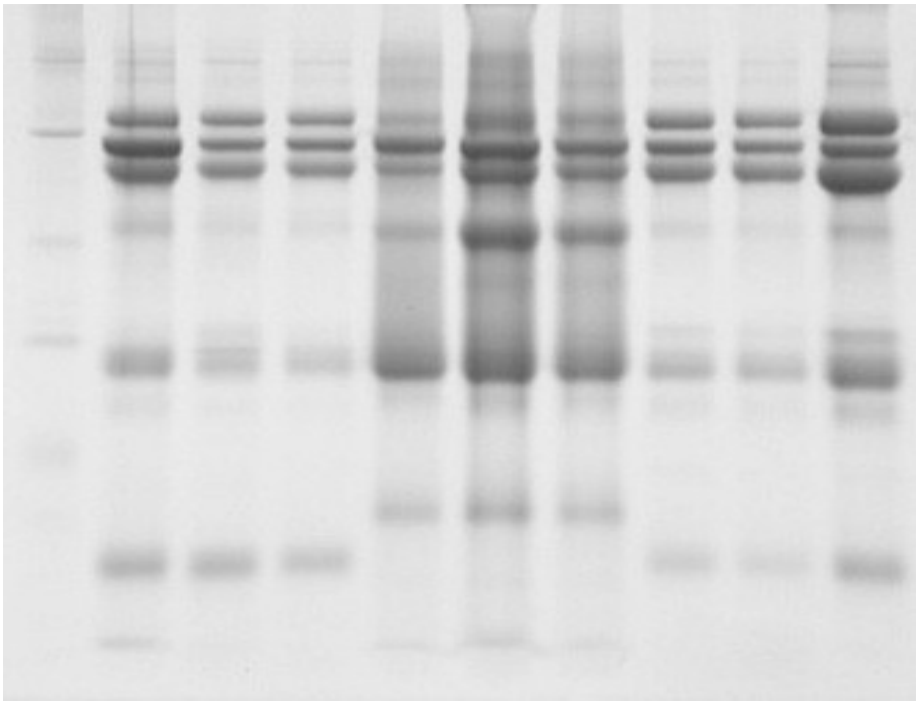
[b]Age, included in all the models; Years since last childbirth, ordinal data; Total fat mass, increments of 10 kg; %-Calories from saturated fat: increments of 5%.

changes of about 3 years since last childbirth (OR=2.6), every 10 kg in total body fat mass (OR=3.0), and every 5% in caloric intake from saturated fat (OR=4.1) increased the odds of being a type I secretor.

It is noteworthy that there were strong correlations among some variables entered into the logistic regression models. Anthropometric variables were positively correlated with one another, with r values ranging from 0.22 to 0.93 (Pearson correlation, $P<0.05$). There were also strong correlations between parity and age at first childbirth ($r=-0.47$, $P<0.001$), and between age at first childbirth and years since last childbirth ($r=-0.52$, $P<0.001$). Age was also positively correlated with the variable "years since last childbirth" in the age range of our study subjects ($r=0.47$, $P<0.0001$), which was significant in each of the regression models. This may be the reason why age was not an independent predictor in multiple logistic regression models. In addition, history of breastfeeding was significantly associated with secretion type in univariate analysis, but was not a predictor that was independent of years since last childbirth in model 1. However, if the variable "years since last childbirth" was not included in the models, breastfeeding (OR=16.0; CI: 3.0-84.1) was an independent predictor, while the influence of total body fat mass (OR=2.2; CI: 1.1-4.4) and %-caloric intake from saturated fat (OR=5.6; 95% CI: 1.7-18.5) on secretion types remained the same as when years since last childbirth was included. This may be because breastfeeding had a strong association with the ordinal variable "years since last childbirth" in our study. Women with fewer years since last childbirth were more likely to have a past history of breastfeeding (P trend<0.0001).

3.3.5 Analyses of secretion types on different menstrual cycle days

Figure 3.1 shows the 1D gel image of 9 NAF samples obtained on three different cycle days from three women. The major protein patterns of NAF remained unchanged in different menstrual phases, suggesting that fluctuating levels of female hormones during the menstrual cycle did not affect the major pattern of NAF secretion.



1A 1B 1C **2A 2B 2C** **3A 3B 3C**

Type II NAF **Type I NAF** **Type II NAF**

Figure 3.1. 1D gel (12.5% polyacrylamide) image of NAF from three women (subjects 1, 2 & 3) on different menstrual cycle days. A, from the follicular phase with low serum estradiol and low serum progesterone; B, from the day with peak serum estradiol; and C, from the luteal phase with peak serum progesterone and high estradiol.

3.4 Discussion

Protein profiles of NAF are a potential source for the discovery of new breast cancer biomarkers. There is some limited evidence from the study of Vizoso *et al*. [55,70] that both the prevalence and the prognosis of breast cancer may be related to a woman's NAF protein profile. Vizoso *et al*. found that recent childbirth was a strong predictor of NAF protein composition [72]. Other possible determinants for protein patterns in NAF have not, to our knowledge, been examined. We studied a variety of potential influences, such as reproductive history, hormone levels, diet, and body composition. We found that body fat mass (not lean body mass) and %-caloric intake from saturated fat are strong predictors of type I NAF and that their effects are independent of a more recent history of childbirth.

Interestingly, other potential factors such as BMI and weight (including lean body mass) differed in women with type I and II NAF in univariate analysis, but were not independent predictors in multivariate analyses when body fat mass was a variable in the model. Of the various types of fat, only saturated fat had an influence on NAF protein profiles. Because body fat mass and fat intake had strong influences on NAF proteomes, the influence of circulating lipids, which can partly reflect fat intake, were also investigated. LDL was a weak predictor for secretion types in univariate analysis, but was not an independent predictor in multivariate analyses. Indices of thyroid function, e.g. T3, T4, and TSH, were not associated with NAF types, consistent with results previously reported by Vizoso *et al.* [73]. It was the total body fat mass, not body fat as a percentage of body weight, that independently predicted secretion types in multivariate analyses. Total lean body mass is also not a predictor of secretion type. These results suggest that dietary fat or fat stores have an influence on NAF protein profiles. Adipose tissue has recently been viewed as an endocrine organ, capable of synthesizing and secreting hormones (such as androgens, estrogens, and leptin), and cytokines (such as interleukin-6 and tumor necrosis factor-α) [74-77]. These substances may act in a paracrine/autocrine manner to regulate breast fluid protein.

Because testosterone and progesterone were shown to up-regulate and estradiol to down-regulate ZAG, apoD and GCDFP-15 in breast cancer cell cultures [60-65], we hypothesized that levels of these hormones might be significant predictors of NAF protein profiles. However, no differences in levels of testosterone and progesterone were detected, and only estradiol was marginally higher in type I compared to type II secretors ($P=0.08$) in univariate analyses. Nevertheless, in an age-adjusted logistic regression model, a significantly higher prevalence of type I NAF was found in women within the highest quartile of estradiol levels compared to those within the lowest quartile ($P<0.01$). These data suggest a threshold effect of circulating estradiol levels on NAF types. A more detailed dose-response study is needed to establish the relationship between estradiol and type I proteins. The weak stimulatory effect of estradiol on type I proteins observed in our human study, while not consistent with *in vitro* results from breast cancer cell culture studies, is in line with *in vivo* effects in rhesus monkeys, where tamoxifen, an anti-estrogen, decreased the mRNA expression of apoD (a type I NAF protein) [66]. Considering potential autocrine/paracrine effects, it is also possible that steroid levels in NAF, which were not measured in this study due to insufficient sample volume, may be a better predictor of NAF profiles than hormone levels in serum. Additionally, no changes in the major protein patterns of NAF were observed during different phases of the menstrual cycle, suggesting that cyclic hormonal fluctuations do not change the secretion pattern of the major proteins in NAF, and that if an effect of circulating hormones does exist, the pattern may be the result of long-term exposure.

ApoD and ZAG are involved in the metabolism and disposition of fat. ApoD is a component of the HDL complex and serves as a carrier protein for transporting cholesterol and other lipids from tissues via the blood stream to the liver for further metabolism and disposition. Liu *et al.* found an interaction of apoD with leptin receptor B in hypothalamic neurons, and an association of hypothalamic apoD expression with body fat and circulating levels of leptin [78]. ZAG is also a carrier protein for various hydrophobic molecules, including lipids, and is well recognized as a lipid mobilizing

factor [79]. ApoD and ZAG may also regulate body fat accrual [78], and their secretion into the circulation may be part of an effort to remove and metabolize excess fat [80]. These data are consistent with our finding that higher body fat mass and dietary saturated fat intake were associated with elevated levels of the two type I proteins, apoD and ZAG, in NAF.

The finding that more recent childbirth was strongly associated with type II NAF ($P<0.0001$) is not surprising, since type II proteins are mostly milk-associated proteins. Our results are consistent with those of Vizoso *et al.,* who reported that type II NAF was found mostly in women who had given birth within the last four years [72]. The prevalence of type II secretors, however, was higher in our study population (33%, Table 3.1), even after excluding women who had given birth within the last four years (20%, Table 3.1), when compared to the 9% reported by Vizoso *et al.*. This might be due to differences in the selection of study subjects. Our subjects were healthy women between 30 and 40 years of age, while Vizoso and coworkers selected subjects from a hospital-based population, who were between 20 and 50 years of age. Current usage of oral contraceptives was also a predictor for type II secretion in the study of Vizoso *et al*., whereas we excluded current users of all contraceptive medications. In addition, our study showed that past usage of contraceptive medications did not affect the NAF protein profile.

The proteomes of type II NAF and breast milk are qualitatively similar, though there are quantitative differences (Figure 1.2 and Results). We found that caseins are more abundant than albumin in breast milk, whereas albumin was more abundant than caseins in type II NAF. Moreover, caseins were not always visible in type II NAF, suggesting a gradual shift from milk fluid to type II NAF after childbirth. This, coupled with the presence of a mixed type NAF in some women, leads us to speculate that type I and type II NAF may not be a constant physiological state in women. Longitudinal studies are needed to confirm this possibility. Our observation that breastfeeding *per se* is associated

with differences in NAF protein profiles, but is not independent from years since last child birth is consistent with a possible time-dependent change from type II to type I NAF.

The few studies examining proteins in NAF as potential breast cancer markers have found mixed and sometimes contradictory results. For example, GCDFP-15 has been proposed as a plasma marker of proliferative breast disease [81], and was also found to be elevated in serum of patients with metastatic breast cancer [82], but, in another study, it was reported to be decreased in the NAF of breast cancer patients [25]. This inconsistency may be simply attributable to a difference in the prevalence of type I NAF (which contains GCFDP-15) and type II NAF in the two study populations, which is modifiable by pregnancy history, body fat mass, and dietary fat intake, as shown in our study, and by current usage of birth control medications and age, as was shown in other studies [17,72]. This underscores the importance of an understanding of the presence and relevance of type I and II NAF in pathological and physiological states before breast cancer biomarkers can be effectively identified.

The major strength of our study is a well-defined study population with a variety of variables available for analysis. However, because of the narrow age range of the study subjects, who were all premenopausal, our findings may have limited inference to postmenopausal women.

In conclusion, the secretion of NAF proteins was influenced strongly by body fat mass, reproductive history, dietary fat intake, and possibly estradiol levels. Obesity is an important public health issue, and also plays a role in breast cancer risk. Further studies with larger sample sizes and longitudinal observations may help determine the relevance of NAF proteomes in estimating breast cancer prevention and prognosis.

CHAPTER 4: CHARACTERIZATION OF PROTEOMES IN NAF

4.1 Introduction

Proteins are the building blocks of cells and the major executors of cellular functions. While genomics based technologies, such as mRNA microarrays, contribute substantially to the understanding of mechanisms for cancer development, there are still limitations in the information that can be provided by this type of studies. This is because genes eventually have to be expressed through proteins, and abundance of proteins and levels of mRNA do not necessarily highly correlate [83]. Additionally, post-translational modifications of proteins such as phosphorylation, glycosylation and acetylation, which are not detectable at the mRNA level, play important roles in protein stability, localization, and functions [84]. Proteins represent more direct, accessible and relevant targets for the early detection of cancer, prediction of prognosis and evaluation of therapeutic efficacy than nucleic acid. Conventional biochemical/immunological assays have been commonly used for the analysis of proteins. However, these methods are not suitable for screening a large variety of protein molecules. Current advances in proteomic technologies, such as protein microarrays and mass spectrometry-based methods that allow the high throughput identification of proteins, have stimulated tremendous interest in characterization of protein markers relevant to cancer development.

A widely used technique for protein characterization is mass spectrometry, which is generally composed of an ion source, a mass analyzer, and an ion detector. An ion source is where molecules of interest are ionized. Commonly used ionization techniques include matrix assisted laser desorption/ionization (MALDI) and electron-spray ionization (ESI). After ionization, a mass analyzer sorts out charged particles such as proteins and peptides according to their masses, and produces a spectrum composed of peaks that are characteristic of the molecules being analyzed. Tandem mass spectrometer (MS/MS), which contains two mass analyzers, can provide further structural information of a

protein by fragmenting peptides of interest that are generated by the first mass analyzers. Typically, the first mass analyzer gives the mass to charge (m/z) patterns of peptide digests of a specific protein, which is also known as the mass fingerprinting of the protein. A selected peptide is then sent to the fragmentation chamber, and the fragmentation pattern of this peptide is then measured by the second mass analyzer. The mass difference, or distance in atomic mass units between the peaks of the second spectrum, allows each amino acid to be identified and the sequence of the selected peptide characterized. The identification of the protein is then deduced through a mass matching process, based on the spectra that are generated by these two mass analyzers.

In addition to mass spectrometry, high throughput antibody-based proteomic techniques, such as protein microarrays and tissue microarrays, have also been widely used in identifying markers that are associated with phenotypes and progression of cancer. Antibody-based techniques allow the simultaneous identification and quantitation/semi-quantitation of multiple proteins in hundreds of samples in a short period of time, through binding to the antibody probes. However, this approach is limited by the availability of antibodies with high specificity and affinity for the target. In addition, antibody-based methods require prior knowledge of proteins of interest, and therefore, can only be used in hypothesis-driven investigations [85,86].

Both antibody-based and mass spectrometry-based approaches have been utilized for the search of breast cancer biomarkers in different types of biological samples, including breast tissue, serum/plasma, ductal lavage, and nipple aspirate fluid [85-90]. In particular, proteomic analysis of nipple aspirate fluid has been intensely pursued, because the procedure for obtaining NAF is non-invasive and involves minimal risk, in comparison to ductal lavage and procedures for acquiring breast tissues by biopsy or breast surgery. Additionally, this fluid is rich in proteins that are secreted from the target tissue of breast cancer, free of contaminations by secretions from other tissues. Markers that are indicative of pathological changes of the breast will probably appear first in NAF, with

concentrations much higher than those in blood, which makes NAF an ideal source for the early detection of breast cancer.

Potential tumor markers, including CEA, PSA, c-erbB-2 (Her2/neu), bFGF, and vEGF (vascular endothelial growth hormone), have been detected in NAF using conventional antibody methods such as ELISA and western blot [91-94]. Recent advances in proteomic technologies, especially in mass spectrometry, have stimulated interest in characterizing proteomes of NAF. Interest has mostly focused on proteins of medium to low abundance, which may be more relevant as potential biomarkers for breast cancer risk. Surface Enhanced Laser Desorption/Ionization Time-of-flight Mass Spectrometry (SELDI-TOF MS) has been utilized to study the protein pattern of NAF in both healthy women and breast cancer patients [95-97]. Varnum *et al.* identified 64 proteins, by two-dimensional liquid chromatography (2D-LC) coupled with tandem mass spectrometry, using a pooled sample from ten women [43]. Alexander *et al.* characterized 41 proteins in one NAF sample, using two dimensional gel electrophoresis (2D gel) and Matrix Assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry (MALDI-TOF MS). Using the same techniques, they further characterized 3 proteins associated with breast cancer risk by qualitative comparison of protein expression in 10 controls and 10 breast cancer cases [25]. Pawlik *et al.* identified 5 proteins that were differentially expressed in 18 cases vs. 4 control samples, using isotope coded affinity tag and tandem mass spectrometry (MS/MS) [1]. A total of 87 proteins were identified in these prior studies, as listed in Table 4.1.

Two different tandem mass spectrometers have been utilized in this study, namely, MALDI-TOF/TOF MS and ESI-Q-TRAP. MALDI-TOF/TOF MS uses matrix-assisted laser ionization as ion source. Mass of ions was calculated based on the time it takes for each peptide to exit the flight path (time-of-flight, proportional to mass/charge ratio of peptides). ESI-QTRAP, on the other hand, is a hybrid triple quadrupole/linear ion trap mass spectrometer that uses an electron-spray as the ion source. In both types of mass

Table 4.1. Literature review of proteins identified in NAF

Protein ID	Function	MW	Ref[a]
Albumin	Binds to other proteins, water, ions, fatty acids, hormones, etc.	69.3	1, 3
Alpha 1B glycoprotein	Immuno-recognition, member of human immunoglobulin superfamily	54.3	3
Alpha-1-acid glycoprotein 1	Modulates the immune system during the acute-phase reaction	23.7	1
Alpha-1-antitrypsin	Inhibitor of serine proteases	46.7	1, 3
Alpha2-actin	Involved in various types of cell motility	42	3
Alpha2-HS-glycoprotein	Promotes endocytosis, and influences the mineral phase of bone.	39.3	1,2,3
Alpha-antichymotrypsin	Inhibit neutrophil cathepsin G and mast cell chymase	47.7	3
Alpha-fetoprotein	Binds copper, nickel, and fatty acids and bilirubin	68.7	1
Alpha-fibrinogen, chain	Polymerize into fibrin and act as a cofactor in platelet aggregation	95	1
Alpha-importin 2 subunit (similar)	Functions in nuclear protein import as an adapter protein for nuclear receptor KPNB1	57.9	1
Alpha-lactoalbumin	Lactose synthesis	16.2	3
Alpha-L-fucosidase	Hydrolyzes the alpha-1,6-linked fucose	54.1	3
Alpha-S1-casein	Transports calcium phosphate in milk	21.7	3
Aminopeptidase N	Catalyzes the removal of unsubstituted N-terminal amino acids from peptides	55.9	3
Antigen p97	Involved in iron cellular uptake	80.2	3
Anti-thrombin III	Inactivates several enzymes of the coagulation system	15.7	1
ApoD	Transport a variety of hydrophobic ligands including cholesterol and progesterone	21.3	1,3
Apolipoprotein A-I	Participates in the reverse transport of cholesterol from tissues to the liver for excretion	30.8	1
Apolipoprotein A-IV	May have a role in chylomicrons and VLDL secretion and catabolism	45.4	1
Apolipoprotein E	Mediates the binding, internalization, and catabolism of lipoprotein particles	36.1	1,3
Apolipoprotein H	Binds to negatively charged substances such as heparin	38.3	1
BCL2	May be involved in targeting and fusion of retrograde transport vesicles with the ER	26.1	1
Beta-2 microglobulin	Beta-chain of major histocompatibility complex class I molecules	13.7	1,3
Beta-actin	Involved in various types of cell motility.	41.7	3
Beta-casein	Milk micelle formation	25.4	3
Breast cancer anti-estrogen resistance 3	Regulates breast cancer cell proliferation	92.3	1
Butyrophilin	May function in the secretion of milk-fat droplets	59	3
Cancer-associated serine protease protecting peptide	Involved in immunedefense suppression, and serine protease protection	2	3

Protein ID	Function	MW	Ref[a]
Cathepsin-D	Acid protease active in intracellular protein breakdown	44.6	3
CD14	Mediate the innate immune response to bacterial lipopolysaccharide	40.1	3
CD34	Adhesion molecule involved in early hematopoiesis	40.7	3
CD5 antigen-like	May play a role in the regulation of the immune system.	38.1	1
Ceruloplamsin	Ferroxidase activity, amine oxidase activity, copper transport and homeostasis	116.2	3
Clusterin/apo J	Binds to cells, membranes and hydrophobic proteins.	52.5	1,3
Coagulation factor II/thrombin	Involved in coagulation	70	3
Collagen α1	Type I collagen is a member of group I collagen	138.9	3
Complement C7	C7 is a constituent of the membrane attack complex.	93.5	3
Complement component 3	Plays a central role in the activation of the complement system	188.6	3
Complement component C4A	Plays a central role in the activation of the complement system.	193	1
Complement factor B precursor	Factor B which is part of the alternate pathway of the complement system	85.5	1, 3
Complement factor D	Involved in complement activation	27	3
Enhancer of zeste homolog	May be involved in the regulation of gene transcription and chromatin structure	85.3	1
Ephrin	Binds to the receptor tyrosine kinases EPHA2, EPHA4, EPHA5, EPHA6 and EPHA7	23.8	3
Fibrinogen gamma-B chain	Yields monomers that polymerize into fibrin and acts as a cofactor in platelet aggregation.	50.2	3
Gelsolin	Calcium-regulated, actin-modulating protein	85.7	3
Hemopexin	Binds heme and transports it to the liver for breakdown and iron recovery	51.7	2
Hepatoglobin	Combines with free plasma hemoglobin, preventing loss of iron	45.2	1,3
Ig A1	Ig alpha is the major immunoglobulin class in body secretions.	37.6	3
Ig A2	Ig alpha is the major immunoglobulin class in body secretions.	36.4	3
Ig gamma chain	Immune response	37	3
Ig J chain	Link two monomer units of either IgM or IgA	15.6	1,3
Ig Kappa light chain	Immune response	13	1,3
Ig Kappa chain	Immune response	12.7	1,3
Ig lamda chain	Immune response	11.2	1,3
Ig lamda light chain	Immune response	12	1,3
Ig M	Immune response	49.6	3
Iroquois-class homeodomain protein	Not known	49.6	3
Kappa-casein	Stabilizes micelle formation, preventing casein precipitation in milk.	20.3	3

Protein ID	Function	MW	Ref[a]
Keratin 1	Expressed in terminally differentiating epidermis.	65.9	3
Keratin 2a	Probably contributes to terminal cornification.	65.9	3
Kininogen (prekallikrein)	Cleaves Lys-Arg and Arg-Ser bonds,releases bradykinin	71.4	1
Lactoferrin	Iron binding transport proteins; antimicrobial activity	78.3	3
Leucine-rich α-2-glycoprotein	Not known	38.2	3
Lipophilin B	Binds androgens and other steroids	9.98	2
Metastasis-associated protein 1	Involved in the regulation of gene expression by modifying histone proteins	80.8	1
Nebulin-related anchoring protein isoform C	Involved in anchoring the terminal actin filaments in the myofibril to the membrane	197.1	1
Nephrocystin 1	Plays a role in the control of epithelial cell polarity	83.3	1
Osteopontin	Involved in enhancing production of interferon-gamma, interleukin-12 and interleukin-10	35.4	3
Poly-Ig receptor	Binds polymeric IgA and IgM, and transport across the cell	83.3	3
Prolactin-inducible protein	Inducible by prolatin and androgens	16.6	1,3
Prostasin	Possesses a trypsin-like cleavage specificity	36.4	3
Pyruvate kinase isozyme M2	Catalize the reaction: ATP + pyruvate = ADP + phosphoenolpyruvate.	57.8	3
Ras-related protein	May be involved in vesicular trafficking and neurotransmitter release	22.5	1
Retinoic acid receptor responder	Highly expressed in skin	18.6	3
Retinol binding protein precursor	Delivers retinol from the liver stores to the peripheral tissues	23	1
S100 A11	Calcium binding	11.7	3
S100 A9	Expressed by macrophages in acutely inflammated tissues and in chronic inflammations	13.2	3
Thymosin β-4	Plays an important role in the organization of the cytoskeleton	4.9	3
Transcobalamin I	Vitamin B12-binding protein. Transports cobalamin into cells	48.2	3
Transferrin	Iron binding transport proteins	77	1,3
Transthyretin	Thyroid hormone-binding protein	15.9	1
Tumor necrosis factor receptor (55 kDa)	Receptor for TNFSF2/TNF-alpha and homotrimeric TNFSF1/lymphotoxin-alpha	51	3
Tumor-associated antigen 90K	Promotes intergrin-mediated cell adhesion.	65.3	1, 3
Urine protein 1	Binds phosphatidylcholine, phosphatidylinositol, and polychlorinated biphenyls	10	3
VitD binding protein	Carrier of the vitamin D sterols	52.9	1, 2
Zinc finger protein 328	May be involved in transcriptional regulation	94.4	1
Zinc-alpha glycoprotein	Stimulates lipid degradation	33.9	1, 3

[a]1: Alexander *et al.*, 2D gel coupled with MALDI, ELISA; 2: Pawlik *et al.*, ICAT coupled with 1D gel and LC-MS/MS; 3: Varnum *et al.*, LC-MS/MS.

spectrometers, mass spectra of peptides were first obtained in a survey scan. Peaks of interest, for example with top intensities, were selected for further fragmentation in the collision cell, generating the MS/MS spectra for the selected peaks. Information from both the first MS and the second MS/MS scans was used to identify proteins, using specific matching algorithms.

As described in Chapter 3, we identified two distinct types of protein profiles in the NAF of healthy women. The heterogeneity in NAF profiles may be influenced by physiological, nutritional, or pathological conditions, as discussed in Chapter 3. Briefly, we have found that body fat mass and dietary intake of saturated fat have strong influences on the proteomes of NAF. In addition, limited studies found that type II NAF is higher in cancer patients (50%) compared to that in healthy women (10%) [72]. Therefore, to better understand the significance of any specific proteins in NAF as potential biomarkers for breast cancer, we chose to profile their proteomes separately, without pooling samples for analyses.

In this study, prior to the characterization of NAF proteomes by mass spectrometry, all samples were subjected to pattern identification for major proteins using 1D gel and were classified as type I or type II NAF, as detailed in Chapter 3. After identification of NAF types, two strategies were utilized for protein characterization by mass spectrometry. In the first strategy, two-dimensional gel electrophoresis was used to separate proteins first by charge (isoelectric point) and then by molecular weight. Proteins were first focused at their isoelectric points on gel strips with immobilized pH gradient under a high voltage, and subsequently separated on SDS-PAGE gels by molecular weight. After staining, proteins separated on 2D gels were visualized. Visible spots were excised, trypsin-digested, and subjected to structure characterization by mass spectrometry.

Since 2D gel was the cornerstone for protein separation in our study, a well-resolved gel was essential for subsequent protein identification. One of the critical factors for

obtaining a well-resolved gel was removal of contaminants that may interfere with 2D gel electrophoresis. These impurities included salts, lipids and carbohydrates. Several partial purification techniques were devised for removing these components (see section 4.2.2.1 Sample purification).

In the second strategy, 1D gel electrophoresis was used to separate proteins by molecular weight. After staining, the entire gel strip was excised band-by-band and digested by trypsin. The peptic digest from each band was individually analyzed by reverse phase nanoLC coupled directly with a highly sensitive electrospray MS/MS (QTRAP).

The purpose of this study was to qualitatively characterize the proteomes of NAF. Quantitative analysis would require different approaches and was not the goal of this dissertation research.

4.2. Materials and Methods

4.2.1 Sample collection

NAF samples were collected from healthy donors (as described in the Study Design and Methods section in Chapter 2) and kept at -80°C until analysis. All NAF samples were first classified as either type I or type II by 1D SDS-PAGE (as described in Chapter 3) before proteome characterization.

4.2.2 2D gel coupled with MALDI-TOF/TOF

4.2.2.1 Sample purification

For type I NAF, five partial purification methods were used, experimentally, for the removal of contaminants such as lipids, carbohydrates, and nucleotides in the samples, and for improving the resolution of NAF proteins on 2D gel. In the first approach, acetone was used to precipitate proteins. Five hundred μl of acetone with 50 mM

dithiothreitol (DTT) was added to 5 µl of NAF. The solution was incubated at -20°C for 30 min, and spun at 14000 g for 5 min. Supernatants were discarded, and precipitates were dried with speed-vacuum for further analysis. In the second approach, trichloroacetic acid (TCA) precipitation followed by acetone wash was used. Five hundred µl of 20% TCA, 50 mM DTT in acetone was added to the NAF sample. The solution was incubated at -20°C for 30 min, and spun at 14,000 g for 5 min. Supernatants were discarded, and precipitates were washed twice with acetone, and dried with speed-vacuum. In the third approach, petroleum ether precipitation was used. Five hundred µl of petroleum ether was added to 5 µl of NAF sample. The solution was incubated at -20°C for 30 min, and spun at 14000 g for 5 min. Supernatants were discarded, and precipitates were dried with speed-vacuum. In the fourth approach, ultracentrifugation was used. NAF samples were dissolved in 2D sample buffer (containing immobilines pH 3-10, 2M thiourea, 5M urea, 4% CHAPS, and 2% CEBP), and centrifuged at 120,000 g for 4 hours. The top lipid layer and bottom precipitates were discarded, and the middle layer was collected for 2D analysis. In the fifth approach, Cleanascite lipid adsorption and clarification reagent (Biotech Support Group, NJ) was used. Cleanascite reagent was mixed with the NAF sample at a 1:1 ratio (vol/vol). The mixture was incubated at 4°C for one hour, and centrifuged at 1,000 g for 15 min. Precipitates were discarded, and the supernatants were collected for further 2D analysis.

4.2.2.2 2D gel analyses of NAF samples

After partial purification, protein concentration was determined by a Bradford assay (Biorad, CA). NAF samples (100 µg) were dissolved in 200 µl 2D sample buffer that contained immobilines (pH 3-10), 2M thiourea, 5M urea, 4% CHAPS, and 2% CEBP. The first dimension separation of proteins by iso-electric focusing was conducted on a IPG strip (11 cm, pH 3-10) at 18°C. The IPG strip was first rehydrated using 2D sample buffer containing the sample at 50 V for 12 h. Proteins that were adsorbed onto the IPG strip during rehydration were then focused for a total of 30,000 voltage-hours under a maximal of 8,000 volts (IPGphor, Amersham Pharmacia Biotech, NJ). After the first

dimension was completed by iso-electric focusing, the IPG strip was shaken at 100 rpm for 15 min in a buffer solution containing 2% (w/v) DTT, 40 mM Tris-HCl (pH 8.8), 8 M urea, 10% (v/v) glycerol, and 2% (w/v) SDS, to reduce proteins. The reduced proteins on the strip were alkylated by incubating in the above buffer with DTT replaced by 2.5% (w/v) iodoacetamide. After the reduction and alkylation steps, the strip was placed on top of an SDS-PAGE gel (Biorad, 11x 9 cm, 8-16%), and proteins were separated by their molecular weight at 200 volts for 55 minutes. SDS-PAGE gels were then stained with Sypro Ruby (Biorad, CA) overnight. After staining, gel images were acquired using the ProExpress Imaging System (PerkinElmer, MA).

4.2.2.3 In-gel tryptic digestion of proteins

Spots visible under UV light from 2D gel were picked manually on a UV transluminator. The gel plugs were transferred as to eppendorf tubes. The plugs were subjected to in-gel trypsin digestion as follows: The plugs were dehydrated with 100% acetonitrile and then reduced with 50 mM DTT in 50 mM NH_4HCO_3 for 30 min, followed by alkylation with 50 mM iodoacetamide in 20 mM NH_4HCO_3 solution for an additional 30 min. Liquid was removed, and pH adjusted to around 8.0 by the addition of 10 mM NH_4HCO_3 to the samples. The liquid was removed, and the samples dehydrated with 100% acetonitrile. After dehydration, the gel plugs were rehydrated with 10 µl of trypsin solution (10 ng/L bovine modified sequencing grade trypsin from Promega, WI) in 25 mM NH_4HCO_3. Samples were then digested for 8 h at 37°C.

4.2.2.4 MALDI-TOF/TOF analysis of tryptic digests

For MALDI-TOF/TOF analysis, solution containing tryptic peptides from a 2-D gel plug was lyophilized and resuspended in 10 µl of 50% acetonitrile/0.1% formic acid solution. Sample was then mixed with a matrix solution (25 mM α-cyano-4-hydroxycinnamic acid in 50% acetonitrile/0.1% formic acid) (1:1, vol/vol). One µl of the mixture was manually spotted on the sample plate for mass analysis using MALDI-TOF/TOF (4700 Proteomics Analyzer, Applied Biosystems, CA). Mass spectra of the peptides were first obtained in

the mass range between 800 and 4000 using the positive ion reflector mode. For each sample analysis, the top five most intense peptide ions were selected and subjected to fragmentation and sequencing using the MS/MS mode.

4.2.3 1D gel and LC-MS/MS

Protein samples (25 μg) were dissolved in 1D sample buffer (containing 100 mM Tris, pH 6.8, 2% SDS, 5% β-mercaptoethanol, 15% glycerol, and enough bromophenol blue to make the buffer dark) and applied to SDS-PAGE gels (10 cm, 12.5% polyacrylamide). Proteins were separated according to molecular weight (200 V, 55 min). SDS-PAGE gels were stained with Sypro Ruby overnight. Each gel strip was divided into 15 fractions, and each fraction was cut into 2x2 mm pieces. The gel pieces from each fraction were trypsin-digested as detailed in section 4.2.2.3. The tryptic digests were lyophilized until analyses by MS.

LC-MS/MS analyses of peptide mixtures were conducted with a LC-MS/MS equipped with an autosampler (FAMOS^{TM}, LC Packings/Dionex, CA), nano-LC system (LC Packings, CA), a nanoelectrospray interface (Applied Biosystems, CA), and a triple quadrupole/linear ion trap mass spectrometry (4000 Q TRAP, Applied Biosystems, CA). The samples were first loaded onto LCPackings C18 precolumn (300 μm×1mm, particle size 5 μm), washed for 5 min with the loading solvent (5% ACN/H_2O/0.05% formic acid), and then injected onto a LCPackings PepMap C18 column (75 μmx150 mm, particle size 5 μm) for nano-LC separation at a flow rate of 200 nl/min. The solvents used were: (A) a solution containing 5: 95: 0.09 (vol by vol) of water, CAN, and FA, and (B) a solution containing 20: 80: 0.09 (vol by vol) of water, CAN, and FA. The initial gradient was linear from 100% solvent A to a mixture of 40% solvent A and 60% B in 45 min, followed by ramping solvent B to 90% in1 min, which was then maintained for an additional 5 min. The column was reconditioned with 100% solvent A for 15 min after each analysis.

All mass data were obtained under the linear ion trap (LIT) scan modes. An EMS (Enhanced MS) scan was used as a survey scan, followed by an ER (Enhanced Resolution) scan to confirm mass and charge status. A dependent scan, EPI (enhanced product ion) scan, was used to obtain high quality MS/MS spectra on specific ions of interest.

The following IDA (information dependent acquisition) criteria were used for obtaining MS/MS data. The intensity threshold was set to 50,000 counts per second. The top three most intense ions with highest intensity were chosen for MS/MS analysis. The former target ions were excluded after 2 occurrences within a 60 sec window. Rolling collision energy, with collision energy spread of 5 ev, was used for each peptide based on its m/z and charge state. MS/MS parameters used were: NanoSpray source in EPI scan mode with 30 V declustering potential, unit resolution for Q1, fill time of the trap (Q3) for 50 ms and LIT scan rate at 4,000 amu/s. The other parameters were set as follows: curtain gas at 10 psi, ion source at 2.5 kV, temperature at 150°C, gas 1 at 25 psi, gas 2 at 65 psi, and CAD gas at high.

4.2.4 Database search for MALDI-TOF/TOF and LC-MS/MS data

Mass spectra obtained using MALDI-TOF/TOF were processed and submitted for analyses using the Global Protein Server Workstation (Applied Biosystems). Protein searchs were performed against the NCBI nonredundant protein database, based on both mass fingerprints and MS/MS data. Protein identification for the LC-MS/MS data was performed using the MASCOT search engine (Matrix Science, London, UK) against the NCBI nonredundant protein database. Protein identification was considered positive if confidence was greater than 99% and at least two peptide sequences with MS/MS spetra matching to the protein.

4.3 Results

Similar to serum, NAF is a complex biological fluid composed of a mixture of proteins, the concentrations of which span over 12 orders of magnitude. Therefore, some procedures for pre-fractionation and separation of proteins are necessary for effective identification of proteins in NAF samples. Proteins are generally separated by charge, molecular weight, and hydrophobicity. 2D gel was the preferred method for protein separation before MS identification, because the separation is based on both charge and molecular weight. Using a standard protocol for 2D gel preparation, good separations and well-resolved gel images were obtained for type II NAF. However, the same protocol was not effective for type I NAF, as severe streaking was observed on the 2D gels.

4.3.1 2D gel and MALDI-TOF/TOF

4.3.1.1 Protein purification before 2D gel analyses

The most commonly used purification procedure for proteins before 2D analyses involves precipitation of proteins from biological fluids using either acetone or a mixture of acetone and TCA. This procedure removes lipids, lipid soluble substances and other small molecules. The procedure worked well for all type II NAF samples that we have analyzed (N=4), but did not work well for any of the type I NAF samples (N=4), as shown in Figure 4.1. There was severe streaking of the proteins spots in the directions of both isoelectric focusing (IEF) and molecular weight separation of all type I samples.

The major streaking spot in the IEF direction for type I NAF (Figure 4.1 type I) contained apo-D when analyzed by MALDI-TOF/TOF MS. In the second SDS-PAGE gel direction, many gel plugs on the path of streaking were picked, analyzed by MALDI-TOF/TOF MS and found to contain apo-D as well. It appeared that protein(s), e.g. apo-D, precipitated during the isoelectric focusing step, and caused both horizontal and vertical streakings. Apo-D is very hydrophobic and, therefore, may have poor solubility in the sample buffer

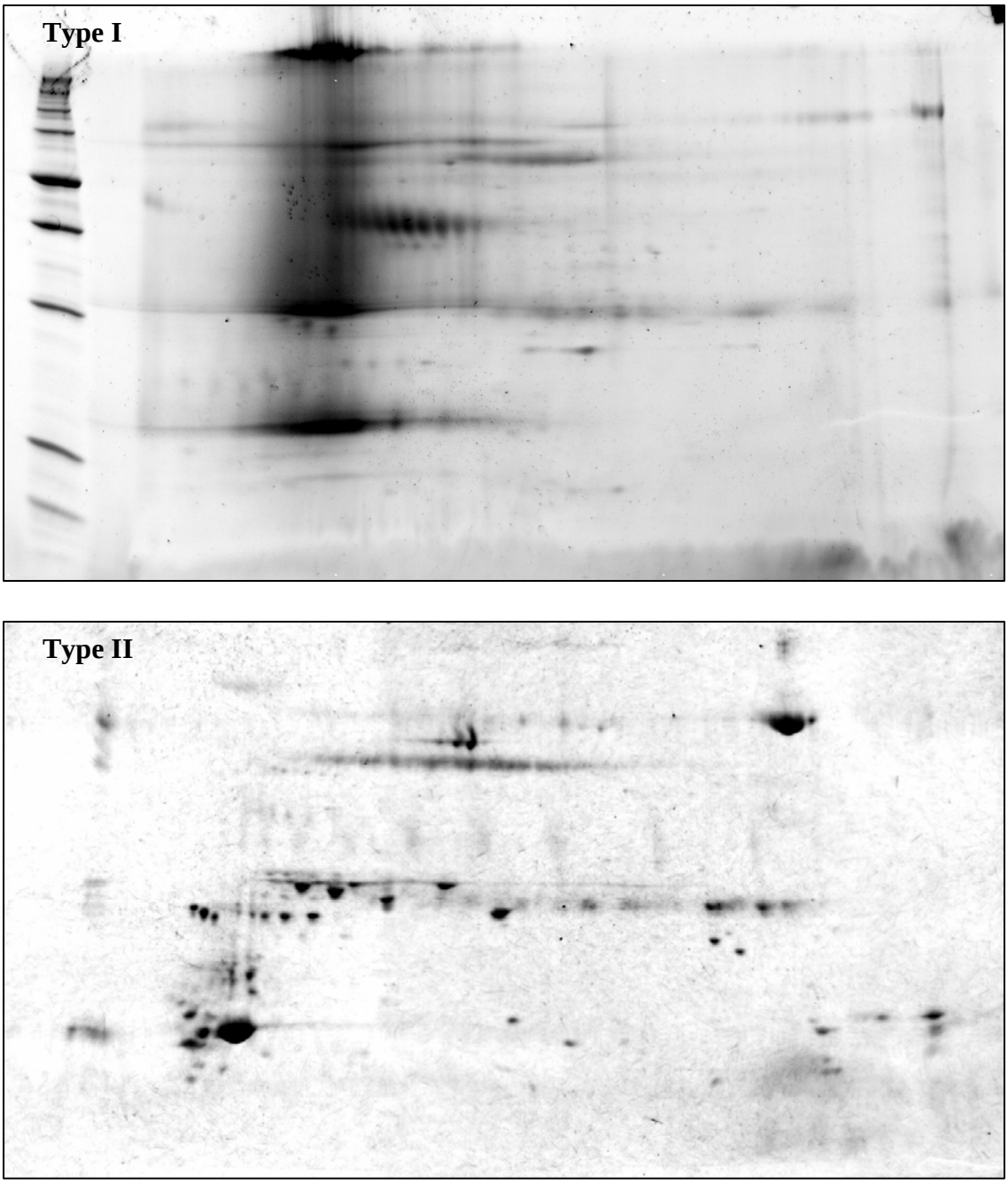


Figure 4.1. 2D gel images of type I and type II NAF

and IPG strips during isoelectric focusing. Because of the poor resolution of type I NAF on 2D gel, a limited number of focused protein spots could be visualized and picked for the identification of proteins using MS. In contrast, protein spots from 2D gel of type II NAF were well focused and could be easily visualized and excised for the identification of proteins by MS.

To improve the performance of type I NAF on 2D gel, various clean-up methods were used, including acetone precipitation (Figure 4.2A), acetone/TCA precipitation (Figure 4.2 B), petroleum ether precipitation (Figure 4.2C), ultracentrifugation (Figure 4.2D), and delipidation by Cleanascite reagent (Figure 4.2E). None of the above methods could clear the streaking on 2D gels of type I NAF (Figure 4.2A-E). Acetone/TCA precipitation did yield better resolution than other methods, and therefore, it was used to separate proteins before 2D gel analysis in this study.

4.3.1.2 MALDI-TOF/TOF analysis for 2D spots

Protein spots visible from all 2D gels, including many of the streaking spots from type I NAF, were excised, digested by trypsin and characterized by MALDI-TOF/TOF. A total of 53 proteins were identified, among which 15 were common to both types, 3 were unique to type I, and 35 were unique to type II (Table 4.2). The name, molecular weight, and major function of proteins, method of structural identification, and the NAF type for proteins identified are listed in Table 4.3.

4.3.2 1D gel and LC-MS/MS

Due to difficulty in resolving proteins from type I NAF on 2D gel, an alternative strategy for protein identification was used. Because both type I and II NAF resolved well in 1D SDS PAGE gel, the entire 1D gel strip from each sample was sectioned slice-by-slice. The tryptic digest from each slice was further separated by a nano-LC coupled to MS/MS. Using this strategy, a total of 75 proteins was identified, among which 25 were common to both types, 12 were unique to type I NAF, and 38 were unique to type II NAF (Table 4.2). The names, molecular weights, and major function of these proteins, the method of

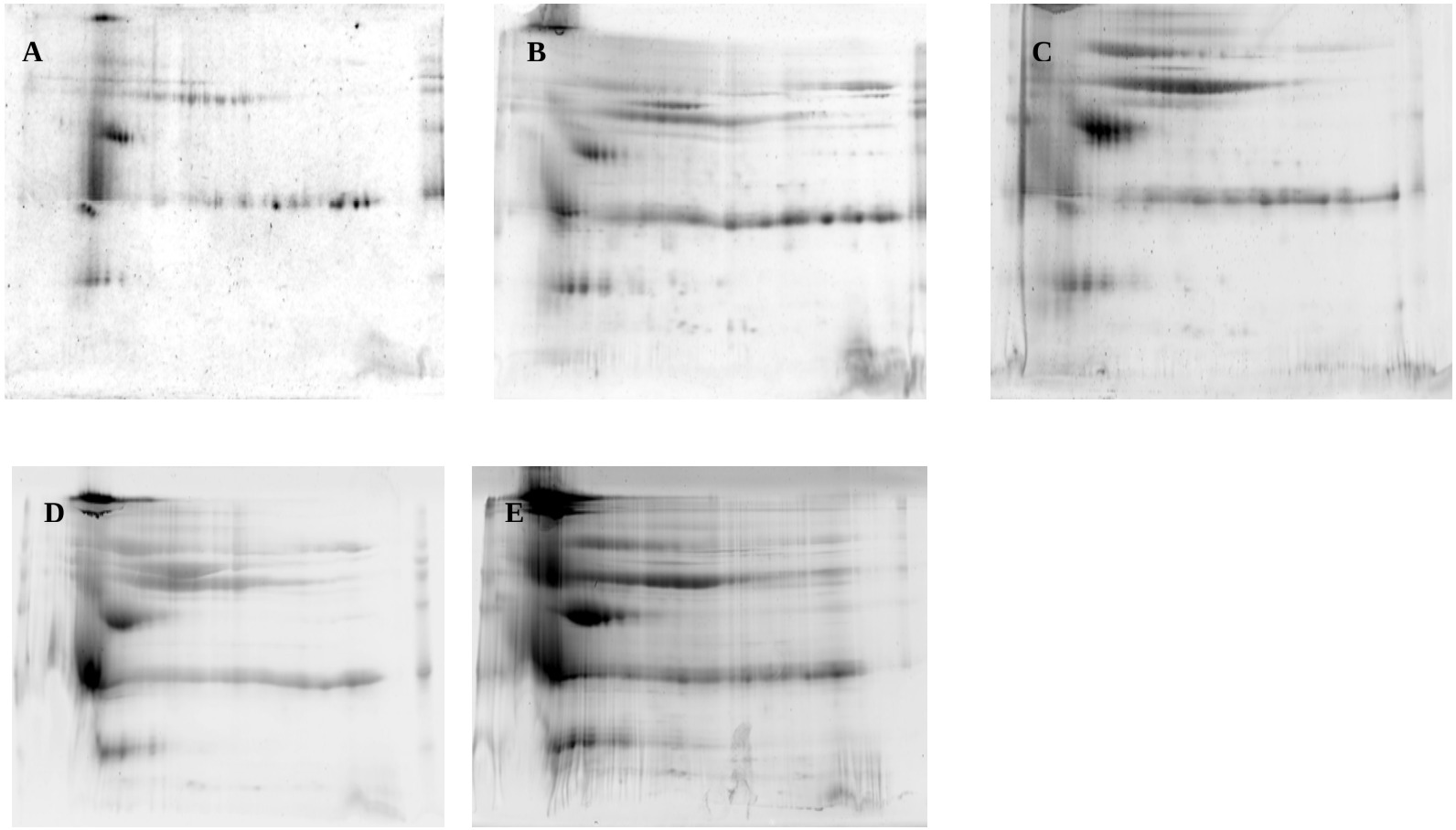


Figure 4.2. 2D gel images for type I NAF after different precipitation strategies: A, acetone precipitation; B, acetone/TCA precipitation; C, petroleum ether precipitation; D, ultracentrifugation; E, Cleanascite lipid adsorption and clarification reagent

Table 4.2. Number of proteins characterized by method of identification and NAF type

Method of Identification	Number of proteins identified			
	Type I only	Type II only	Both types	Total
2D gel and MALDI-TOF/TOF	3	35	15	53[a]
1D gel and LC-MS/MS	12	38	25	75[a]
Total	13	57	29	99

[a]29 proteins were detected by both methods

identification, and the associated NAF type are shown in Table 4.3.

4.3.3 Comparison of two strategies and summary of proteins identified

Among the 99 proteins identified, 24 were detected by 2D gel coupled with MALDI-TOF/TOF only, 46 were characterized by 1D gel coupled with LC-MS/MS only, and 29 were successfully detected by both methods (Figure 4.3).

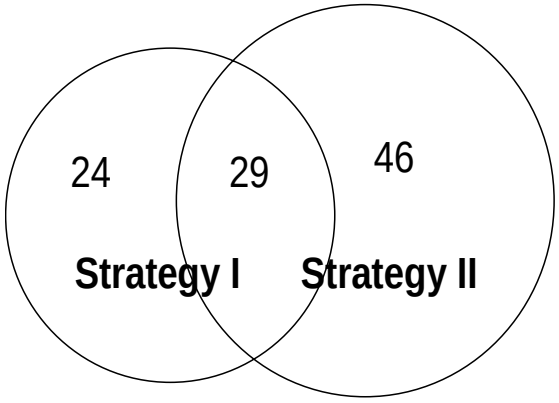


Figure 4.3. Number of proteins identified by two different strategies: Strategy I, 2D gel coupled with MALDI-TOF/TOF; Strategy II, 1D gel coupled with LC-MS/MS.

When the proteins identified by both strategies were combined, a total of 99 different proteins were characterized, among which 29 were common to both types, 13 were

unique to type I NAF, and 57 unique to type II NAF (Table 4.2). Protein ID, major function, molecular weight, NAF type, and method of identification of these proteins were summarized and listed in Table 4.3. These proteins were segregated into 9 groups based on their major functions. Out of the 99 proteins identified, 33 are involved in immune response, 19 are carrier proteins for various steroids, nutrients and other biochemicals, 18 are membrane proteins and proteins involved in structure maintenance, 9 are proteases and their inhibitors, 6 are involved in milk formation, 5 are enzymes, 3 are involved in fibrolysis/coagulation, and the remaining 6 have unclassified/unknown functions.

4.4 Discussion

Two strategies, i.e. 2D gel coupled with MALDI-TOF/TOF and 1D gel coupled with LC-MS/MS, have been utilized for characterizing NAF proteins in our study samples. The first strategy, 2D gel coupled with MALDI-TOF/TOF, performed well in the separation and identification of proteins in type II NAF, but had limited use for characterizing proteins for type I NAF. This may be attributable to the hydrophobic nature of some major proteins, such as apoD and prolactin-inducible protein, in type I NAF. We showed that 1D gel coupled with LC-MS/MS worked well for both types of NAF. The 1D-gel-LC-MS/MS yielded more protein IDs compared to the 2D-gel-MALDI-TOF/TOF (75 vs. 54). In addition, when proteins identified using both strategies were combined, 29 proteins were identified by both methods, 24 proteins were only identified respectively by strategy I and 46 were only identified by strategy II, suggesting that these two methods complement each other (Figure 4.3).

As discussed in Chapter 1, the formation of milk in the breast lobule and duct is an integrated activity in which various types of mammary cells and tissues contribute in a coordinated fashion. Based on the proteins identified in NAF (Table 4.3), we suggest that

Table 4.3. List of NAF proteins identified

Protein ID	Function	MW	Type	Method*
Immune response				
Alpha-2-HS-glycoprotein precursor	Promotes endocytosis, and possesses opsonic properties	39.3	1	2
HLA-B/MHC class 1 antigen	Immune recognition	20.3	1	1
Ig A2	Ig alpha is the major immunoglobulin class in body secretions.	16.3	1	2
Ig gamma	Immune response	16.3	1	2
Small inducible cytokine a17 precursor	Chemotactic factor for T-lymphocytes	10.5	1	2
Alpha 1B glycoprotein	Immuno-recognition, member of human immunoglobulin superfamily	54.3	2	1
Alpha-1-acid glycoprotein 1	Modulates the immune system during the acute-phase reaction	23.7	2	1,2
Alpha-1-acid glycoprotein 2	Modulates the immune system during the acute-phase reaction	23.8	2	1
Cardiotrophin-like cytokine factor 1 precursor	Cytokine with B-cell stimulating capability	25.2	2	2
Complement component C6 precursor	Involved in the formation of the lytic C5b-9m complex	104.8	2	2
Ig Kappa chain V-II region	Immune response	12.7	2	1,2
Ig lambda chain v-iii region	Immune response	11.2	2	2
IL-25	Induces activation of NF-kappa-B and stimulates chemokine IL-8	20.3	2	1,2
Lysozyme C	Bacteriolytic function; enhance the activity of immunoagents	16.5	2	1,2
Macrophage mannose receptor 1	Mediates the endocytosis of glycoproteins by macrophages.	168.9	2	1,2
Monocyte differentiation antigen CD14	Mediate the innate immune response to bacterial lipopolysaccharide	40.1	2	2
Protein S100 A8	Expressed by macrophages in chronic inflammations	10.9	2	1
Protein S100-A9/Calgranulin-B	Expressed by macrophages during inflammations	13.2	2	2
Beta-2 microglobulin	Beta-chain of major histocompatibility complex class I molecules	13.7	1,2	1
C3	Plays a central role in the activation of the complement system	188.6	1,2	1,2
C4	Plays a central role in the activation of the complement system	193	1,2	1
Tumor antigen 90k	Promotes intergrin-mediated cell adhesion	65.3	1,2	2
HLA-A2	Immune recognition	19.3	1,2	1
Ig A1	Ig alpha is the major immunoglobulin class in body secretions.	36.4	1,2	1,2
Ig gamma 1 chain c	Immune response	36.1	1,2	2
Ig J chain	Link two monomer units of either IgM or IgA	15.6	1,2	1,2
Ig kappa chain	Immune response	12.7	1,2	2

Protein ID	Function	MW	Type	Method
Ig Kappa light chain variable region	Immune response	12.7	1,2	1
Ig lambda	Immune response	11.2	1,2	1,2
Ig M	Immune response	49.6	1,2	2
Ig mu chain c region	Immune response	49.6	1,2	2
Poly Ig receptor	Binds polymeric IgA and IgM, and transport across the cell	83.3	1,2	1,2
Protein S100-A6	Expressed when quiescent fibroblasts are stimulated to proliferate	10.2	1,2	2
Carrier protein				
Hemoglobin subunit beta	Oxygen transport from the lung to the peripheral tissues	16	1	2
Lipophilin-B precursor	Bind androgens and other steroids, overexpressed in breast cancer	10	1	2
Uteroglobin precursor	Binds phosphatidylcholine, phosphatidylinositol, and progesterone	10	1	2
ApoE	Binding, internalization, and catabolism of lipoprotein particles	36.2	2	2
ApoH/beta-2 glycoprotein	Binds to various kinds of negatively charged substances such as heparin	38.3	2	2
Fatty acid binding protein, adipocyte	Lipid transport protein in adipocytes.	14.8	2	1
Fatty acid synthase	Catalyzes the formation of long-chain fatty acids	275.9	2	1
Ferritin light subunit	Stores iron in a soluble, nontoxic, readily available form.	16.4	2	1
Transferrin	Iron binding transport proteins	77	2	1,2
Transthyretin	Thyroid hormone-binding protein	15.9	2	1
VitD binding protein	Carrier of the vitamin D sterols	52.9	2	1
Albumin	Binds to other proteins, water, ions, fatty acids, hormones, etc.	69.3	1,2	1,2
ApoD	Transport cholesterol and progesterone	21.3	1,2	1,2
Calreticulin	Calcium binding chaperone	48.1	1,2	1
Clusterin	Binds to cells, membranes and hydrophobic proteins	52.5	1,2	1,2
Fatty acid binding protein, heart	Transports of long-chain fatty acids and their acyl-CoA esters	14.7	1,2	1,2
Fatty acid-binding protein, epidermal	High specificity for fatty acids, esp. for C18 chain length.	15.5	1,2	1,2
Lactoferrin	Iron binding transport proteins; antimicrobial activity	78.3	1,2	1,2
Zinc alpha glycoprotein	Binds to hydrophobic molecules and stimulates lipid degradation	33.9	1,2	1,2
Membrane-binding and structure maintenance				
Butyrophilin subfamily 3 member A1	single-pass type I membrane protein	57.7	1	2
Annexin A2 isoform 1	Calcium-regulated membrane-binding protein	40.7	2	1

Protein ID	Function	MW	Type	Method
Annexin A2 isoform 2	Calcium-regulated membrane-binding protein	38.8	2	1
Annexin A2/Annexin II/Lipocortin II	Calcium-regulated membrane-binding protein	40.7	2	2
Annexin I	Promotes membrane fusion and involved in exocytosis.	38.9	2	1
Annexin V	Anticoagulant protein, inhibitor of the thromboplastin-specific complex	35.8	2	1
Leucine-rich repeat-containing protein 24	single-pass membrane protein	55.2	2	2
Alpha actin-2	Involved in various types of cell motility	42	2	2
Beta-actin/ACTB	Involved in various types of cell motility.	41.7	2	1,2
Dynamin II	Binds and hydrolyzes GTP	97.9	2	1
Gelsolin	Actin-depolymerizing protein	18.5	2	2
Gamma-actin	Involved in various types of cell motility	42.1	2	1
Histone H3.4 (H3t)	Core component of nucleosome	15.5	2	2
Histone H4	Core component of nucleosome	11.4	2	2
Myosin heavy chain	Muscle contraction	223.6	2	2
Osteopontin	Elevated in women with metastatic breast cancer	35.4	2	2
Tenascin precursor	Extracellular matrix protein	240.9	2	2
Tropomyosin 4	Binds to actin filaments, plays a central role in muscle contraction.	28.6	2	1
Protease and inhibitor				
Cystatin C precursor	Inhibitor of cysteine proteinases, inhibit TGF-B initiated metastasis	15.8	1	2
Gamma-glutamyltranspeptidase 1 precursor	Initiates extracellular glutathione (GSH) breakdown	61.4	1	2
Aminopeptidase N/CD13	Broad specificity aminopeptidase, involved in angiogenesis of cancers	109.5	2	2
Cathepsin B precursor	Involved in intracellular degradation and turnover of proteins	37.8	2	2
Cathepsin D	Acid protease active in intracellular protein breakdown.	44.6	2	2
Cystatin B	Intracellular thiol protease inhibitor	11.5	2	1
Plasma protease c1 inhibitor	Activation of the C1 complex is under control of the C1-inhibitor	55.2	2	2
Alpha-1 antitrypsin	Inhibitor of serine proteases	46.7	1,2	1,2
Alpha-1-antichymotrypsin	Inhibit neutrophil cathepsin G and mast cell chymase	47.7	1,2	1,2
Milk formation				
Alpha-lactoalbumin	Lactose synthesis	16.2	2	1,2
Alpha-S1-casein	Transports calcium phosphate in milk	21.7	2	1,2

Protein ID	Function	MW	Type	Method
Kappa casein	Stabilizes micelle formation, preventing casein precipitation in milk.	20.3	2	1,2
Butyrophilin subfamily 1 member A1	May function in the secretion of milk-fat droplets	59	2	2
Lactadherin/Milk fat globule-EGF factor 8	Specific ligand for the alpha-v/beta-3 and alpha-v/beta-5 receptors	43.1	2	2
Beta casein	Milk micelle formation	25.4	1,2	1,2
Enzymes				
Aldehyde oxidase	Belongs to the xanthine dehydrogenase family	147.9	2	2
Ceruloplamsin	Ferroxidase activity, amine oxidase activity, and copper transport	116.2	2	1,2
Enlose 1	Glycolysis, growth control, hypoxia tolerance and allergic responses	47.5	2	1
Xanthine dehydrogenase	Catalyzes the purine degradation pathway	148.8	2	1,2
Bile salt activated lipase	Catalyzes fat and vitamin absorption	78.6	1 ,2	1,2
Coagulation/fibrolysis				
Von Willebrand factor precursor	Involved in blood coagulation	309.3	1	2
Coagulation factor XII precursor	Participates in the initiation of blood coagulation and fibrinolysis	67.8	2	2
Plasminogen	Dissolves the fibrin of blood clots and acts as a proteolytic factor	90.6	1,2	2
Other/not known				
Decidual protein induced by progesterone	Not known	23.4	1	2
AP-1 complex subunit mu-1	Subunit of clathrin-associated adaptor protein complex 1	48.6	2	2
Ataxin-1	Binds RNA in vitro. May be involved in RNA metabolism	87.1	2	2
leucine rich alpha-2 glycoprotein	Secreted protein found in plasma, function not clear	38.2	2	2
Ubiquitin C	Protein degradation	17.1	2	1
Prolactin inducible protein	Inducible by prolatin and androgens	16.6	1,2	1,2

* method: 1, 2D gel coupled with MALDI-TOF/TOF; 2, 1D gel coupled with LC-MS/MS.

many of the mechanisms involved in milk formation are involved in the production of NAF, as discussed below.

As shown in Table 4.3, NAF proteins may play important roles in a wide variety of cellular functions in the breast, for example, immune response, protein degradation, and milk formation. One third of the NAF proteins characterized (33 out of 99) were involved in immune responses. Proteins involved in acquired and innate immune responses are a part of the repertoire of proteins in NAF. The presence of a large number of immunoglobulins, especially secretory IgA, is indicative of humoral immunity. These are likely present in NAF for defense against microorganisms. The detection of several macrophage-associated proteins, complements, bacteriolytic lysozyme, and cytokines in NAF indicate the presence of elements of cellular and innate immunity in breast fluid. Proteins such as tumor antigen 90K stimulate host defense mechanism against pathogens and tumor cells. Members of the S100 protein family, which are calcium binding proteins expressed by macrophages, have also been found in NAF. Elevated levels of S100 proteins in serum and tissues have been associated with various types of cancers, including breast and prostate cancer [98,99]. Thus, a variety of proteins associated with known immune surveillance mechanisms have been found in NAF. Further comparison of profiles of the immune response proteins between breast cancer cases and controls may reveal risk markers for breast cancer.

The second largest group of proteins present in NAF was carrier/transport proteins, including binding proteins for various hormones, lipids, vitamins, iron, and calcium. The breast is a target organ for hormones of endocrine origins (such as sex glands) or paracrine origins (such as local adipocytes). Therefore, the presence of proteins capable of binding and regulating the activities of various hormones, such as androgens, progesterone, and thyroid hormones in NAF, is expected. These proteins may affect the proliferation of breast glands by regulating the availabilities and activities of endocrine and paracrine substances. Among these carrier proteins, lipophilin-B and uteroglobin

both belong to the secretory lipophilin family, and may be involved in the development and metastasis of various types of cancers including breast cancer [100]. Further studies on the association between their expression in NAF and breast cancer risk are warranted. One third group of proteins was mainly membrane-binding proteins and proteins involved in the maintenance of cellular structure. Several members of the annexin family, a group of calcium-regulated membrane-binding proteins, have been found in NAF. Annexins play important roles in signal transduction, cell proliferation and differentiation. Decreased expression of annexins, such as annexin I and II, have been associated with cancer progression [101,102]. Annexin I and dynamin may also participate in the formation and secretion of NAF. Another protein of interest in this group is gelsolin, an actin-binding protein which is important for cell mobility and has been suggested as a prognostic marker for breast cancer [103]. Interestingly, proteins in this group were present predominantly in type II NAF, suggesting that sloughing of membrane proteins and endoplasmic proteins occurs simultaneously with the secretion of other proteins in type II NAF. Further studies are needed to reveal the relevance of this observation to breast cancer risk.

Another group of proteins found in NAF included protease and protease inhibitors, a diversified group involved in a wide variety of biological activities. Proteases and their corresponding inhibitors, such as cathepsin B and cystatin D, are responsible for degradation and maintaining homeostatic functions of proteins, including some matrix proteins. It is not surprising that many proteins in this group, such as cathepsin D (aspartic protease) and cathepsin B (thiol protease), play important roles in the invasion and metastasis of various types of cancers [104,105]. Elevated circulating levels of these proteins have been reported for patients with advanced-stage breast cancer [106,107]. Proteases in this group have also been the targets for certain anti-tumor drugs, such as metalloproteinase inhibitors [108]. Several of the protease inhibitors are involved in inflammation, coagulation/fibrolysis, and protection of tissues against bacteria-derived enzymes.

Based on the major proteins representative of type I and II NAF, we speculate that the dominant secretion pathways for type I and type II NAF differ. First, major type I proteins were also abundant in other apocrine types of tissues, such as sweat glands [57], and fluid from gross cystic disease of the breast [109]. ApoD and GCDFP-15 have been used as biomarkers of apocrine secretion in many studies [110,110]. Second, major type I proteins are more hydrophobic and are often involved in lipid metabolism, consistent with the lipid-secreting function of the apocrine tissues. Last, women with type I NAF were, on average, more obese than those with type II profiles suggesting that body fat played a role in the secretion of type I NAF. It is, therefore, hypothesized that type I NAF may be the result of a gradual up-regulation of apocrine secretion post pregnancy, while type II NAF is primarily derived from transcytosis and exocytosis transport.

In addition to the proteins of high abundance, many of the medium- to low-abundant proteins also demonstrated different distributions between type I and type II NAF. A group of carrier proteins for various nutrients, such as vitamin D binding protein, fatty acid binding proteins, and calcium and iron-binding proteins, are common components of breast milk. These nutrition-associated proteins were also found predominantly in type II NAF. A number of membrane-binding proteins and proteins for maintaining cellular structures were also found predominantly in type II NAF. Among these proteins, annexin I promotes membrane fusion and is involved in exocytosis, and dynamin, is a GTPase that is involved in endocytosis of newly formed vesicles. It is not surprising that a few typical milk-associated proteins, including lactalbumin, casein proteins, lactadherin and butryphilin, were also found only in type II NAF. The presence of different profiles of medium- to low-abundant proteins further supported the hypothesis that the dominating pathways that are involved in the secretion of the two NAF types may differ. A more thorough analysis of the spectra differences between type I and type II may help to provide a better understanding of the mechanisms for NAF formation.

Although type I and type II NAF displayed a major difference in the profiles of the

highly abundant proteins as well as many medium- to low-abundant proteins, they had at least 29 proteins in common using two different strategies for protein identification. The most abundant proteins that are characteristic of type I NAF on 1D gel, such as prolactin-inducible protein, apoD, and ZAG, were also found but with lower abundance in type II NAF. Similarly, some typical type II proteins on 1D gel, such as lactoferrin, were also detected in trace amounts in type I NAF. This is consistent with our speculation that type I and type II may be in equilibrium as a function of time since last pregnancy, as discussed in Chapter 3. Briefly, type II NAF is more prevalent in women with a recent history of pregnancy and the opposite is true for type I, and the presence of a mixed type further supports this hypothesis. NAF contents of non-lactating women represent a balance of secretory activities through apocrine vs. transcytosis/exocytosis pathways.

In summary, we have developed successful methods for the identification of NAF proteins. Using two different state-of-the-art analytical tools, 99 proteins have been characterized in NAF, encompassing a variety of biological functions, such as immune response, binding of steroids and other hormones, membrane binding and signal transduction, and protein degradation. A number of proteins identified may play important roles in defense mechanisms against breast cancer, proliferation of breast glands, and breast cancer metastasis and invasion. Several of these proteins (such as tumor antigen 90 K, S100 proteins, cathepsin D, and lipophilin B) have also been found in serum and reported to be associated with breast cancer [98,111], suggesting that NAF is a good source for discovering breast cancer biomarkers. From the protein profiles identified, it is likely that the formation of NAF and breast milk may share common mechanisms, in which the coordination of multiple pathways is involved. However, the dominant pathways responsible for the secretion of type I and type II NAF may differ. The regulation of protein profiles in NAF may be influenced by the physiologic state of the breast. The results from this study will serve as the basis for further research on potential biomarkers for human breast cancer.

CHAPTER 5: SUMMARY AND FUTURE DIRECTIONS

5.1 Secretor status

Our study of 238 healthy premenopausal women between 30-40 years of age indicates that the secretory activity of the breast is under the influence of multiple factors, especially diet and reproductive background. An earlier menarche and past history of childbirth were associated with being a secretor of NAF. Later age at first childbirth was also associated with being a secretor. Our study also showed, for the first time, that dietary lactose and milk intake, a modifiable factor, was associated with the secretion of NAF. The odds for being a secretor of NAF increased two fold with an increment of 10 g in dietary lactose consumption. However, dietary intake of total calories, proteins, fats or carbohydrates did not affect the secretion of NAF. This result suggests that dietary intake of lactose may be used to increase the secretion volume of NAF.

5.2 Patterns of major proteins in NAF

Two major protein profiles were identified in NAF, with about 60% of the samples being type I, 30% being type II, and 10% having a mixture of both type I and II. Type I is enriched with proteins that are also found in gross cystic disease fluid of the breast; these are mostly lipid mobilizing proteins, such as zinc-alpha glycoprotein and apoD. Type II NAF is abundant in milk-associated proteins such as lactalbumin and lysozyme. However, these differences may be more quantitative than qualitative, as revealed by analyses using mass spectrometry. Comparison of characteristics of women with these two different types of NAF profiles revealed that women with type I NAF were, on average, older, had higher body fat mass, higher consumption of saturated fat, longer years since last childbirth, and were less likely to have breastfed their babies ($P<0.05$). After mutual adjustment using age-adjusted logistic regression models, secretion of type I NAF was predicted independently by more years since last childbirth (OR=2.6, CI 1.3 to 5.2), higher total body fat mass (OR=3.0, CI 1.5 to 6.1), and higher % dietary calorie intake

from saturated fat (OR=4.1, CI 1.1 to 14.6). The level of the above characteristics of women with mixed type NAF generally fell between women with type I and those with type II NAF. Women with type I and type II NAF did not differ significantly in their circulating levels of testosterone, progesterone, or estradiol. Future studies should examine the prevalence of various proteins in NAF in breast cancer cases and controls.

5.3 Proteomic analyses

Using 2D gel coupled with MALDI-TOF/TOF and 1D gel coupled with LC-MS/MS, a total of 99 proteins were characterized in the NAF of healthy, 30 to 40 year-old women. Several methods were tested to improve the resolution of type I NAF proteins on 2D gel, with limited success. Proteins identified included proteins involved in immune response, carrier proteins for nutrients and hormones, proteases and their inhibitors, matrix proteins, and membrane proteins. There appeared to be quantitative differences in these proteins in NAF samples from different women. The significance of these quantitative differences in relation to breast cancer risk will be the focus of future studies. The long term objective is to discover biomarkers for earlier detection of breast cancer than can be achieved by mammography.

REFERENCE LIST

1. Pawlik, T. M., Hawke, D. H., Liu, Y., Krishnamurthy, S., Fritsche, H., Hunt, K. K., and Kuerer, H. M. Proteomic analysis of nipple aspirate fluid from women with early-stage breast cancer using isotope-coded affinity tags and tandem mass spectrometry reveals differential expression of vitamin D binding protein. BMC. Cancer, *6:* 68, 2006.

2. Li, J., Zhao, J., Yu, X., Lange, J., Kuerer, H., Krishnamurthy, S., Schilling, E., Khan, S. A., Sukumar, S., and Chan, D. W. Identification of biomarkers for breast cancer in nipple aspiration and ductal lavage fluid. Clin. Cancer Res., *11:* 8312-8320, 2005.

3. MacMahon, B. Epidemiology and the causes of breast cancer. Int. J. Cancer, *118:* 2373-2378, 2006.

4. McManaman, J. L., Reyland, M. E., and Thrower, E. C. Secretion and fluid transport mechanisms in the mammary gland: comparisons with the exocrine pancreas and the salivary gland. J. Mammary. Gland. Biol. Neoplasia., *11:* 249-268, 2006.

5. King, E. B., Barrett, D., King, M. C., and Petrakis, N. L. Cellular composition of the nipple aspirate specimen of breast fluid. I. The benign cells. Am. J. Clin. Pathol., *64:* 728-738, 1975.

6. Petrakis, N. L. ASPO Distinguished Achievement Award Lecture. Studies on the epidemiology and natural history of benign breast disease and breast cancer using nipple aspirate fluid. Cancer Epidemiol. Biomarkers Prev., *2:* 3-10, 1993.

7. Petrakis, N. L. Physiologic, biochemical, and cytologic aspects of nipple aspirate fluid. Breast Cancer Res. Treat., *8:* 7-19, 1986.

8. Miller, W. R., Elton, R. A., and Loudon, N. Oral contraceptive steroids and breast secretions. Br. J. Obstet. Gynaecol., *96:* 967-972, 1989.

9. Higgins, S. A., Matloff, E. T., Rimm, D. L., Dziura, J., Haffty, B. G., and King, B. L. Patterns of reduced nipple aspirate fluid production and ductal lavage cellularity in women at high risk for breast cancer. Breast Cancer Res., *7:* R1017-R1022, 2005.

10. Huang, Y., Nagamani, M., Anderson, K. E., Kurosky, A., Haag, A. M., Grady, J. J., and Lu, L. J. A strong association between body fat mass and protein profiles in nipple aspirate fluid of healthy premenopausal non-lactating women. Breast Cancer Res. Treat., 2006.

11. Buehring, G. C. Screening for breast atypias using exfoliative cytology. Cancer, *43:* 1788-1799, 1979.

12. Petrakis, N. L., Lim, M. L., Miike, R., Lee, R. E., Morris, M., Lee, L., and Mason, L. Nipple aspirate fluids in adult nonlactating women--lactose content, cationic Na+, K+, Na+/K+ ratio, and coloration. Breast Cancer Res. Treat., *13:* 71-78, 1989.

13. Wynder, E. L., Hill, P., Laakso, K., Littner, R., and Kettunen, K. Breast secretion in Finnish women: a metabolic epidemiologic study. Cancer, *47:* 1444-1450, 1981.

14. Wynder, E. L., Lahti, H., Laakso, K., Cheng, S. L., DeBevoise, S., and Rose, D. P. Nipple aspirates of breast fluid and the epidemiology of breast disease. Cancer, *56:* 1473-1478, 1985.

15. Petrakis, N. L., King, E. B., Lee, M., and Miike, R. Cerumen phenotype and proliferative epithelium in breast fluids of U.S.-born vs. immigrant Asian women: a possible genetic-environmental interaction. Breast Cancer Res. Treat., *16:* 279-285, 1990.

16. Petrakis, N. L., Ernster, V. L., Sacks, S. T., King, E. B., Schweitzer, R. J., Hunt, T. K., and King, M. C. Epidemiology of breast fluid secretion: association with breast cancer risk factors and cerumen type. J. Natl. Cancer Inst., *67:* 277-284, 1981.

17. Wrensch, M. R., Petrakis, N. L., Gruenke, L. D., Ernster, V. L., Miike, R., King, E. B., and Hauck, W. W. Factors associated with obtaining nipple aspirate fluid: analysis of 1428 women and literature review. Breast Cancer Res. Treat., *15:* 39-51, 1990.

18. Baltzell, K. A., Wrensch, M., and Sison, J. D. A descriptive study of variables associated with obtaining nipple aspirate fluid in a cohort of non-lactating women. BMC. Womens Health, *6:* 15, 2006.

19. Lee, M. M., Wrensch, M. R., Miike, R., and Petrakis, N. L. The association of dietary fat with ability to obtain breast fluid by nipple aspiration. Cancer Epidemiol. Biomarkers Prev., *1:* 277-280, 1992.

20. Petrakis, N. L., Barnes, S., King, E. B., Lowenstein, J., Wiencke, J., Lee, M. M., Miike, R., Kirk, M., and Coward, L. Stimulatory influence of soy protein isolate on breast secretion in pre- and postmenopausal women. Cancer Epidemiol. Biomarkers Prev., *5:* 785-794, 1996.

21. Petrakis, N. L. Genetic-environmental interactions in relation to low dose studies: a possible model from breast cancer. Environ. Health Perspect., *42:* 97-102, 1981.

22. Zhang, L., Shao, Z. M., Beatty, P., Sartippour, M., Wang, H. J., Elashoff, R., Chang, H., and Brooks, M. N. The use of oxytocin in nipple fluid aspiration. Breast J., *9:* 266-268, 2003.

23. Klein, P., Glaser, E., Grogan, L., Keane, M., Lipkowitz, S., Soballe, P., Brooks, L., Jenkins, J., Steinberg, S. M., DeMarini, D. M., and Kirsch, I. Biomarker assays in nipple aspirate fluid. Breast J., *7:* 378-387, 2001.

24. Djuric, Z., Visscher, D. W., Heilbrun, L. K., Chen, G., Atkins, M., and Covington, C. Y. Influence of lactation history on breast nipple aspirate fluid yields and fluid composition. Breast J., *11:* 92-99, 2005.

25. Alexander, H., Stegner, A. L., Wagner-Mann, C., Du Bois, G. C., Alexander, S., and Sauter, E. R. Proteomic analysis to identify breast cancer biomarkers in nipple aspirate fluid. Clin. Cancer Res., *10:* 7500-7510, 2004.

26. Buehring, G. C., Letscher, A., McGirr, K. M., Khandhar, S., Che, L. H., Nguyen, C. T., and Hackett, A. J. Presence of epithelial cells in nipple aspirate fluid is associated with subsequent breast cancer: a 25-year prospective study. Breast Cancer Res. Treat., *98:* 63-70, 2006.

27. Wrensch, M. R., Petrakis, N. L., Miike, R., King, E. B., Chew, K., Neuhaus, J., Lee, M. M., and Rhys, M. Breast cancer risk in women with abnormal cytology in nipple aspirates of breast fluid. J. Natl. Cancer Inst., *93:* 1791-1798, 2001.

28. Gruenke, L. D., Wrensch, M. R., Petrakis, N. L., Miike, R., Ernster, V. L., and Craig, J. C. Breast fluid cholesterol and cholesterol epoxides: relationship to breast cancer risk factors and other characteristics. Cancer Res., *47:* 5483-5487, 1987.

29. Chatterton, R. T., Jr., Geiger, A. S., Mateo, E. T., Helenowski, I. B., and Gann, P. H. Comparison of hormone levels in nipple aspirate fluid of pre- and postmenopausal women: effect of oral contraceptives and hormone replacement. J. Clin. Endocrinol. Metab, *90:* 1686-1691, 2005.

30. Rose, D. P. Hormones and growth factors in nipple aspirates from normal women and benign breast disease patients. Cancer Detect. Prev., *16:* 43-51, 1992.

31. Hsiung, R., Zhu, W., Klein, G., Qin, W., Rosenberg, A., Park, P., Rosato, E., and Sauter, E. High basic fibroblast growth factor levels in nipple aspirate fluid are correlated with breast cancer. Cancer J., *8:* 303-310, 2002.

32. Gann, P., Chatterton, R., Vogelsong, K., Dupuis, J., and Ellman, A. Mitogenic growth factors in breast fluid obtained from healthy women: evaluation of biological and extraneous sources of variability. Cancer Epidemiol. Biomarkers Prev., *6:* 421-428, 1997.

33. Sauter, E. R., Garofalo, C., Hewett, J., Hewett, J. E., Morelli, C., and Surmacz, E. Leptin expression in breast nipple aspirate fluid (NAF) and serum is influenced by body mass index (BMI) but not by the presence of breast cancer. Horm. Metab Res., *36:* 336-340, 2004.

34. Chatterton, R. T., Jr., Geiger, A. S., Khan, S. A., Helenowski, I. B., Jovanovic, B. D., and Gann, P. H. Variation in estradiol, estradiol precursors, and estrogen-related products in nipple aspirate fluid from normal premenopausal women. Cancer Epidemiol. Biomarkers Prev., *13:* 928-935, 2004.

35. Sauter, E. R., Tichansky, D. S., Chervoneva, I., and Diamandis, E. P. Circulating testosterone and prostate-specific antigen in nipple aspirate fluid and tissue are associated with breast cancer. Environ. Health Perspect., *110:* 241-246, 2002.

36. Kumar, S. R., Sauter, E. R., Quinn, T. P., and Deutscher, S. L. Thomsen-Friedenreich and Tn antigens in nipple fluid: carbohydrate biomarkers for breast cancer detection. Clin. Cancer Res., *11:* 6868-6871, 2005.

37. Black, M. H. and Diamandis, E. P. The diagnostic and prognostic utility of prostate-specific antigen for diseases of the breast. Breast Cancer Res. Treat., *59:* 1-14, 2000.

38. Sauter, E. R., Klein, G., Wagner-Mann, C., and Diamandis, E. P. Prostate-specific antigen expression in nipple aspirate fluid is associated with advanced breast cancer. Cancer Detect. Prev., *28:* 27-31, 2004.

39. Zhao, Y., Verselis, S. J., Klar, N., Sadowsky, N. L., Kaelin, C. M., Smith, B., Foretova, L., and Li, F. P. Nipple fluid carcinoembryonic antigen and prostate-specific antigen in cancer-bearing and tumor-free breasts. J. Clin. Oncol., *19:* 1462-1467, 2001.

40. Lithgow, D., Nyamathi, A., Elashoff, D., Martinez-Maza, O., and Covington, C. C-reactive protein in nipple aspirate fluid: relation to women's health factors. Nurs. Res., *55:* 418-425, 2006.

41. Sauter, E. R., Lininger, J., Magklara, A., Hewett, J. E., and Diamandis, E. P. Association of kallikrein expression in nipple aspirate fluid with breast cancer risk. Int. J. Cancer, *108:* 588-591, 2004.

42. Qin, W., Zhu, W., Wagner-Mann, C., and Sauter, E. R. Nipple aspirate fluid expression of urokinase-type plasminogen activator, plasminogen activator inhibitor-1, and urokinase-type plasminogen activator receptor predicts breast cancer diagnosis and advanced disease. Ann. Surg. Oncol., *10:* 948-953, 2003.

43. Varnum, S. M., Covington, C. C., Woodbury, R. L., Petritis, K., Kangas, L. J., Abdullah, M. S., Pounds, J. G., Smith, R. D., and Zangar, R. C. Proteomic characterization of nipple aspirate fluid: identification of potential biomarkers of breast cancer. Breast Cancer Res. Treat., *80:* 87-97, 2003.

44. Sauter, E. R., Chervoneva, I., Diamandis, A., Khosravi, J. M., Litwin, S., and Diamandis, E. P. Prostate-specific antigen and insulin-like growth factor binding protein-3 in nipple aspirate fluid are associated with breast cancer. Cancer Detect. Prev., *26:* 149-157, 2002.

45. Isaacs, C., Cavalli, L. R., Cohen, Y., Pennanen, M., Shankar, L. K., Freedman, M., Singh, B., Liu, M., Gallagher, A., Rone, J. D., Dickson, R. B., Sidransky, D., and Haddad, B. R. Detection of LOH and mitochondrial DNA alterations in ductal lavage and nipple aspirate fluids from hngh-risk patients. Breast Cancer Res. Treat., *84:* 99-105, 2004.

46. Zhu, W., Qin, W., Bradley, P., Wessel, A., Puckett, C. L., and Sauter, E. R. Mitochondrial DNA mutations in breast cancer tissue and in matched nipple aspirate fluid. Carcinogenesis, *26:* 145-152, 2005.

47. Krassenstein, R., Sauter, E., Dulaimi, E., Battagli, C., Ehya, H., Klein-Szanto, A., and Cairns, P. Detection of breast cancer in nipple aspirate fluid by CpG island hypermethylation. Clin. Cancer Res., *10:* 28-32, 2004.

48. Zhu, W., Qin, W., Ehya, H., Lininger, J., and Sauter, E. Microsatellite changes in nipple aspirate fluid and breast tissue from women with breast carcinoma or its precursors. Clin. Cancer Res., *9:* 3029-3033, 2003.

49. Malatesta, M., Mannello, F., Bianchi, G., Sebastiani, M., and Gazzanelli, G. Biochemical and ultrastructural features of human milk and nipple aspirate fluids. J. Clin. Lab Anal., *14:* 330-335, 2000.

50. Permyakov, E. A. and Berliner, L. J. alpha-Lactalbumin: structure and function. FEBS Lett., *473:* 269-274, 2000.

51. Sunehag, A. L., Louie, K., Bier, J. L., Tigas, S., and Haymond, M. W. Hexoneogenesis in the human breast during lactation. J. Clin. Endocrinol. Metab, *87:* 297-301, 2002.

52. SHATTON, J. B., GRUENSTEIN, M., SHAY, H., and Weihouse, S. Enzyme of galactose synthesis in mammary gland and mammary tumors of the rat. J. Biol. Chem., *240:* 22-28, 1965.

53. Emerman, J. T., Bartley, J. C., and Bissell, M. J. Interrelationship of glycogen metabolism and lactose synthesis in mammary epithelial cells of mice. Biochem. J., *192:* 695-702, 1980.

54. Lee, M. M., Petrakis, N. L., Wrensch, M. R., King, E. B., Miike, R., and Sickles, E. Association of abnormal nipple aspirate cytology and mammographic pattern and density. Cancer Epidemiol. Biomarkers Prev., *3:* 33-36, 1994.

55. Sanchez, L. M., Vizoso, F., Diez-Itza, I., and Lopez-Otin, C. Identification of the major protein components in breast secretions from women with benign and malignant breast diseases. Cancer Res., *52:* 95-100, 1992.

56. Hu, S., Xie, Y., Ramachandran, P., Ogorzalek Loo, R. R., Li, Y., Loo, J. A., and Wong, D. T. Large-scale identification of proteins in human salivary proteome by liquid chromatography/mass spectrometry and two-dimensional gel electrophoresis-mass spectrometry. Proteomics., *5:* 1714-1728, 2005.

57. Tada, T., Ohkubo, I., Niwa, M., Sasaki, M., Tateyama, H., and Eimoto, T. Immunohistochemical localization of Zn-alpha 2-glycoprotein in normal human tissues. J. Histochem. Cytochem., *39:* 1221-1226, 1991.

58. Ghafouri, B., Kihlstrom, E., Stahlbom, B., Tagesson, C., and Lindahl, M. PLUNC (palate, lung and nasal epithelial clone) proteins in human nasal lavage fluid. Biochem. Soc. Trans., *31:* 810-814, 2003.

59. Bundred, N. J., Scott, W. N., Davies, S. J., Miller, W. R., and Mansel, R. E. Zinc alpha-2 glycoprotein levels in serum and breast fluids: a potential marker of apocrine activity. Eur. J. Cancer, *27:* 549-552, 1991.

60. Haagensen, D. E., Stewart, P., Dilley, W. G., and Wells, S. A. Secretion of breast gross cystic disease fluid proteins by T47D breast cancer cells in culture--modulation by steroid hormones. Breast Cancer Res. Treat., *23:* 77-86, 1992.

61. Simard, J., de Launoit, Y., Haagensen, D. E., and Labrie, F. Additive stimulatory action of glucocorticoids and androgens on basal and estrogen-repressed apolipoprotein-D messenger ribonucleic acid levels and secretion in human breast cancer cells. Endocrinology, *130:* 1115-1121, 1992.

62. Boutteville, C., Revillion, F., Vandewalle, B., and Lefebvre, J. [Hormonal regulation of GCDFP-15 secretion in breast tumors. Study of cultured cells]. Bull. Cancer, *76:* 805-811, 1989.

63. Simard, J., Hatton, A. C., Labrie, C., Dauvois, S., Zhao, H. F., Haagensen, D. E., and Labrie, F. Inhibitory effect of estrogens on GCDFP-15 mRNA levels and secretion in ZR-75-1 human breast cancer cells. Mol. Endocrinol., *3:* 694-702, 1989.

64. Simard, J., Dauvois, S., Haagensen, D. E., Levesque, C., Merand, Y., and Labrie, F. Regulation of progesterone-binding breast cyst protein GCDFP-24 secretion by estrogens and androgens in human breast cancer cells: a new marker of steroid action in breast cancer. Endocrinology, *126:* 3223-3231, 1990.

65. Murphy, L. C., Tsuyuki, D., Myal, Y., and Shiu, R. P. Isolation and sequencing of a cDNA clone for a prolactin-inducible protein (PIP). Regulation of PIP gene expression in the human breast cancer cell line, T-47D. J. Biol. Chem., *262:* 15236-15241, 1987.

66. Zhou, J., Ng, S., Adesanya-Famuiya, O., Anderson, K., and Bondy, C. A. Testosterone inhibits estrogen-induced mammary epithelial proliferation and suppresses estrogen receptor expression. FASEB J., *14:* 1725-1730, 2000.

67. Topper, Y. J. and Freeman, C. S. Multiple hormone interactions in the developmental biology of the mammary gland. Physiol Rev., *60:* 1049-1106, 1980.

68. Ray, D. B., Horst, I. A., Jansen, R. W., and Kowal, J. Normal mammary cells in long term culture. I. development of hormone-dependent functional monolayer cultures and assay of alpha-lactalbumin production. Endocrinology, *108:* 573-583, 1981.

69. Teng, C. T., Pentecost, B. T., Chen, Y. H., Newbold, R. R., Eddy, E. M., and McLachlan, J. A. Lactotransferrin gene expression in the mouse uterus and mammary gland. Endocrinology, *124:* 992-999, 1989.

70. Vizoso, F., Plaza, E., Vazquez, J., Serra, C., Lamelas, M. L., Gonzalez, L. O., Merino, A. M., and Mendez, J. Lysozyme expression by breast carcinomas, correlation with clinicopathologic parameters, and prognostic significance. Ann. Surg. Oncol., *8:* 667-674, 2001.

71. Webster, J. and Oxley, D. Peptide mass fingerprinting: protein identification using MALDI-TOF mass spectrometry. Methods Mol. Biol., *310:* 227-240, 2005.

72. Vizoso, F., Sanchez, L. M., Diez-Itza, I., Luz, L. M., and Lopez-Otin, C. Factors affecting protein composition of breast secretions from nonlactating women. Breast Cancer Res. Treat., *23:* 251-258, 1992.

73. Vizoso, F., Diez-Itza, I., Sanchez, L. M., Tuya, A. F., Ruibal, A., and Lopez-Otin, C. Relationship between serum prolactin levels and protein composition of breast secretions in nonlactating women. J. Clin. Endocrinol. Metab, *79:* 525-529, 1994.

74. Bray, G. A. The underlying basis for obesity: relationship to cancer. J. Nutr., *132:* 3451S-3455S, 2002.

75. Bray, G. A. Medical consequences of obesity. J. Clin. Endocrinol. Metab, *89:* 2583-2589, 2004.

76. Irwin, M. L., McTiernan, A., Bernstein, L., Gilliland, F. D., Baumgartner, R., Baumgartner, K., and Ballard-Barbash, R. Relationship of obesity and physical activity with C-peptide, leptin, and insulin-like growth factors in breast cancer survivors. Cancer Epidemiol. Biomarkers Prev., *14:* 2881-2888, 2005.

77. Quinkler, M., Sinha, B., Tomlinson, J. W., Bujalska, I. J., Stewart, P. M., and Arlt, W. Androgen generation in adipose tissue in women with simple obesity--a site-specific role for 17beta-hydroxysteroid dehydrogenase type 5. J. Endocrinol., *183:* 331-342, 2004.

78. Liu, Z., Chang, G. Q., and Leibowitz, S. F. Apolipoprotein D interacts with the long-form leptin receptor: a hypothalamic function in the control of energy homeostasis. FASEB J., *15:* 1329-1331, 2001.

79. Bing, C., Bao, Y., Jenkins, J., Sanders, P., Manieri, M., Cinti, S., Tisdale, M. J., and Trayhurn, P. Zinc-alpha2-glycoprotein, a lipid mobilizing factor, is expressed in adipocytes and is up-regulated in mice with cancer cachexia. Proc. Natl. Acad. Sci. U. S. A, *101:* 2500-2505, 2004.

80. Gohda, T., Makita, Y., Shike, T., Tanimoto, M., Funabiki, K., Horikoshi, S., and Tomino, Y. Identification of epistatic interaction involved in obesity using the KK/Ta mouse as a Type 2 diabetes model: is Zn-alpha2 glycoprotein-1 a candidate gene for obesity? Diabetes, *52:* 2175-2181, 2003.

81. Manni, A., Santen, R. J., Boucher, A. E., Lipton, A., Harvey, H., Drago, J., Rohner, T., Haagensen, D., Glode, M., and Santner, S. J. Evaluation of CEA and GCDFP-15 plasma level during hormonally induced cancer stimulation. Anticancer Res., *4:* 141-144, 1984.

82. Haagensen, D. E., Jr., Dilley, W. G., Mazoujian, G., and Wells, S. A., Jr. Review of GCDFP-15. An apocrine marker protein. Ann. N. Y. Acad. Sci., *586:* 161-173, 1990.

83. Anderson, L. and Seilhamer, J. A comparison of selected mRNA and protein abundances in human liver. Electrophoresis, *18:* 533-537, 1997.

84. Tyers, M. and Mann, M. From genomics to proteomics. Nature, *422:* 193-197, 2003.

85. Kallioniemi, O. P., Wagner, U., Kononen, J., and Sauter, G. Tissue microarray technology for high-throughput molecular profiling of cancer. Hum. Mol. Genet., *10:* 657-662, 2001.

86. Bertucci, F., Birnbaum, D., and Goncalves, A. Proteomics of breast cancer: principles and potential clinical applications. Mol. Cell Proteomics., *5:* 1772-1786, 2006.

87. Dua, R. S., Isacke, C. M., and Gui, G. P. The intraductal approach to breast cancer biomarker discovery. J. Clin. Oncol., *24:* 1209-1216, 2006.

88. Pawlik, T. M. and Kuerer, H. M. The evolving role of proteomics in the early detection of breast cancer. Int. J. Fertil. Womens Med., *50:* 212-216, 2005.

89. Abramovitz, M. and Leyland-Jones, B. A systems approach to clinical oncology: focus on breast cancer. Proteome. Sci., *4:* 5, 2006.

90. Davis, M. A. and Hanash, S. High-throughput genomic technology in research and clinical management of breast cancer. Plasma-based proteomics in early detection and therapy. Breast Cancer Res., *8:* 217, 2006.

91. Foretova, L., Garber, J. E., Sadowsky, N. L., Verselis, S. J., Joseph, D. M., Andrade, A. F., Gudrais, P. G., Fairclough, D., and Li, F. P. Carcinoembryonic antigen in breast nipple aspirate fluid. Cancer Epidemiol. Biomarkers Prev., *7:* 195-198, 1998.

92. Sauter, E. R., Daly, M., Linahan, K., Ehya, H., Engstrom, P. F., Bonney, G., Ross, E. A., Yu, H., and Diamandis, E. Prostate-specific antigen levels in nipple aspirate fluid correlate with breast cancer risk. Cancer Epidemiol. Biomarkers Prev., *5:* 967-970, 1996.

93. Inaji, H., Koyama, H., Motomura, K., Noguchi, S., Tsuji, N., and Wada, A. Immunohistochemical survey of pS2 expression in intraductal lesions associated with invasive ductal carcinoma of the breast. Int. J. Cancer, *55:* 883-886, 1993.

94. Kawamoto, M. Breast cancer diagnosis by lactate dehydrogenase isozymes in nipple discharge. Cancer, *73:* 1836-1841, 1994.

95. Sauter, E. R., Shan, S., Hewett, J. E., Speckman, P., and Du Bois, G. C. Proteomic analysis of nipple aspirate fluid using SELDI-TOF-MS. Int. J. Cancer, *114:* 791-796, 2005.

96. Paweletz, C. P., Trock, B., Pennanen, M., Tsangaris, T., Magnant, C., Liotta, L. A., and Petricoin, E. F., III Proteomic patterns of nipple aspirate fluids obtained by SELDI-TOF: potential for new biomarkers to aid in the diagnosis of breast cancer. Dis. Markers, *17:* 301-307, 2001.

97. Pawlik, T. M., Fritsche, H., Coombes, K. R., Xiao, L., Krishnamurthy, S., Hunt, K. K., Pusztai, L., Chen, J. N., Clarke, C. H., Arun, B., Hung, M. C., and Kuerer, H. M. Significant differences in nipple aspirate fluid protein expression between healthy women and those with breast cancer demonstrated by time-of-flight mass spectrometry. Breast Cancer Res. Treat., *89:* 149-157, 2005.

98. Hermani, A., Hess, J., De Servi, B., Medunjanin, S., Grobholz, R., Trojan, L., Angel, P., and Mayer, D. Calcium-binding proteins S100A8 and S100A9 as novel diagnostic markers in human prostate cancer. Clin. Cancer Res., *11:* 5146-5152, 2005.

99. Cross, S. S., Hamdy, F. C., Deloulme, J. C., and Rehman, I. Expression of S100 proteins in normal human tissues and common cancers using tissue microarrays: S100A6, S100A8, S100A9 and S100A11 are all overexpressed in common cancers. Histopathology, *46:* 256-269, 2005.

100. Culleton, J., O'Brien, N., Ryan, B. M., Hill, A. D., McDermott, E., O'Higgins, N., and Duffy, M. J. Lipophilin B: A gene preferentially expressed in breast tissue and upregulated in breast cancer. Int. J. Cancer, *120:* 1087-1092, 2007.

101. Ornstein, D. K. and Tyson, D. R. Proteomics for the identification of new prostate cancer biomarkers. Urol. Oncol., *24:* 231-236, 2006.

102. Hayes, M. J. and Moss, S. E. Annexins and disease. Biochem. Biophys. Res. Commun., *322:* 1166-1170, 2004.

103. Thor, A. D., Edgerton, S. M., Liu, S., Moore, D. H., and Kwiatkowski, D. J. Gelsolin as a negative prognostic factor and effector of motility in erbB-2-positive epidermal growth factor receptor-positive breast cancers. Clin. Cancer Res., *7:* 2415-2424, 2001.

104. Talieri, M., Papadopoulou, S., Scorilas, A., Xynopoulos, D., Arnogianaki, N., Plataniotis, G., Yotis, J., and Agnanti, N. Cathepsin B and cathepsin D expression in the progression of colorectal adenoma to carcinoma. Cancer Lett., *205:* 97-106, 2004.

105. Downs, L. S., Jr., Lima, P. H., Bliss, R. L., and Blomquist, C. H. Cathepsins B and D activity and activity ratios in normal ovaries, benign ovarian neoplasms, and epithelial ovarian cancer. J. Soc. Gynecol. Investig., *12:* 539-544, 2005.

106. Lah, T. T., Cercek, M., Blejec, A., Kos, J., Gorodetsky, E., Somers, R., and Daskal, I. Cathepsin B, a prognostic indicator in lymph node-negative breast carcinoma patients: comparison with cathepsin D, cathepsin L, and other clinical indicators. Clin. Cancer Res., *6:* 578-584, 2000.

107. Garcia, M., Platet, N., Liaudet, E., Laurent, V., Derocq, D., Brouillet, J. P., and Rochefort, H. Biological and clinical significance of cathepsin D in breast cancer metastasis. Stem Cells, *14:* 642-650, 1996.

108. Butler, G. S. and Overall, C. M. Proteomic validation of protease drug targets: pharmacoproteomics of matrix metalloproteinase inhibitor drugs using isotope-coded affinity tag labelling and tandem mass spectrometry. Curr. Pharm. Des, *13:* 263-270, 2007.

109. Haagensen, D. E., Jr., Mazoujian, G., Dilley, W. G., Pedersen, C. E., Kister, S. J., and Wells, S. A., Jr. Breast gross cystic disease fluid analysis. I. Isolation and radioimmunoassay for a major component protein. J. Natl. Cancer Inst., *62:* 239-247, 1979.

110. Viacava, P., Naccarato, A. G., and Bevilacqua, G. Apocrine epithelium of the breast: does it result from metaplasia? Virchows Arch., *431:* 205-209, 1997.

111. Iacobelli, S., Sismondi, P., Giai, M., D'Egidio, M., Tinari, N., Amatetti, C., Di Stefano, P., and Natoli, C. Prognostic value of a novel circulating serum 90K antigen in breast cancer. Br. J. Cancer, *69:* 172-176, 1994.

VITA

Yafei Huang was born in Hubei, China on April 20th 1976 to Xidong Huang and Jinrong Kong. She has two daughters, Marissa Liu, and Serena Liu, with husband, Ping Liu, who is a current graduate student at the University of Texas Medical Branch at Galveston TX.

In September 1993, Yafei Huang was admitted to School of Medicine, Beijing University of Chinese Medicine in Beijing China, to study both modern medicine and traditional Chinese medicine. She graduated in July 2000 from Beijing University of Chinese Medicine with Bachelor of Medicine and Master of Science degrees. In August 2000, she entered Graduate School of Biomedical Sciences at the University of Texas Medical Branch at Galveston TX. During her graduate training at UTMB, Yafei Huang received a Predocotoral Fellowship Award from U.S. Department of Defense Breast Cancer Research Program (2003-2006), was honored by Katherina Siebert Award for Excellence in Oncologic Research (2006), James E. Beall II Memorial Scholarship (2006), Shirley Patricia Parker Scholarship (2005), Margaret Saunders Travel Award (2004), Arthur V. Simmang Academic Scholarships (2004), Don W. Micks Scholarship (2004), and Elias Hochman Research Award (2003) from the University of Texas Medical Branch, and was also an honoree of "Who's Who among Students in American Colleges and Universities" (2005) from Who's Who Organization.

PUBLICATIONS

Manuscripts

1. **Yafei Huang**, Shimin Cao, Karl E. Anderson, James J. Grady, Manubai Nagamani and Lee-Jane W. Lu, Decreased circulating levels of tumor necrosis factor α (TNF-α) in postmenopausal women during consumption of soy containing isoflavones, J Clin Endocrinol Metab. 2005 Jul;90(7):3956-62.

2. **Yafei Huang**, Manubai Nagamani, Karl E. Anderson, Alexander Kurosky, Anthony M Haag, James J. Grady, and Lee-Jane W. Lu, A strong association between body fat mass and protein profiles in nipple aspirate fluid of healthy premenopausal non-lactating women, Breast Cancer Res Treat. 2007 Jul;104(1):57-66.

3. **Yafei Huang**, Manubai Nagamani, Karl E. Anderson, James J. Grady, and Lee-Jane W. Lu, Dietary lactose intake as a strong predictor for secretor status of nipple aspirate fluid in healthy premenopausal non-lactating women. (completed)

4. **Yafei Huang**, Alexander Kurosky, and Lee-Jane W Lu, Proteomic analysis of nipple aspirate fluid to identify breast cancer biomarkers (in preparation)

5. Fatima Nayeem, **Yafei Huang**, Karl E Anderson, James J Grady, Manubai Nagamani, Lee-Jane W Lu, Correlates of Sex Hormone Binding Globulin in Premenopausal Women (in preparation)

Abstracts

1. **Yafei Huang**, Shimin Cao, Karl E. Anderson, James J. Grady, Manubai Nagamani and Lee-Jane W. Lu, Decreased circulating levels of tumor necrosis factor α (TNF-α) in postmenopausal women during consumption of soy containing isoflavones. Presented at the 2003 annual meeting of the American Association for Cancer Research, Abstract # 1514;

2. **Yafei Huang**, Alexander Kurosky, Zheng Wu, Anthony Haag and Lee-Jane W Lu, Proteomic analysis of nipple aspiration fluid in non-lactating pre-menopausal women by two-dimensional gel electrophoresis and matrix-assisted laser desorption/ionization mass spectrometry (MALDI MS). Presented at the 2003 annual meeting of the Endocrine Society. Manuscript in preparation, Abstract # 85027.

3. **Yafei Huang**, Manubai Nagamani, James J. Grady, Karl E. Anderson, Lee-Jane W. Lu. Factors affecting secretor status of nipple aspirate fluid in healthy non-lactating women. Presented at the 2004 AACR Frontiers in Cancer Prevention Meeting. Abstract # 312

4. **Yafei Huang**, Manubai Nagamani, Lee-Jane W. Lu. Factors affecting secretor status and secreting types of nipple aspirate fluid in non-lactating premenopausal women. Presented at the 2004 annual meeting of the Endocrine Society. Abstract # 850030

5. **Yafei Huang**, Manubai Nagamani, Karl E. Anderson, James J. Grady, Lee-Jane W. Lu. Obesity as a predictor for secretor status and type of fluid obtained by nipple aspiration in healthy non-lactating premenopausal women. Accepted for presentation at the Era of Hope 2005 Department of Defense Breast Cancer Research Program Meeting.

Permanent address:

Water Utility
Xishui County
Hubei Province, China, 436200

This dissertation was typed by Yafei Huang