In warm memory of my parents

Acknowledgements

The work with this thesis took place from 2005 to 2010 at the Department of Community Medicine, the University in Tromsø, and was sponsored by a university grant. I have had five inspiring, challenging and educational years at "ISM" among knowledgeable and supportive colleagues.

I would like to express my gratitude toward my supervisor Eiliv Lund for giving me the opportunity to work with the Norwegian Women and Cancer study, and for his enthusiastic and unceremonious guidance.

Although not formally appointed supervisors, I consider Kjersti Bakken and Vanessa Dumeaux my mentors in pharmaco-epidemiological and molecular epidemiological research, respectively. I greatly appreciate every feed-back and support. I also thankfully acknowledge the assistance from the always obliging statisticians Marita Melhus and Tonje Braathen, and all helpful contributions from my co-authors.

I owe deep gratitude to the NOWAC factotum Bente Augdal, whose rational and down-to-earth manner effectively calms down green PhD-students, and whose systematic capabilities keep the NOWAC data in order. I am grateful for additional support from all NOWAC-associated colleagues, and most particularly Karina Standahl Olsen, Lotta Rylander, Jostein Johansen, Guri Skeie, Kristin Sørensen and Merethe Kumle. As an infrequent participant in the "morning coffee club" I want to thank fellow PhD-students Signe, Anne Mette and Tove for this forum for life reflection.

I thank Sameline Grimsgaard for inspiring me to apply for a PhD-grant and for urging me to make the choice of research field my own decision. I also thank my colleagues in pharmacy and pharmaceutical education, my friends and my dear family.

Last but not least, like everyone working with NOWAC data I am deeply grateful for all the women who have contributed to my research by sending their health information and blood samples to the Department of Community Medicine in Tromsø.

Marit

Table of content

Abstract

Endogenous and exogenous sex hormones are known risk factors for hormone dependent cancers like breast cancer. Use of postmenopausal hormone therapy (HT) increased through the 1990s until large randomized trials showed an unfavourable risk-benefit ratio for HT use. The media coverage of the Women's Health Initiative publication in 2002 was the main opinion-former. Developments in high throughput technology have provided the opportunity to investigate relationships between sex hormones and gene expression in a population based cohort like the Norwegian Women and Cancer study (NOWAC).

The aim of the thesis was to describe changes in the prevalence of self-reported HT use among Norwegian women, to examine the association between HT use and endogenous sex hormone levels and to explore the relationship between sex hormones and gene expression utilizing the NOWAC postgenome biobank.

The research was performed in a cross-sectional design within NOWAC. In a multiangular approach using several data sources and a broad spectre of analytical methods, female sex hormones were described according to consumption (paper I), plasma levels (paper II) and gene expression (paper III). Paper I is a pharmaco-epidemiological study of HT use while paper II validates the variables current HT use and menopausal status. Paper III represents a feasibility study, investigating gene expression patterns related to sex hormones in a population based sample of women.

The analyses showed increasing HT use toward 2002 followed by a steep decline toward 2005. The extensive long-term use found in 2005 has subsequently declined according to data from the Norwegian Prescription Database. Plasma hormone concentrations verified that the NOWAC questionnaires provide valid information on HT use and menopausal status. Both endogenous and exogenous sex hormones were associated with distinct gene expression profiles in peripheral blood and a novel estrogen signature was detected. Although further confirmation through analysis of an independent data set is needed, the results show that gene expression profiling is functional in an epidemiological context.

List of papers

Paper I

Waaseth M, Bakken K, Lund E. Patterns of hormone therapy use in the Norwegian Women and Cancer study (NOWAC) 1996-2005. Maturitas 2009;63:220-6.

Paper II

Waaseth M, Bakken K, Dumeaux V, Olsen KS, Rylander C, Figenschau Y and Lund E. Hormone replacement therapy use and plasma levels of sex hormones in the Norwegian Women and Cancer Postgenome Cohort - a cross-sectional analysis. BMC Women's Health 2008;8:1.

Paper III

Waaseth M, Dumeaux V, Olsen KS, Rylander C, and Lund E. Sex hormones and gene expression in peripheral blood from postmenopausal women – the NOWAC postgenome study. (Submitted)

Abbreviations

BMI	Body Mass Index
CI	Confidence Interval
CVD	Cardiovascular Disease
DNA	Deoxyribonucleic acid
E_2	Estradiol
FDR	False Discovery Rate
FSH	Follicle stimulating hormone
HERS	The Heart and Estrogen/Progestin Replacement Study
HRT	Hormone replacement therapy
HT	Hormone therapy
IARC	The International Agency for Research on Cancer
LH	Luteinizing hormone
mRNA	Messenger-RNA
MWS	The Million Women Study
NHS	The Nurses Health Study
NorPD	The Norwegian Prescription Database
NOWAC	The Norwegian Women and Cancer study
OC	Oral contraceptives
P_4	Progesterone
RCT	Randomized Controlled Trial
RNA	Ribonucleic acid
SHBG	Sex Hormone Binding Globuline
T	Testosterone
T_4	Thyroxine
WEST	The Women's Estrogen for Stroke Trial
WHI	The Women's Health Initiative
WISDOM	The Women's International Study of long Duration Oestrogen after Menopause

1 Introduction

Hormone therapy (HT), also known as hormone replacement therapy (HRT), has been used against climacteric complaints among menopausal women for almost 60 years. In Norway the first hormone therapy product was marketed in 1953 (Etifollin®). Since then, HT consumption has varied according to shifting interpretations of its risks and benefits, based on progressions in research.

During the 1990s the protruding view was that HT would benefit almost any menopausal woman. This was based on results from observational studies, both case-control studies and cohort studies, e.g. the Nurses Health Study (NHS). In the early 1990s combined results from these studies showed a protective effect of HT on cardiovascular disease (CVD) and mortality[1, 2]. These results were supported by favourable effects of HT on known risk factors for CVD, like plasma levels of cholesterol and fibrinogen[3]. Another acknowledged indication for HT use was osteoporosis prevention[4]. The risk of endometrial cancer associated with use of unopposed estrogen was counteracted by adding progestogen to treatments intended for women with intact uterus[5, 6]. Some reports published from the early 1980's both described a potential risk of breast cancer[7] and suggested a more watchful approach towards HT's alleged beneficial effects on heart disease[8, 9]. However, these reports probably did not reach clinicians to the same extent as positive reports, due to selective marketing strategies performed by the pharmaceutical companies producing HT. Instead, the proclaimed favourable risk-benefit ratio led to a marked increase in HT use. In Norway the increase was particularly steep[10], as shown by sales figures (Figure 1.1). Based on the results of the early observational studies, several large randomized controlled trials (RCTs) were initiated to test the hypotheses in an experimental context. In 1998, results from the Heart and Estrogen/progestin Replacement Study (HERS) reported no benefit of HT as secondary prevention of CVD among women with established coronary disease[11]. The results from the Women's Estrogen for Stroke Trial (WEST) in 2001 showed a possible increased risk of a second stroke among women randomized to HT treatment[12]. The Women's Health Initiative (WHI) study, a large RCT conducted among healthy women, similarly showed no benefit of HT for the primary prevention of CVD[13]. On the contrary, when considering all

outcomes measured, there seemed to be an unfavourable overall risk-benefit ratio, particularly among users of combined estrogen/progestogen therapy. Hence, the study was prematurely stopped for ethical reasons. Because of the results from WHI, the Women's International Study of long Duration Oestrogen after Menopause (WISDOM) was also prematurely closed. With a mean follow-up of only 11,9 months (ten years planned), they were unable to conclude about long term effects of HT, but confirmed the results of WHI for short term use[14]. Since the WHI report in 2001, large observational studies, i.e. the Million Women Study (MWS)[15] and the Norwegian Women and Cancer study (NOWAC)[16], have confirmed the increased risk of breast cancer associated with HT use.

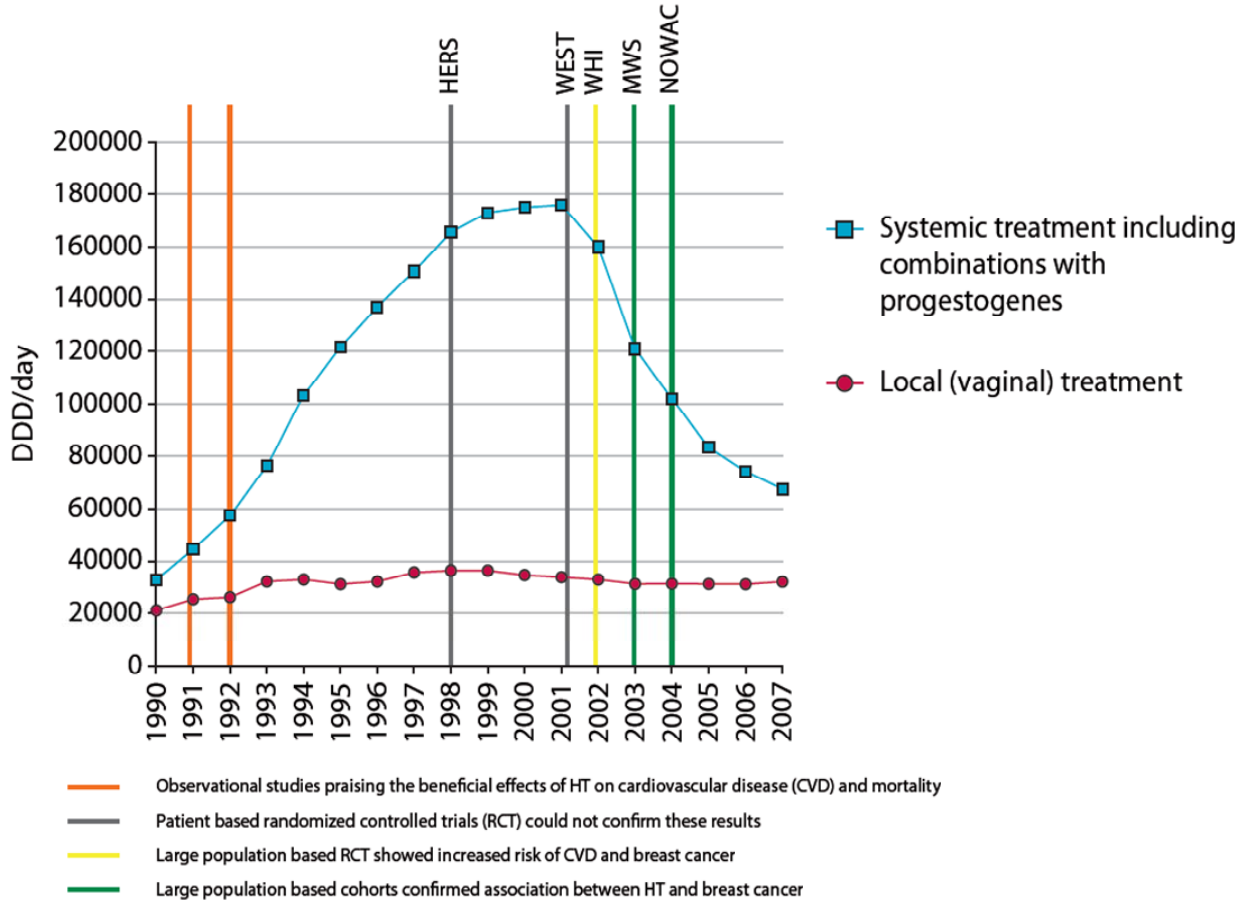

Figure 1.1 Use of HT in Norway 1990-2007[1, 2, 11-13, 15, 16]

The growing evidence of disease risk, particularly breast cancer and CVD, associated with the use of HT has led to altered treatment guidelines in several countries including Norway. HT may be used for relief of vasomotor symptoms (e.g. hot flushes) for the shortest duration and at the lowest effective dose[17-19]. Despite convincingly documented effect on bone resorption, HT is no longer the preferred prophylaxis against osteoporosis for postmenopausal women in Norway[20]. The attitude towards HT use has changed among both patients and physicians and consequently the prescription practice is altered in several countries[21-24].

Attempts to explain the paradoxical discrepancy between the early observational studies and the WHI trial regarding CVD risk has resulted in two main hypotheses; 1) the observational studies were biased (prescriber bias, healthy user bias or misclassification of exposure), or 2) the so called timing hypothesis which proposes that early initiation of HT at the inception of menopause will delay the onset of cardiovascular events, and that the WHI trial participants were too old[25]. The discussion is ongoing.

1.1 Female sex hormones

Biosynthesis

Estrogens and progesterone (P_4) are the main female endogenous sex hormones. Together with low levels of testosterone (T), these steroids are synthesized from cholesterol (Figure 1.2). The biosynthesis takes place mainly in the ovaries, though estrogens are also produced in breast, placenta and adipose tissue. Estrogens include estradiol (E_2), Estrone (E_1) and estriol (E_3), ranked by declining affinity for estrogen receptor binding. Follicle stimulating hormone (FSH) and Luteinizing hormone (LH) constitutes the gonadotropins that stimulate steroid hormone synthesis. Figure 1.3 shows the regulatory system of the Hypothalamus-Pituitary-Ovary axis with its stimulatory and feed-back mechanisms.

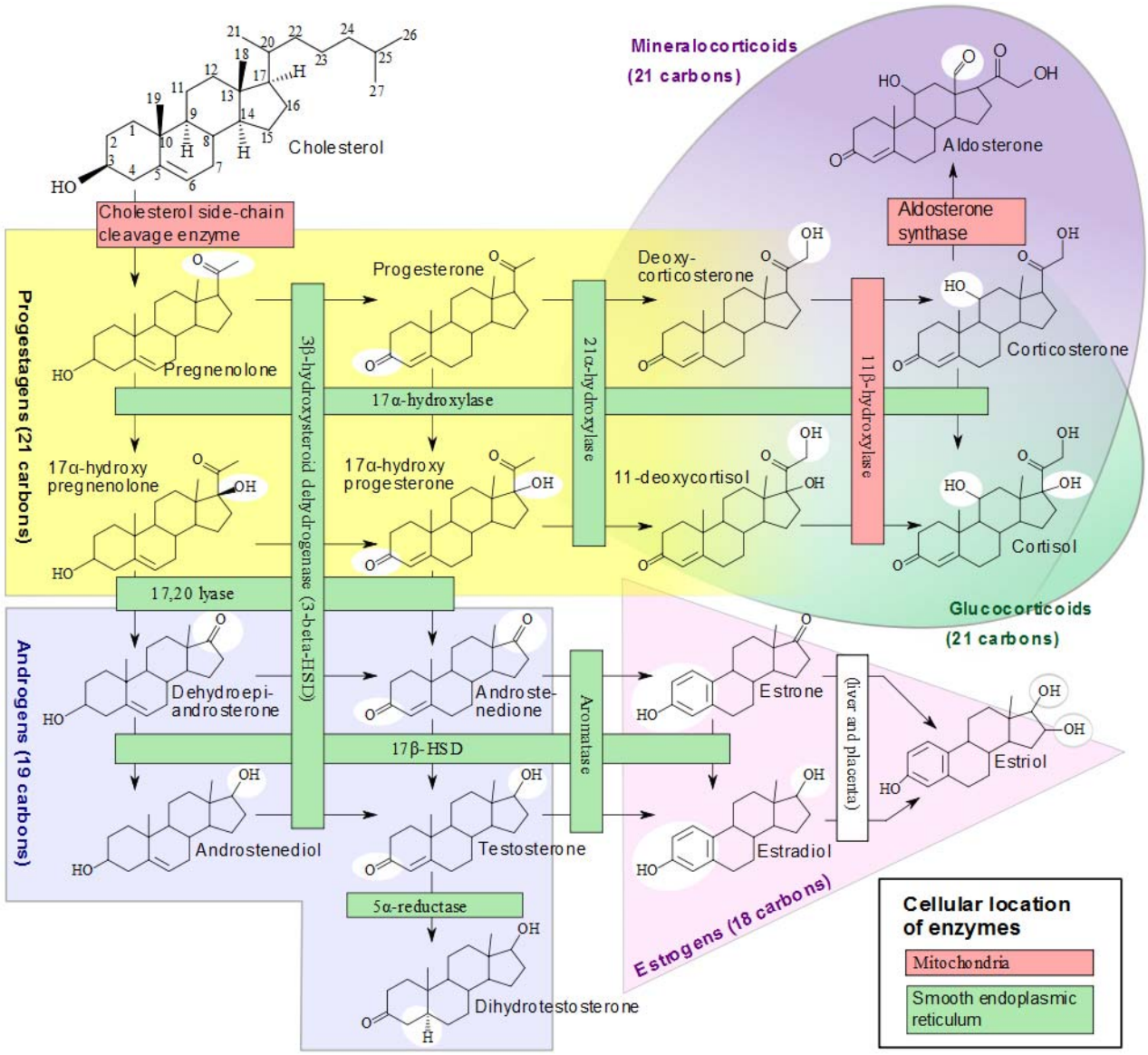

Figure 1.2 Steroid biosynthesis

Enzymes, their cellular location, substrates and products in human steroidogenesis. The major classes of steroid hormones: progesterone, androgens, estrogens, mineralocorticoids and glucocorticoids. HSD: Hydroxysteroid dehydrogenase. (Source: Michael Häggstrøm, http://commons.wikimedia.org/wiki/File:Steroidogenesis.png)

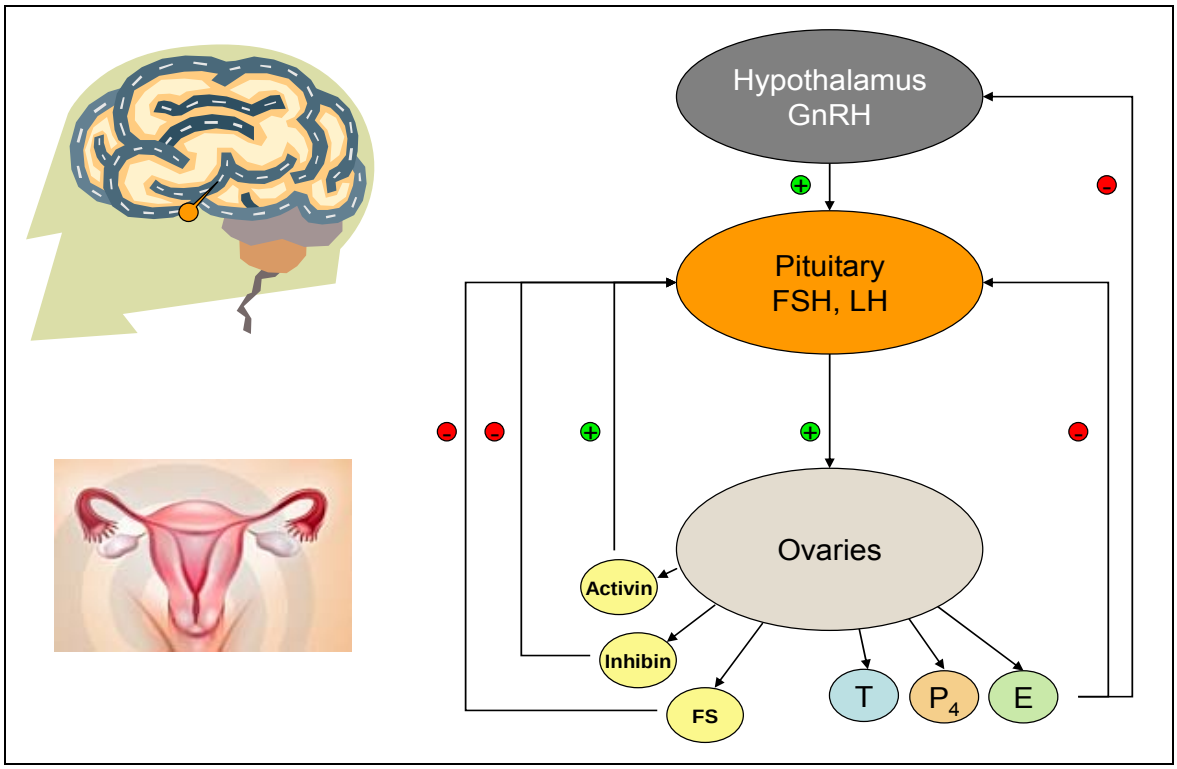

Figure 1.3 The Hypothalamus-Pituitary-Ovary axis

GnRH: Gonadotropin releasing hormone, FSH: Follicle stimulating hormone, LH: Luteinizing hormone, FS: Follistatin, T: Testosterone, P_4: Progesterone, E: Estrogens.

Exogenous female sex hormones

Exogenous female sex hormones, i.e. sex hormones from sources outside the biosynthesis, include oral contraceptives (OC) and postmenopausal HT. HT preparations can be categorized according to hormone content, dose and administration rout. An overview of HT preparations marketed in Norway is shown in Table 1.1. The estrogen component in products for systemic administration is micronised 17-β-estradiol or the valerate ester of 17-β-estradiol. Except for Etifollin®, no products contain etinylestradiol and Etifollin® was withdrawn from the Norwegian market in January 2000. Estriol is mainly used as vaginal products. Progestogens are synthetic steroids with progesterone like effects. The progestogen component in combined products varies (Table 1.1). However, when looking at the most frequently used combinations (Activelle®, Kliogest® and Trisekvens®), norethisterone is by far the most prevalent progestogen. Tibolone is a synthetic steroid with estrogen, androgen and progestogen properties.

Table 1.1 HT preparations marketed in Norway

Hormone content and Administration route	Details	Trade name	Dose E_2/Progestogen*	High/Low estrogen
Etinylestradiol		Etifollin	50 µg/0 µg	H
Estradiol				
Oral		Progynova	1 mg/0 µg	L
		Progynova	2 mg/0 µg	H
		Climen	2 mg/0 µg	H
Transdermal		Climara, Evorel, Estradot and Estraderm	> 50 µg/0 µg	H
			≤ 50 µg/0 µg	L
Vaginal		Vagifem	25 µg/0 µg	L
		Estring	7,5 µg/0 µg	L
Estriol				
Oral		Oestriol	1 and 2 mg /0 µg	L
		Ovesterin	1 and 2 mg /0 µg	L
Vaginal		Ovesterin	0.5 mg/0 µg	L
Estradiol and progestogens combined				
Oral	continuous	Activelle	1 mg/0.5 mg N	L
		Kliogest	2 mg/1 mg N	H
		Climodien	2 mg/2 mg D	H
		Indivina	1 mg/2.5-5 mg M	L
		Indivina	2 mg/5 mg M	H
	sequential	Trisekvens	2 mg (1mg, 6 days)	H
		Trisekvens forte	4 mg (1mg, 6 days)	H
		Cyclabil	2 mg/0.25 mg Lev	H
		Totelle Sekvens	2 mg/0.5 mg T	H
		Novofem	1 mg/1 mg N	L
Transdermal	continuous	Estalis	50 µg/250 µg N	L
	sequential	Estalis sekvens	50 µg/250 µg N	L
		Estracomb	50 µg/250 µg N	L
Tibolone				
Oral		Livial	0/2.5 mg	
Progestogens				
Oral	Medroxyprogesterone	Provera/Perlutex	0/5-10 mg	
	Norethisterone	Primolut-N	0/5 mg	
Vaginal	Progesterone	Crinone	0/8%	
		Progesteron NAF	0/25 mg	

* N=Norethisterone, D=Dienogest, M=Medroxyprogesterone, Lev=Levonorgestrel, T=Trimegaston

Transport

Sex hormone binding globulin (SHBG) is the main transport protein for sex steroid hormones in blood, although steroids also bind to albumin but with lower affinity. Due to their lipophilic structure, the biologically active free fraction of steroid hormones in blood is small.

Metabolism

Steroids are metabolized through reduction/oxidation of functional groups and subsequent conjugation[26, 27]. This two-step process takes place primarily in the liver and renders the metabolites more hydrophilic, thereby facilitating elimination through kidneys and/or bile. Metabolites are generally less active or inactive compared with the original substance. Figure 1.4 shows the main metabolic pathways for estrogens. The intermediate hydroxylation step may occur at several of the carbons in the estrogen molecule, forming a variety of metabolites, although 2- and 16α-hydroxylation are the quantitatively most important pathways.

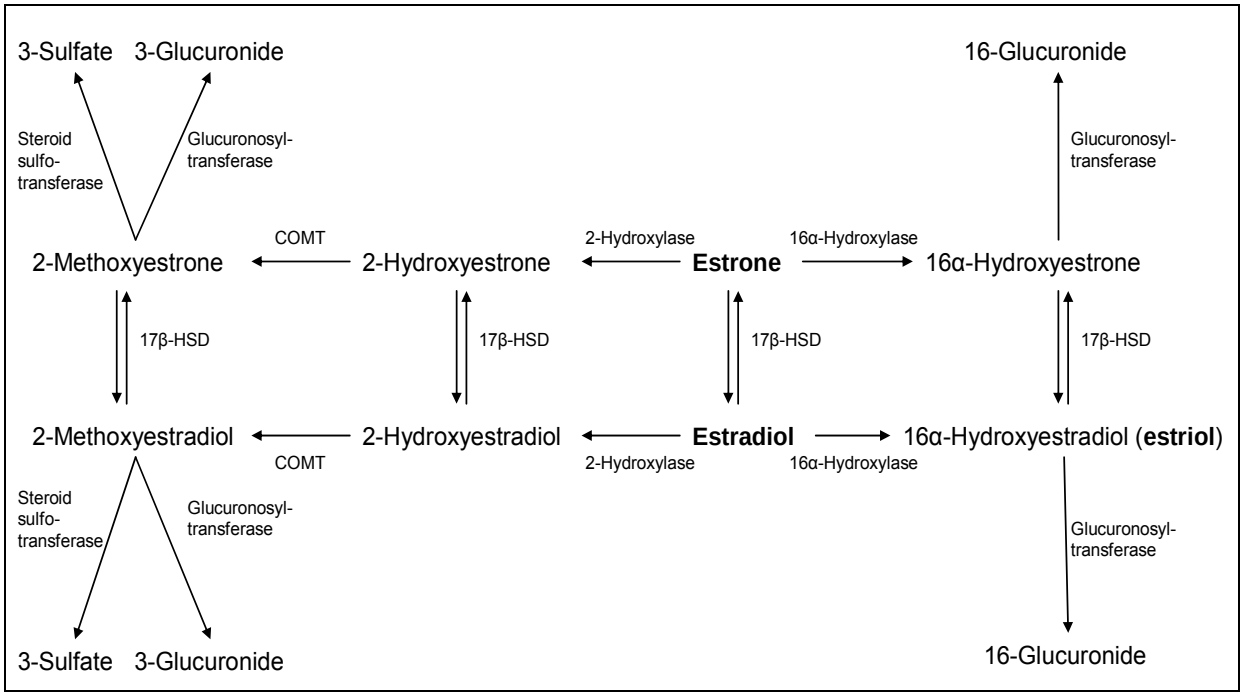

Figure 1.4 Estrogen metabolism

17β-HSD: 17β-Hydroxysteroid-dehydrogenase, COMT: Catechol-O-Methyltransferase

Orally administered steroids undergo extensive first-pass metabolism in the liver before reaching circulation. In other respects metabolism of exogenous sex hormones follows the same pathways as endogenous sex steroids. However, metabolism of synthetic steroids like norethisterone and tibolone will result in "non-biological" metabolites with varying steroid activity. Two of the three tibolone metabolites (3α-hydroxy-tibolone and 3β-hydroxy-tibolone) have estrogen effects while the third (Δ^4-isomer) has progestogen and androgen effects.

Mechanism of action

The effects of sex steroids are mediated through both genomic and non-genomic pathways (Figure 1.5). In the classic genomic action, the steroid ligand binds to its nuclear receptor and the ligand-receptor-complex, in concert with other transcription factors and co-activators or –suppressors, interacts with DNA to initiate or inhibit transcription. Estrogens bind to estrogen receptors (ER α and/or β), Progesterone binds to the progesterone receptor (PR) and Testosterone binds to the androgen receptor (AR). The so called non-genomic actions of sex hormones is mediated through association with ion-channels or G-protein coupled receptors (e.g. estrogen binding GPCR30) with or without binding to membrane associated steroid receptors[28]. Upon ligand binding, membrane associated steroid receptors initiate cytoplasmic signalling pathways (e.g. MAPK, PI3K and AKT)[29] which influence cell functions directly or indirectly through gene transcription. Some of these non-genomic actions of steroid receptors are suggested to cause breast cancer tumours becoming resistant toward estrogen antagonists[29]. The main effects of estrogens are seen in the reproductive organs and bones. However, research into non-genomic steroid pathways, inter alia, has expanded the list of sex hormone target tissues to include the vascular system, central nervous system, gastrointestinal tract, immune system, skin, kidney, and lung[30].

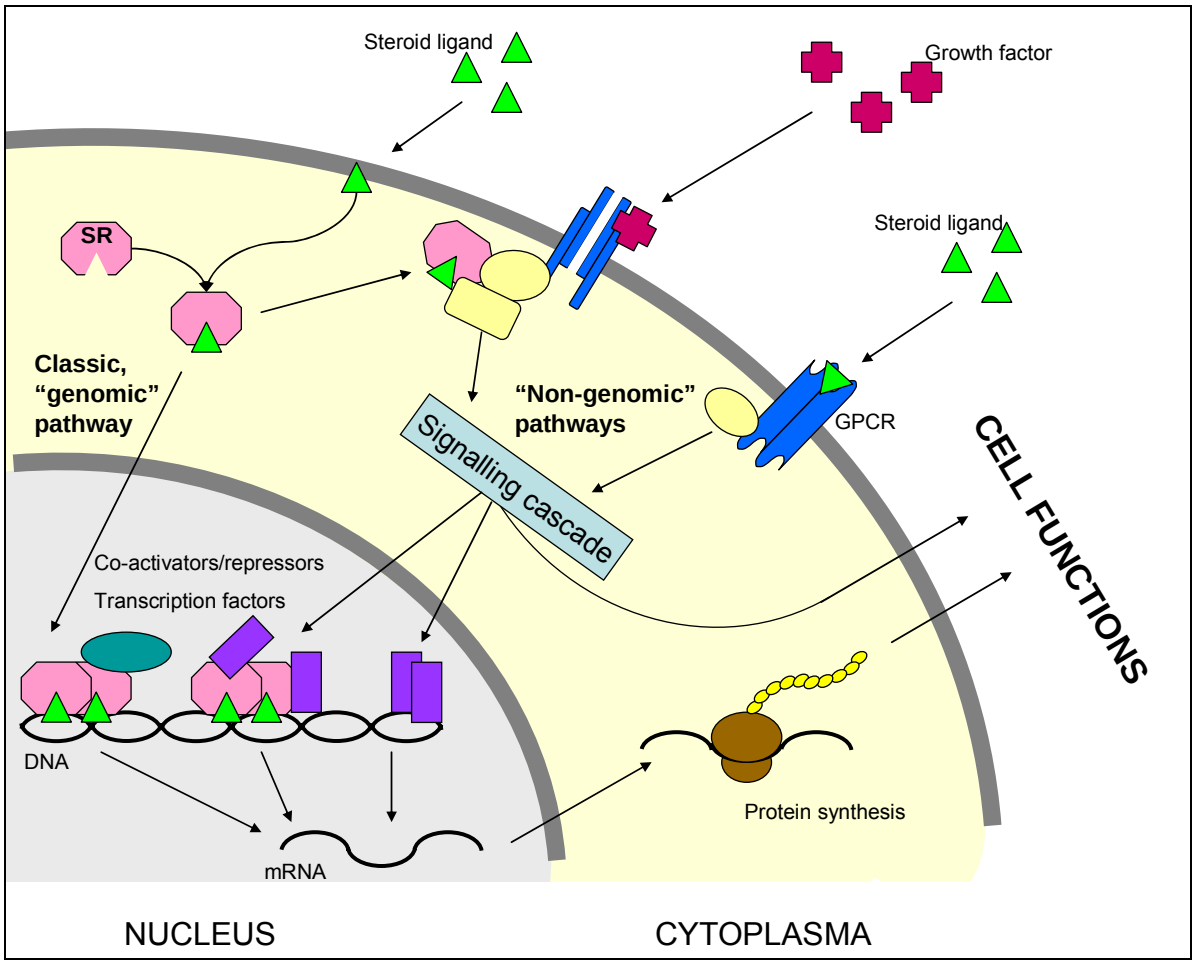

Figure 1.5 Steroid receptor signalling pathways[28, 29]

SR: Steroid receptor, GPCR: G-protein coupled receptor

The menopausal transition

In the menstrual cycle the carefully orchestrated sex hormone surges are regulated through balanced influence of inhibin, activin and follistatin on FSH[31, 32]. Inhibin suppresses FSH secretion while activin has a stimulatory effect on FSH (Figure 1.3). In the aging woman, declining levels of inhibin due to ovarian follicle depletion leads to increasing levels of FSH followed by a temporary increase in circulating E_2[32]. After further depletion of follicles, FSH eventually fails to uphold E_2 levels leading to termination of the menstrual cycle, i.e. menopause.

1.2 Female sex hormones and cancer

Steroid hormones induce growth and development in their target tissues. By causing cell proliferation they are important in the aetiology of hormone related cancers in these tissues. Breast cancer is the most common cancer diagnosis in women; the age adjusted incidence rate in Norway was 72.8 per 100 000 person-years in 2008[33]. The suggestion that estrogen may be a major aetiologic factor in breast cancer carcinogenesis was first published in the early 1970s[34]. Several studies have shown an association between serum or plasma concentrations of steroid hormones and the risk of hormone related cancers like breast cancer[35-38]. Similarly, exogenous hormones has been found to increase the risk of breast cancer[7, 13, 15, 16] and in 2005 the International Agency for Research on Cancer (IARC) classified combined estrogen-progestogen products (OC and HT) as human carcinogens[39].

1.3 Gene expression

With a few exceptions (e.g. mature erythrocytes and thrombocytes) all cells in an individual carry a complete set of the individual's genes. The genes are stored in the DNA molecule (deoxyribonucleic acid) in the cell nucleus. DNA is a double stranded helix in which the two strands are kept together by hydrogen bonds between bases, i.e. base pairs (Figure 1.6). The genetic code is defined by the base sequence, and the complete human genome sequence was first published in 2001[40].

The central dogma of molecular biology describes the transfer of information from DNA to cell function in three steps: transcription, translation and post-translational modification (Figure 1.7). Within this context, gene expression is defined by the abundance and assortment of gene transcripts, i.e. mRNA (messenger ribonucleic acid), in a biologic sample. RNA is a single stranded molecule, a mirror image of the coding sequence of the gene in the DNA molecule. The human genome encompass only just 30 000 genes which are the source of a much larger number of proteins (estimates varies from around 100 000 to several millions).

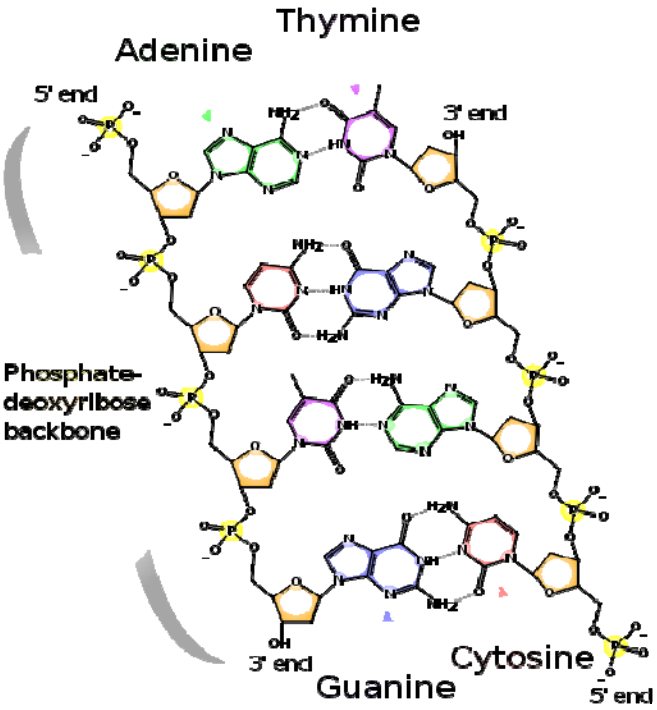

Figure 1.6

Chemical structure of DNA

Dotted lines represent hydrogen bonds between the bases in the two strands. Adenine (A) is always bound to Thymine (T) and guanine (G) to cytocine (C). RNA contains only a single strand in which Thymine is replaced by Uracil (U).

(Source: Madeleine Price Ball, http://commons.wikimedia.org/wiki/File:DNA_chemical_structure.svg)

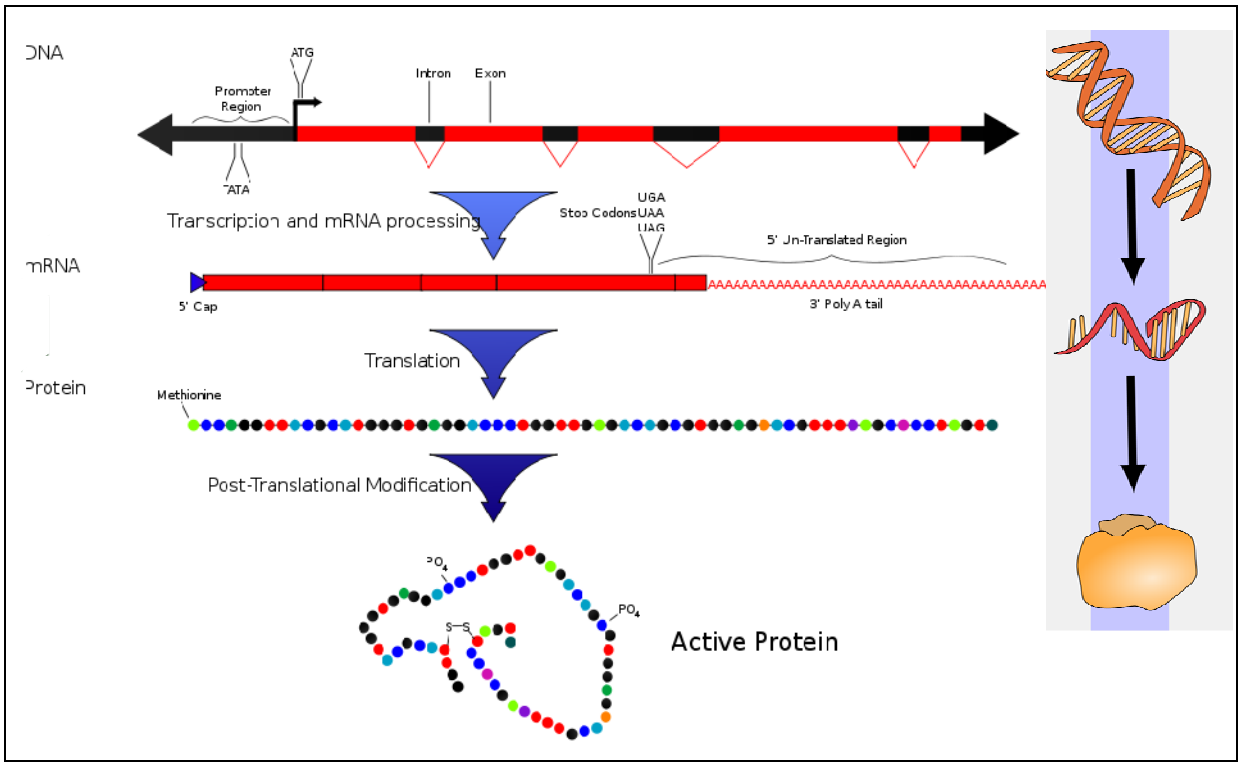

Figure 1.7 Overview of the central dogma of molecular biology

A gene in the double stranded DNA is transcribed into single stranded mRNA, the template for protein synthesis. (Adapted from Mike Jones, http://en.wikipedia.org/wiki/File:Cdmb.svg#file, and "Squidonius", http://en.wikipedia.org/wiki/File:Molbio-Header.svg)

High throughput technology like microarray provides the ability to simultaneously measure the expression of large sets of genes in biologic samples, a process called gene expression profiling or transcriptomics. Microarray is *"…a supporting material (as a glass or plastic slide) onto which numerous molecules or molecular fragments usually of DNA or protein are attached in a regular pattern for use in biochemical or genetic analysis."* (MedlinePlus medical dictionary, http://www.nlm.nih.gov/medlineplus). When measuring RNA, complementary single stranded copies of DNA sequences representing different genes constitute the probes on the microarray. In a full genome scan all human genes (~30 000) are represented on the chip. Purified and labelled RNA from each blood sample is hybridized onto a microarray, i.e. the single stranded RNA molecules in the sample binds to their complementary single stranded DNA copies in the probes by hydrogen bonds (Figure 1.6). Which genes are being expressed and to what extent, will vary according to cell specific and tissue specific function, access to nutrients, environmental exposure, intra- and intercellular signalling, disease, etc. The basic principles of how gene expression is measured by microarray are shown schematically in Figure 1.8.

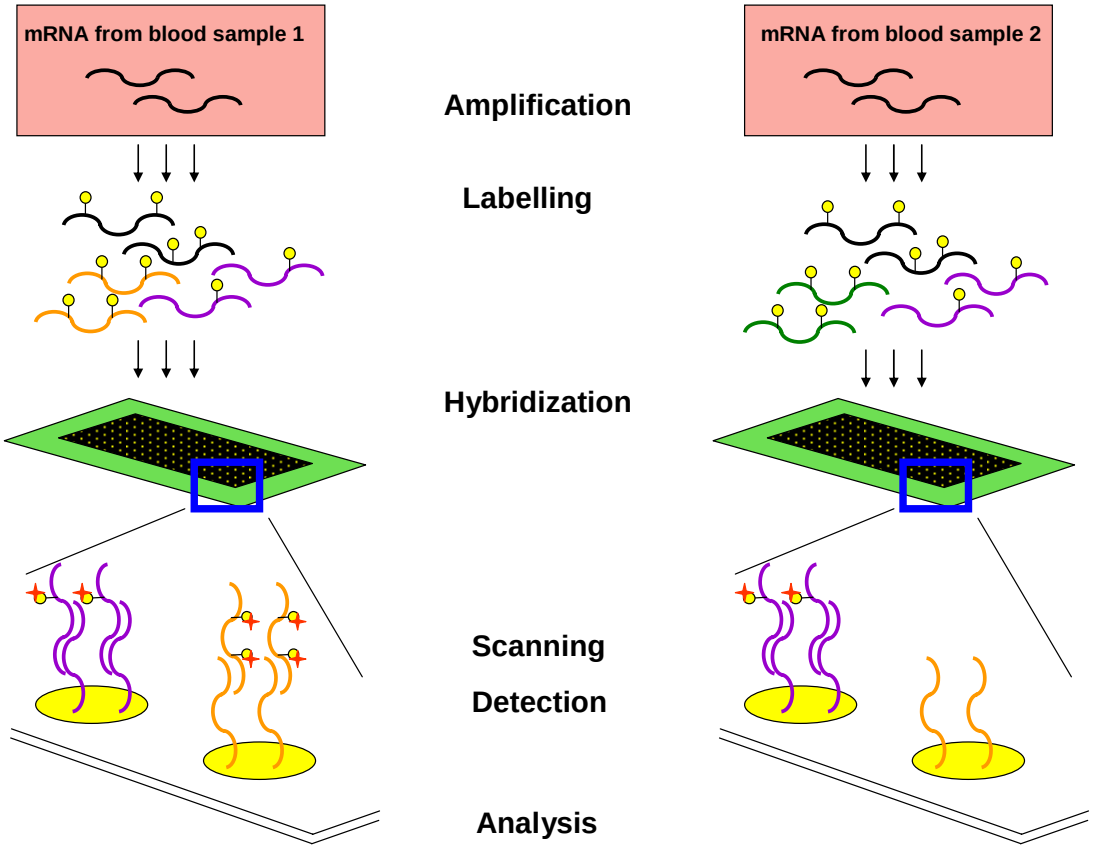

Figure 1.8 Basic microarray principles

Blood sample collection and mRNA isolation precede these steps. Gene expression is defined through measuring the light intensity from each spot on the array (chemiluminescence labelling of mRNA). The microarray analysis thus provides both a qualitative and quantitative measure of gene expression.

1.4 Molecular epidemiology

In classic epidemiologic research the major concern is to find associations between exposure and disease with the aim to provide health improving advice for the public. Less emphasis is put on the aetiological mechanisms behind these associations. This philosophy has been referred to as "black box epidemiology"[41]. Techniques from molecular biology provide complementary information on biologic mechanisms. The application of such techniques within an epidemiological context broadens the potential to answer scientific questions regarding the relationship between risk factors and disease outcome[42, 43]. It can be viewed as

a peek into the “black box”. Terms like “integrative epidemiology” and “systems epidemiology” have been used as labels for such novel epidemiological designs.

Cancer is a condition where the cell growth is out of control. In the established cancer cell a wide spectre of abnormal functions will be mirrored in the gene expression profile and for instance different breast cancer types can be classified according to gene expression analysis of tumour tissue[44]. A gene expression signal will presumably be weaker in peripheral blood than in tumour, and early stage disease will produce a weaker signal than late stage. In cancer research and for breast cancer in particular, there has been great optimism concerning the application of high throughput technology in the development of improved diagnostic and/or prognostic tools in clinical settings[45]. A large proportion of the research in this area so far has been carried out on immortalized cell-lines or tumour biopsies. There has been less focus on normal variation in human gene expression, particularly in large populations. However, this is important for our ability to define what characterises the disease state gene expression. NOWAC is one of very few cohorts in which such research is possible in a population context[46]. Describing the gene expression pattern related to prevalent exposures in a population based study sample provides a basis for future research achievements regarding cancer aetiology and diagnostic/prognostic tests.

A major challenge will be to detect weak signals in peripheral blood. Previous research has shown that it is possible to detect effects on gene expression from physiological, pathological, behavioural and environmental factors in human peripheral blood[47]. Peripheral blood cells are estimated to express about 80% of the genome and over 80% of genes expressed in main organ tissues are also expressed in blood[48]. Blood samples are readily collected in epidemiological studies and the development of RNA preserving blood collection tubes have made possible the collection of both RNA and conventional plasma biomarkers into the NOWAC postgenome biobank.

1.5 Female sex hormones and gene expression

Analysing gene expression in relation to endogenous and exogenous sex hormones might provide information on the link between sex hormones and breast cancer.

Estradiol is by far the most thoroughly investigated female sex hormone in relation to gene expression. However, research in this field is mainly conducted on breast cancer cell lines or tissue, or animals (e.g. genetically modified mice). Knowledge on the influence of normal variation of sex hormone levels on gene expression is practically nonexistent, and with the diversity of tissues expressing steroid receptors and the discovery of non-genomic steroid signalling pathways, predicting the effects of sex hormones on gene expression is exceedingly difficult.

In view of the known effects of sex steroids on target tissues one might expect to see an influence on gene expression related to cell growth in these tissues. Also, sex hormones may be associated with the expression of steroid metabolising enzymes or with proteins involved in the synthesis of other hormones, for instance FSH through feed-back mechanisms (the Hypothalamus-Pituitary-Ovary axis), or transport proteins (SHBG). Groups of genes previously reported to be associated with either female sex hormones or menopausal characteristics would be worth examining in a gene expression analysis. Another important factor, particularly when investigating the blood transcriptome, is the immune system. Cancer affects both the adaptive and innate immune system in various ways[49]. Sex steroids have been investigated regarding influence on blood cell count and function, and there seem to be some effects although a fair amount of data is inconclusive[50].

The explorative nature of this research field suggests a biangular analysis approach, i.e. one should look for new gene signatures but also test previously published signatures for differential expression associated with sex hormones. Also, the work should be viewed as a feasibility study, investigating whether gene expression signatures are at all detectable in a population based setting.

2 Aims

The overall aim of the thesis was to describe changes in the prevalence of self-reported HT use among Norwegian women, to examine the association between HT and endogenous sex hormone levels and to explore the relationship between sex hormones and gene expression through utilizing the NOWAC postgenome biobank.

More specifically:

To describe the changes in patterns of HT use from 1996 to 2005 (paper I)

To describe plasma sex-hormone levels in relation to menopausal status and HT use and thereby validate these two variables (paper II)

To describe the pattern of gene expression relative to endogenous and exogenous sex hormones in postmenopausal women (paper III)

3 Materials and methods

3.1 The Norwegian Women and Cancer study

The Norwegian Women and Cancer Study (NOWAC) is a national population based cohort study with the paramount objective to investigate the relationship between risk factor exposure and cancer, particularly breast cancer[51]. Questionnaire data on lifestyle and health are collected at 4-6 year intervals. Participants, women aged 30-70 years, are randomly drawn from the Central Population Register. Since the start in 1991 about 172,000 women have been enrolled in NOWAC.

Participants have become enrolled in three main steps; 1991, 1995-1997 and 2003-2007 (Figure 3.1, red boxes), with distribution of a second questionnaire from 1998 to 2002 (Figure 3.1, green boxes). Distribution of a third questionnaire was initiated in 2001 (Figure 3.1, yellow boxes).

All questionnaires contain core questions regarding socio-economic status, reproduction, use of exogenous sex hormones, lifestyle (e.g. smoking, alcohol consumption and physical activity), self rated health, familial breast cancer and anthropometric measures (height/weight). Beyond this, the questionnaires may vary between mailings with regards to both length (mainly four or eight pages) and type of questions. Due to the stepwise enrolment and questionnaire diversity, the NOWAC study population consists of several sub samples. The 50,000 women who donated a blood sample from 2002 and onwards constitute the NOWAC postgenome cohort (Figure 3.1, blood droplets).

The studies were approved by The Regional Committee for Medical and Health Research Ethics (REK Nord) and the Norwegian Data Inspectorate.

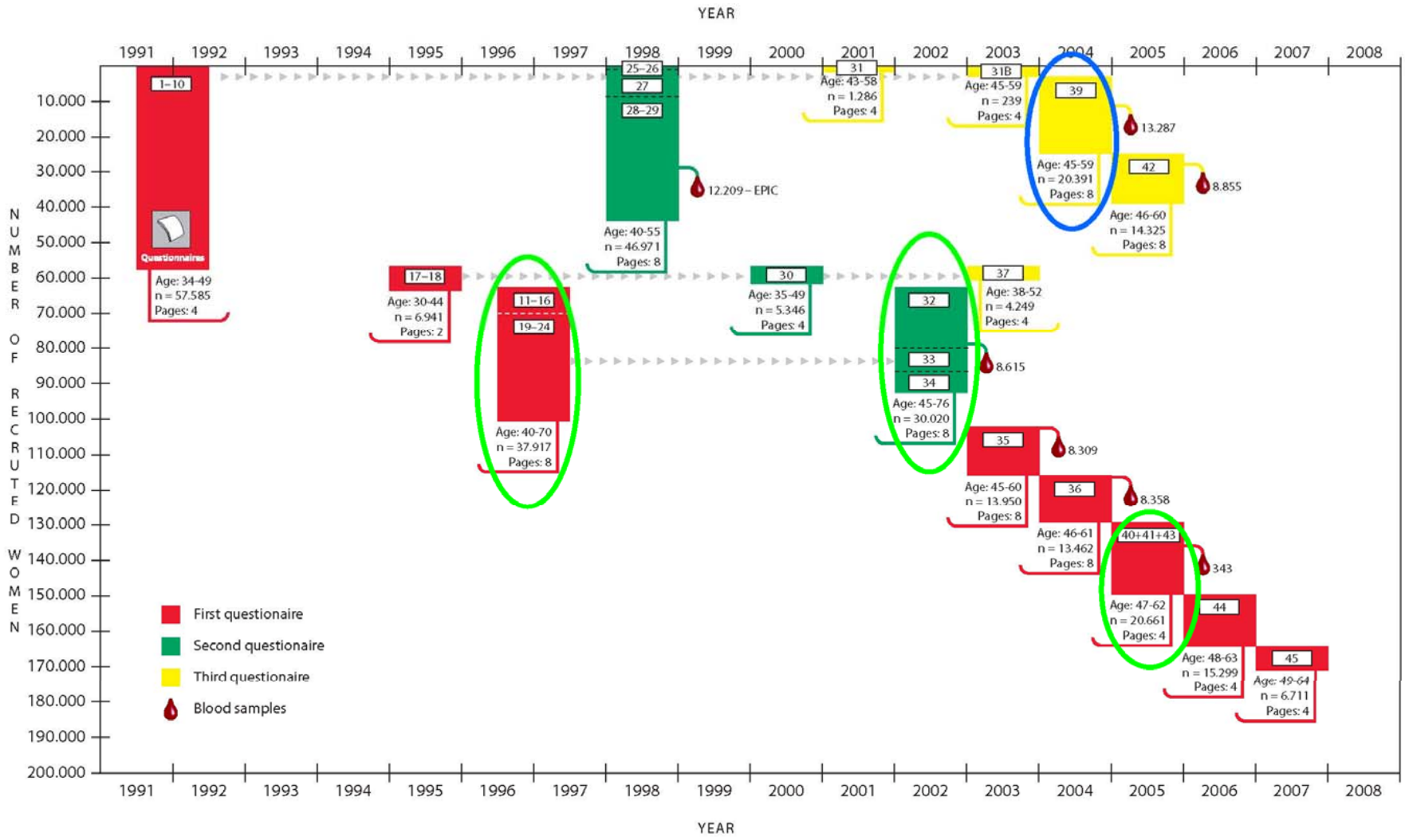

Figure 3.1 Enrolment in the Norwegian Women and Cancer study.

Blood sampling

The mailing of invitations to donate blood was organized in groups of 500. Each woman received a blood collection kit to bring to her local physicians office. Blood was drawn into two blood collection tubes; one for measurements of biomarkers like plasma hormone concentration (collection tube with anticoagulant (citrate)) and one for measurements of gene expression in whole blood (collection tube with RNA stabilizing agent). All blood samples were accompanied by a two-page questionnaire covering menopause, height, weight, present use of medication (including HT) and dietary supplements, as well as variables related to blood draw (hour, posture, fasting etc.).

3.2 Study populations and design

All three papers represent descriptive studies conducted in a cross-sectional design. Paper I is a pharmaco-epidemiological analysis of drug utilization comparing prevalence of HT use at three time points. The study is based on questionnaire data collected in 1995/96 (series 20 and 23), 2002 (series 32 and 33) and 2005 (series 41 and 43), see Figure 3.1, light green circles. The study population was restricted to North Norwegian women 48 to 62 years old to obtain comparable groups.

Papers II and III are based on questionnaires and blood samples from one of the blood collection groups in series 39 (Figure 3.1, blue circle). Paper II is a cross-sectional study where plasma biomarkers are used to validate questionnaire information. Paper III represents an incipient systems epidemiology approach by adding information on gene expression to the phenotype data in paper II.

Copies of the questionnaires that provided the information for these reports as well as the accompanying letters of invitation, information and HT pamphlets can be found in appendix I and II.

3.3 Central variables

Menopausal status

In paper I the women were classified as pre-/perimenopausal or postmenopausal based on their answers to the question “Are your menstrual periods still regular?”. A woman who answered “no” was to give the reason why and the age at which her periods stopped. Women who answered “yes” were classified as pre-/perimenopausal. Women who answered “no” and/or stated a reason for and/or the age at menopause were classified as postmenopausal. Women who were uncertain due to hysterectomy, HT use, disease etc. were classified as postmenopausal if they were 53 years or older, otherwise as unknown menopausal status. In paper II the main categorization was performed on basis of the two-page blood sample questionnaire which does not contain questions on age at menopause or reasons for stopped periods. Women with self-reported irregular menses who were ≥53 years were classified as postmenopausal, otherwise as premenopausal. HT users were included among the postmenopausal women for the comparison of endogenous hormone levels. In paper II plasma concentrations of E2 and FSH were used to validate the menopause classification according to both the blood (two-page) questionnaire and the standard (eight-page) questionnaire. For the remaining analyses in paper II a combination of questionnaire data and plasma concentrations were used to define menopausal status (Figure 2 c in paper II). Women defined as postmenopausal in paper II were included in the microarray analyses in paper III. Microarrays were also run for five women who were originally excluded from paper II due to uncertainty regarding HT use. One of the five was subsequently defined as premenopausal based on plasma hormone concentrations and excluded from paper III.

HT

HT use was defined according to four main features based on questionnaire data:

- Current, former or ever use versus never use
- Administration, i.e. systemic (oral or transdermal) or local (vaginal)
- Content, i.e. hormone substance, estrogen dose (high, low or no estrogen) and regimens (continuous or sequential)

- Duration of use (continuous variable or categorized as <1 year, 1 to <5 years and ≥5 years)

Points 2) and 3) are outlined in Table 1.1 for all products reported used.

Endogenous hormones measurements

Plasma concentrations of estradiol (E_2), progesterone (P_4), testosterone (T), Follicle Stimulating Hormone (FSH) and Sex Hormone Binding Globulin (SHBG) were measured by immunometry at the Department of Medical Biochemistry, University Hospital of North Norway, Tromsø, Norway. In paper II mean plasma concentrations were compared across different categories of HT use and body mass index (BMI). In paper III gene expression was compared between high (fourth quartile) and low (first quartile) hormone concentrations.

Whole blood gene expression

Gene expression was measured by microarray technology, full genome scan, using the Applied Biosystems platform at the Department of Clinical Medicine, the University of Tromsø, Norway. After exclusions based on quality controls etc. the final expression matrix consisted of 285 samples (i.e. women) and 16,185 gene probes. The number of actual genes will be slightly lower because one gene can be represented by more than one probe on the microarray chip. For further details see paper III or Dumeaux et al. (PLoS Genetics 2010)[52]. Gene expression was compared between current HT users and non-users and between high and low hormone concentrations.

3.4 Statistical methods

In paper I and II the analyses was performed using SPSS® for Windows® (Statistical Package for the Social Sciences). The freely available software R version 2.8.1 (http://cran.r-project.org) together with program packages from the Bioconductor project (http://www.bioconductor.org) was used in paper III.

In papers I and II well known statistical methods like logistic regression and analysis of co-variance (ANCOVA) were used to test for significant differences across groups of women.

In paper III differential gene expression was tested through both gene-wise approach (Limma)[53] and gene set analysis (Globaltest)[54]. In the gene-wise approach a linear model is fitted for each gene in the total matrix (n probes = 16,185) assuming no co-variation between genes. The genes found significant in Limma were further inspected for enrichment in functional pathways and networks using web based applications like DAVID (the Database for Annotation, Visualization and Integrated Discovery)[55] and HEFalMp (Human Experimental/Functional Mapper)[56]. In gene set enrichment analysis, a priori defined groups of genes are tested for significant difference in expression between women with high and low sex hormone levels or between HT users and non users. The gene sets represent certain pathways or exposure signatures, e.g. the signatures found in the Limma-analyses in paper III or gene lists found in literature or publicly available databases like KEGG (the Kyoto Encyclopedia of Genes and Genomes)[27]. If the test is significant, the genes in the gene set are, on average, more associated with the response variable than one would expect under the null hypothesis. A subsequent gene plot analysis reveals which genes in the gene set contribute the most to the differential expression, i.e. core genes. Because of the matrix dimensions with number of probes >> number of samples, the analyses was corrected for multiple testing using the false discovery rate (FDR).[57] The FDR is the expected proportion of false rejections among the rejected null hypotheses.

Further details concerning the applied statistical procedures are described in the respective papers.

4 Main Results

4.1 Paper I

HT use increased from 1996 (30.5% current use and 43.3% ever use) to 2002 (38.2% current use and 59.3% ever use), and decreased toward 2005 (14.7% current use and 35.8% everuse). Current HT use among pre-/perimenopausal women was 9.4% in 1996, 13.2% in 2002 and 0.6% in 2005. High-dose combinations were still the most used HT products in 2005, although low-dose estrogen products and tibolone had gained market shares. The mean duration of current HT use was 3.7 years in 1996, 5.8 years in 2002 and 6.3 years in 2005 ($p < 0.05$). In 2005 current HT use was associated with age and menopausal status but not with other formerly known predictors like education length and ever use of oral contraceptives (OC).

4.2 Paper II

Among the postmenopausal women 20% were HT users. Users of systemically-administered E_2-containing HT preparations had plasma E_2 and FSH concentrations comparable to those of premenopausal women. The plasma concentrations of estradiol (E_2) increased with increasing E_2 dose, and use of systemic E_2-containing HT suppressed the concentration of Follicle Stimulating Hormone (FSH). SHBG concentrations increased mainly among users of oral E_2 preparations. Vaginal E_2 application did not influence hormone concentrations.

There was no difference in BMI between HT users and nonusers but increased BMI was associated with increased E_2 and decreased FSH and SHBG concentrations among nonusers.

Menopausal status defined by the two-page questionnaire showed 92% sensitivity (95% CI 89–96%) and 73% specificity (95% CI 64–82%), while the eight-page questionnaire showed 88% sensitivity (95% CI 84–92%) and 87% specificity (95% CI 80–94%). Current HT use showed 100% specificity and 88% of the HT users had plasma E_2 concentrations above the 95% CI of nonusers. Hence, the NOWAC questionnaires provide valid information on current hormone use and menopausal status.

4.3 Paper III

In the attempt to discover new gene signatures, the gene-wise analysis (number of probes = 16 185) did reveal genes significantly associated with the different types of HT; total HT, systemic E_2 or E_2/P, systemic E_2 alone, tibolone and thyroxine (users of other medication excluded). These gene sets represent potential HT expression signatures. According to DAVID and HEFalMp the gene sets do not constitute known gene networks or pathways, except maybe the tibolone gene set. When comparing high and low hormone levels, no noteworthy gene-wise expression difference was found.

The gene set enrichment analysis included 56 gene lists of varying length (number of probes = 3 - 400, see Supplementary Table 1 in paper III). Among these, 22 gene sets were significantly enriched between high and low concentrations of E_2 (HT and thyroxine users excluded). Among them were seven estrogen/HT related gene sets, including our gene list associated with systemic E_2 or E_2/P use, which thereby represents a novel estrogen signature. Seven gene sets were related to immune response. For P_4, 15 gene sets were enriched. Among them were 11 of the gene sets significant for E_2 although with differences in core genes up-regulated in the “high-hormone”-group. The results show that even the low variation in sex hormone levels among postmenopausal women is associated with differences in gene expression.

5 Discussion

Verified by plasma concentrations of sex hormones, the thesis shows that the NOWAC questionnaire information on HT use and menopausal status is valid. The prevalence of HT use shows substantial changes from 1996 to 2005, with a peak in 2002 just before the WHI report was commonly known. The gene expression analyses reveal detectable differential gene expression associated with both endogenous and exogenous sex hormones, illustrated by the finding of a novel estrogen signature.

5.1 Methodological issues

Validity, bias and confounding

All measurements, whether based on questionnaires or plasma samples, are hampered by varying degrees of random and systematic error. The influence of random error on estimates of association between variables is generally controlled through the inclusion of a sufficiently large number of observations. Bias (systematic error) occurs when systematic differences in how study participants are selected or choose to participate (selection bias), or in the way information is collected from the participants (information bias) affect the analysis estimates and conclusions[58]. Bias can only be minimized through careful study design and conduct.

Validity may be defined as absence of systematic error[58]. Internal validity refers to whether the study inferences also pertain to the source population. External validity refers to whether the results are generalizable to people outside the source population. In a population based random sample of study participants like NOWAC selection bias is not a major issue for the internal validity, but self-selection might affect the generalizability of the study. Information bias, i.e. misclassification of participants into exposure or outcome categories due to insufficient methods for data collection, might represent a problem.

Confounding occurs if the association between explanatory and response variables is actually caused by a third factor which is associated with both. Such confounding factors are usually accounted for through adjustments in the statistical models, but this is only possible

if the factor is measured in the study. Potential confounding factors which were assumed to make an impact was adjusted for in the various analyses. However, there may be residual confounding in the models and there may also be unknown and unmeasured factors that could influence the results, particularly in relation to gene expression.

Bias in cross-sectional analyses is usually discussed with regards to the relationship between exposure and disease. The focus of this thesis is not to measure disease risk, but rather to describe the prevalence of factors known to cause disease. Thus, problems like incidence-prevalence bias or length-biased sampling are not relevant issues. However, misclassification of participants and generizability to external populations deserve consideration.

Misclassification (information bias)

Misclassification bias results from incorrect determination of exposure (or outcome) in observational studies, i.e. the participant is classified into the wrong category. In a cohort study like NOWAC potential misclassification bias will be largely non-differential since all participants are measured equally through the same questionnaire or blood collection kit.

Exposure to HT use is not a constant characteristic and misclassification of subjects may lead to misinterpretation of potential associations between HT use and other variables. Since information on HT use is collected in the same way for the three cross-sections in paper I, it should be comparable. The most reliable exposure variable would be current HT use. To help the women recall former use of HT, the NOWAC questionnaires are accompanied by a pamphlet containing photos of the HT preparations marketed in Norway since 1953. However, former HT use will still be less reliable because general awareness of potential benefit or harm from the therapy might influence the aptitude to recall previous use. The conclusions from paper I are thus mainly based on the analyses of current use and paper II confirms the validity of this exposure variable.

If associations between HT use and gene expression is dependent on previous exposure or duration of exposure, the variable “current use” might be impropriate for the analysis in paper III. On the other hand, brief effects of HT on gene expression might have been missed since time since last dose were not taken into consideration. This is however an

unlikely influence since continuous HT use results in steady state hormone concentrations. Still, the cross-sectional design favours detection of longer-lasting associations.

Menopausal status is a central variable when investigating sex hormones. A hysterectomized woman might think she is postmenopausal since her menstrual periods have stopped, but unless she is also oophorectomized, her hormone production may be intact. Sequential HT might give an impression of premenopausal status due to monthly bleedings while continuous HT might suppress monthly bleedings in a woman who is not yet postmenopausal. Women may also be uncertain of their menopausal status due to use of other medications, diseases etc that interfere with their menstruation cycle. In NOWAC, women with uncertain menopausal status who are ≥53 years old are classified as postmenopausal because about 90% of the women above 52 years report that their menstrual periods have stopped. The reason for applying this method for menopausal status classification is to be able to utilize as much of the cohort information as possible in the analyses. Preferably, women should not be classified as postmenopausal unless they have experienced at least 12 months amenorrhea after the final menstrual period. However, this would exclude a fair amount of women with unknown status. In paper II the method for menopausal status classification was validated using plasma concentrations of E_2 and FSH as the gold standard. In reality, the only way to accurately identify menopause is in retrospect because there is no adequate independent biomarker[59]. In want of this information, the hormone concentration cut-offs, defined as the clinical limits used by the laboratory, was the best available alternative. It is disputed whether a single measurement of FSH can give sufficient evidence of menopause because of individual variation in hormone levels[60]. The use of both E_2 and FSH concentrations increases the confidence in the hormonal classification. Paper II shows that the classification method is valid for non-users of HT. HT users had to be excluded from the analysis, seeing as their endogenous hormone levels would reflect HT use and not menopausal status. However, there is no reason to omit application of the menopause classification on this group. They are not biologically different from non-users.

For the remaining analyses in paper II the questionnaire and hormonal classification methods were combined to be further assured of the postmenopausal status. For instance,

an oophorectomized woman will invariably be postmenopausal irrespective of hormonal levels. The study population in paper III was based on this combined classification.

The reduction in HT use since 2002 shown in paper I will make future menopausal classification easier since the proportion of women with uncertain menopausal status will decrease accordingly. On the other hand, future research into different aspect regarding HT will demand larger study groups to gain necessary power to detect potential effects on for instance gene expression.

External validity

The generalizability in NOWAC is generally secured through random selection of participants, using the Central Population Register, and reasonably high response rates. Still, the participants are self-selected and they might differ from non-respondents in important characteristics and result in self-selection bias, or non-response bias. However, the important question is not whether non-responders differ from the responders but whether the responders differ from the source population and, if so, whether this difference is associated with the factors under study, e.g. HT use. For instance socioeconomic status has been found to differ between responders and non-responders. If there is an association between socioeconomic status and HT use, and if women with low socioeconomic status are underrepresented in the study sample, the prevalence of HT use might be overestimated. A multi-angular evaluation of the external validity of the NOWAC study (data from 1991 to 1996) revealed only minor differences between responders and the total sample of women at a 57% response rate[61]. Moreover, increasing the response rate from 50% to 70% by sending a reminder to non-responders did not change the distribution for OC use, fertility or education length. Hence, in paper I, the 58% response rate among first time participants in 1996 should not compromise the validity of the study. We are equally confident about the validity of the 2005 data (66% response rate), particularly given the close similarity between HT-use in NOWAC and HT-prescriptions in NorPD. The 2002 data, representing second time response, lowers the response rate further (45%), but the close similarity to national sales figures supports a sound validity.

The generalizability might be a larger problem for the study sample in papers II and III which comprise third time participants. An unpublished comparison of this study sample

with a sample of first time participants in the same age group (48-62 years) who donated blood during the same month in 2005 showed that the main difference was a higher education level among the first time participants. This is not surprising, as the third time participants answered this question in 1991. Some women might have taken further education after their first questionnaire (age range in 1991: 34-48 years). There was also a difference in duration of HT use in that the first time participants had used HT longer than the third time participants. There were no differences in household income, smoking pattern, OC use, self-reported health or use of medication between these two NOWAC samples. Based on these results combined with the random selection of participants and high response rate (89% in this group, 74% overall in the NOWAC postgenome cohort), the study sample in papers II and III is considered to represent the source population sufficiently.

In addition to self-selection, the exclusion of certain participants from the analysis limits the generalizability. The conclusions from papers II and III are for instance not necessarily valid for pre- or perimenopausal women.

Technological considerations

In an epidemiologic collection of large amounts of data, some information will be lost due to the need for pragmatic collection protocols. Traditionally, this concerns questionnaire composition, length and collection intervals. The building of the NOWAC biobank in later years has extended this problem to include limitations with regards to amount and quality of for example blood samples. There is a trade-off between the strength in numbers and extensive blood collection protocols. Furthermore, it is a challenge to measure as many biomarkers as possible from as little blood as possible.

In paper II it would have been desirable to also measure estrone, androstenedione, dehydroepiandrosterone (DHEA) and particularly norethisterone acetate, the by far most frequently used progestogen on the Norwegian HT market. This was however not feasible with the available amount of plasma. The immunometric assays used to measure endogenous hormone levels are developed primarily for clinical applications, not for research, and citrate plasma is not the best sample matrix. Moreover, the limited variation in sex hormone concentrations among postmenopausal women necessitates a sensitive assay. Mass spectrometry would be preferable for future analyses. On the other hand, potential

variation found with more sensitive assays might not make a difference for the hormonal influence on gene expression. In paper III we compared gene expression in the high and low quartile of hormone concentration, and these groups would hardly change with more sensitive hormone assays. In retrospect, a differentiation into bound and free fraction of steroid hormones may also have added valuable information. For instance a classification into high and low levels of free fraction E_2 might have resulted in a different gene expression pattern from the one found in paper III.

The microarray technology was originally developed for laboratory research with very limited numbers of samples. It is not accommodated to analyse large epidemiologic blood collections. The time needed to perform the analysis and the limited blood sampling standardization introduces technical variability due to variation in sample preparation and processing. In the paper III study population three technical variables (array lot number, RNA extraction date and time between blood draw and storage) explained 46.5% while six biological/exposure variables (age, BMI, fasting, smoking, HT and medication) explained 8.1% of the overall variation in gene expression[52]. Thus, the level of random noise seems to be high. Although the analyses were adjusted for the above mentioned technical variables in addition to standard normalization, the ability to find subtle differences in gene expression is somewhat limited.

Intra-individual variation, i.e. normal variation within each woman, is not taken into consideration. This is also due to limitations in the study protocol, since it does not allow for more than one sample per participant. However, it is reasonable to assume that between-individual variation exceeds the intra-individual variation sufficiently, as has also been shown by previous publications[62, 63]. A recent report suggests that within-individual gene expression measured by microarray analysis remain stable over one month while a small percentage of genes varies over three months[64].

5.2 Interpretation and context

The thesis represents descriptive research in a cross-sectional epidemiologic design. Due to the simultaneous collection of data, the temporal sequence of events is not known and associations between explanatory and response variables cannot be inferred as aetiological

relationships. The work provides insight into the changing prevalence of HT use, plasma sex hormone levels and gene expression in a representative cohort of middle aged women.

The findings in paper I confirm sales figures and suggest that in 2005 climacteric complaints were the sole reason for HT use, in accordance with revised national guidelines. The main concern is the enduring long-term use. Since 2005, HT consumption in Norway has decreased further according to both sales figures[65] and the Norwegian Prescription Database (NorPD)[66]. The 2009 NorPD report shows that among women >45 years the total number of new users per year remained stable (~12 000 women) from 2005 to 2008[66]. Since the total prevalence of women filling HT prescriptions decreased, this means that more women are ceasing treatment. This pattern differ from the one found in paper I where the decrease toward 2005 in prevalence of current HT use was largest in the younger age groups. Hopefully this indicates a decline from the extensive long term use of HT in 2005 seen in paper I, together with increase in yearly individual risk-benefit assessment recommended by national guidelines[18].

Vasomotor symptoms are common in women during the menopausal transition, the prevalence ranging between 40 and 80% in different studies[67]. For some of these women the climacteric complaints affect their quality of life and HT is unsurpassed as the most effective treatment for these symptoms[68, 69]. For those who choose HT, continued research into the effects of this medication is warranted. In later years the main focus for research on the association between HT use and breast cancer has turned from whether HT causes cancer to details regarding the association; differentiation between types of HT, tumour characteristics, prognosis, survival and implications for mammography screening. Unpublished preliminary results from a case-control study nested within the NOWAC cohort show an odds ratio of about 2.4 (N=660) for the association between current HT use and breast cancer, thereby confirming previous results regarding risk of breast cancer[16]. NOWAC data have also shown that the breast cancer risk associated with use of HT is higher among former OC users[70]. Several reports suggest that the decline in HT consumption after 2002 have already resulted in decreased breast cancer incidence[71-73]. This is supported by results from the WHI trial[74]. However, in many countries the rise and fall in HT use coincides with the initiation of mammography screening programs in the same age group. The contribution

of each of these two factors in the decreasing breast cancer incidence trends is difficult to disentangle[75]. Another WHI report shows that HT use seriously influences the sensitivity of mammography as a test for breast cancer[76]. HT should therefore be considered an important confounder in studies evaluating the effect of mammography screening on breast cancer incidence since it is associated with both "treatment" and disease. So far it is not documented that a temporary cessation of HT for a few weeks or months before mammography reduces the HT influence on the test; the effects seem rather prolonged[76, 77].

In the attempt to elucidate the relationship between sex hormones and cancer in a population based context, gene expression profiling shows promising results so far according to the paper III research pilot. HT is associated with statistically significant gene expression profiles and, whether or not women are HT users, their gene expression profiles are associated with plasma concentrations of E_2 and P_4. However, in addition to the challenges related to technology and the handling of vast amounts of data, the very nature of gene expression complicates the interpretations[78]. There are several biological steps between the gene transcript and the resulting active protein. Factors like alternative splicing of transcripts, influence of microRNA and post-translational modifications, all result in a protein diversity which is not detectable through transcriptomics alone. Additionally, epigenetic modifications of DNA (e.g. DNA methylation and histone acetylation) influence on which genes are open for transcription, and represent a pre-transcriptional step in the central dogma of molecular biology (Figure 1.7). By including measurements of for instance microRNA expression future studies will add an important dimension to gene expression analysis. Nevertheless, transcriptomics alone may indicate which biological pathways are active in the woman at the moment of blood draw. Together with the extensive co-variation between genes, this is why the focus of the gene expression analysis in paper III is to detect gene signatures or networks, and not specific genes.

According to the study population in papers II and III, over 50% of middle aged women use some sort of medication, from mild over-the-counter analgesics to immunosuppressants, while about 25% are current smokers. Both general medication use and smoking were found to be associated with gene expression[52]. If gene expression research should ultimately result in diagnostic or prognostic tests for cancer, there will

invariably be a considerable proportion of medication users and smokers among the patients eligible for testing. A reliable gene expression signal should thus be independent of potential effects of medication and smoking. Medication users and smokers were therefore not excluded from the comparison of gene expression between high and low endogenous hormones.

The HT related gene sets from the Limma-analysis were included in the gene set enrichment analysis to see if they would be differentially expressed between high and low endogenous hormone levels. HT and thyroxine users were therefore excluded. A second analysis without exclusions was also performed. This did not materially change the ranking of the gene sets except for the HT gene sets. However, the p-values and FDR increased somewhat. Including more samples should intuitively add to the power, but perhaps the inclusion of these particular samples increased the random noise. Although the analyses seem reasonably stable, the gene expression pattern found in paper III needs confirmation through analysis of an independent data set. This is important both for the interpretation of results from paper III, but also for the reliability of future gene expression analysis on material from the NOWAC biobank.

The inclusion of gene expression analysis in an epidemiologic design is expected to add to the knowledge on mechanisms for disease development, i.e. carcinogenesis. The transcriptome is viewed as an intermediary functional factor in the causal relationship between exposure and outcome, e.g. between sex hormones and breast cancer. If the HT related gene expression profiles from paper III are confirmed in an independent data set, they might be used as markers of HT exposure and as tools for validating HT variables. A future diagnostic test based on gene expression should include gene signatures that accurately identify the diseased individuals in a population regardless of their exposure status. Describing the gene expression pattern related to different exposures might thus provide the ability to account for potential confounding signals in future attempts to extract the disease specific influence on gene expression[79].

6 Conclusions and perspectives

Through a multiangular approach and by use of a broad spectre of analytical methods this thesis describes female sex hormones among postmenopausal women in terms of consumption, plasma levels and gene expression.

The results show a rise and fall in postmenopausal HT use which parallels the sales figures, and plasma hormone concentrations reveal that the NOWAC questionnaires provide valid information on menopause and HT use. Although the gene expression signatures found require confirmation, the work shows that endogenous and exogenous sex hormones are associated with significant gene expression patterns, and that gene expression profiling is feasible in an epidemiological setting.

Future perspectives

HT consumption and risk profile

HT use and female hormone dependent cancer incidence should be monitored further to investigate to what degree the changing patterns of HT use influences cancer incidence. This needs to be assessed in relation to the mammography screening programme which was initiated in parallel to the increase in HT use in Norway.

Some postmenopausal women suffer vasomotor symptoms to an extent that severely affects their quality of life. The reluctance toward HT use might increase the search for alternative, non-hormonal treatments of vasomotor symptoms. Such alternative treatment strategies have so far not proven equivalent effect to HT and are not necessarily safer than HT. It would thus be prudent to investigate which alternatives to HT are used and to what extent, along with potential health effects, among women with climacteric symptoms.

Different HT regimens might have different risk-benefit ratios. Of particular interest is the impact of progestogens. Several reports suggest that the progestogens might be the main culprit in causing breast cancer, and that different progestogens carry different risks[80, 81]. Finding the HT with the lowest risk level, yet with retained effect, would be helpful for women who need treatment for their vasomotor symptoms.

Sex hormones and gene expression

To confirm the associations found between exogenous hormones and gene expression, the gene sets should be tested in an independent, and preferably larger, data set. Improved technology with increased processing standardization and shorter analysis period will hopefully result in data sets with less noise, which renders it possible to detect subtle differences in gene expression with improved precision.

Additional fields where further research is desirable:

- Potential effects on gene expression by different types of medication.
- Gene expression analyses within a prospective epidemiologic design may improve cancer diagnosis and treatment through development of tests.
- Gene expression analysis in a nested case-control design might elucidate the aetiological relationship between sex hormones and hormone dependent cancers.
- The collection of breast biopsies in the NOWAC biobank will enable validation of peripheral blood as a surrogate measure of breast tissue gene expression.
- The transcriptomics research could be extended to include measurements of influencing factors like microRNA and epigenetics (e.g. histone acetylation and DNA methylation). This might improve the biological interpretation.
- Plasma concentrations of IGF-1 and IGFBP-3 were measured in the study sample from papers II and III and could be explored in relation to both questionnaire data and gene expression.

8 References

1. Stampfer MJ, Colditz GA. Estrogen replacement therapy and coronary heart disease: A quantitative assessment of the epidemiologic evidence. *Preventive Medicine* 1991;20(1):47-63.

2. Grady D, Rubin SM, Petitti DB, et al. Hormone Therapy To Prevent Disease and Prolong Life in Postmenopausal Women. *Ann Intern Med* 1992;117(12):1016-37.

3. Effects of estrogen or estrogen/progestin regimens on heart disease risk factors in postmenopausal women. The Postmenopausal Estrogen/Progestin Interventions (PEPI) Trial. The Writing Group for the PEPI Trial. *JAMA* 1995;273(3):199-208.

4. Eriksen EF, Kassem M, Langdahl B. European and North American experience with HRT for the prevention of osteoporosis. *Bone* 1996;19(5, Supplement 1):179S.

5. Whitehead MI, McQueen J, Beard RJ, Minardi J, Campbell S. The Effects of Cyclical Oestrogen Therapy and Sequential Oestrogen/Progestogen Therapy on the Endometrium of Post-Menopausal Women. *Acta Obstetricia et Gynecologica Scandinavica* 1977;56(s65):91-101.

6. Furness S, Roberts H, Marjoribanks J, Lethaby A, Hickey M, Farquhar C. Hormone therapy in postmenopausal women and risk of endometrial hyperplasia. *Cochrane Database of Systematic Reviews* 2009(2).

7. Collaborative Group on Hormonal Factors in Breast C. Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52.705 women with breast cancer and 108.411 women without breast cancer. *The Lancet* 1997;350(9084):1047-59.

8. Petitti D, Perlman JA, Sidney S. Postmenopausal estrogen use and heart disease. *N Engl J Med* 1986;315(2):131-2.

9. Vandenbroucke JP. How much of the cardioprotective effect of postmenopausal estrogen is real? *Epidemiology* 1995;6(3):207-8.

10. Sogaard AJ, Tollan A, Berntsen GKR, Fonnebo V, Magnus JH. Hormone replacement therapy: Knowledge, attitudes, self-reported use -- and sales figures in Nordic women. *Maturitas* 2000;35(3):201-14.

11. Hulley S, Grady D, Bush T, et al. Randomized Trial of Estrogen Plus Progestin for Secondary Prevention of Coronary Heart Disease in Postmenopausal Women. *JAMA* 1998;280(7):605-13.

12. Viscoli CM, Brass LM, Kernan WN, Sarrel PM, Suissa S, Horwitz RI. A Clinical Trial of Estrogen-Replacement Therapy after Ischemic Stroke. *N Engl J Med* 2001;345(17):1243-9.

13. Writing Group for the Women's Health Initiative. Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women: Principal Results From the Women's Health Initiative Randomized Controlled Trial. *JAMA* 2002;288(3):321-33.

14. Vickers MR, MacLennan AH, Lawton B, et al. Main morbidities recorded in the women's international study of long duration oestrogen after menopause (WISDOM): a randomised controlled trial of hormone replacement therapy in postmenopausal women. *BMJ* 2007bmj.39266.425069.AD.

15. Million Women Study Collaborators. Breast cancer and hormone-replacement therapy in the Million Women Study. *The Lancet* 2003;362(9382):419-27.

16. Bakken K, Alsaker E, Eggen AE, Lund E. Hormone replacement therapy and incidence of hormone-dependent cancers in the Norwegian Women and Cancer study. *Int J Cancer* 2004;112(1):130-4.

17. U.S. Preventive Services Task Force*. Postmenopausal Hormone Replacement Therapy for Primary Prevention of Chronic Conditions: Recommendations and Rationale. *Ann Intern Med* 2002;137(10):834-9.

18. Jacobsen S, Harg P. Hormonbehandling av kvinner med menopausale symptomer [Hormonal treatment of women with symptoms of menopause]. *Nytt om legemidler* 2003;26(2):3-4.

19. Recommendations on the need for revision of the CPMP points to consider on hormone replacement therapy. *Committee for medicinal products for human use (CHMP), The European Agency for the Evaluation of Medicinal Products (EMEA)* 2004;CHMP/EWP/1738/04.

20. Jacobsen S, Harg P. Hormonsubsitutsjonsbehandling ikke lenger førstevalg ved osteoporose/osteoporoseprofylakse [Hormone substitution is no longer the preferred treatment of osteoporosis/osteoporosis prophylaxis]. *Nytt om legemidler* 2003;26(7):10.

21. Hersh AL, Stefanick ML, Stafford RS. National Use of Postmenopausal Hormone Therapy: Annual Trends and Response to Recent Evidence. *JAMA* 2004;291(1):47-53.

22. Moen MH, Nilsen S-T, Iversen O-E. A significant change in Norwegian gynecologist's attitude to hormone therapy is observed after the results of the Women's Health Initiative Study. *Acta Obstet Gynecol Scand* 2005;84(1):92 - 3.

23. Hallquist AC, Moen MH, Andrew M, Aursnes I. Attitudes to hormone therapy during menopause among Norwegian women [English summary]. *Tidsskr Nor Laegeforen* 2006;126(9):1195-7.

24. Thunell L, Milsom I, Schmidt J, Mattsson L-A. Scientific evidence changes prescribing practice - a comparison of the management of the climacteric and use of hormone replacement therapy among Swedish gynaecologists in 1996 and 2003. *BJOG: An International Journal of Obstetrics & Gynaecology* 2006;113(1):15-20.

25. Phillips LS, Langer RD. Postmenopausal hormone therapy: Critical reappraisal and a unified hypothesis. *Fertil Steril* 2005;83(3):558-66.

26. Stanczyk FZ, Bretsky P. Biosynthesis, transport, and metabolism of steroid hormones. In: Henderson BE, Ponder B, Ross RK, eds. Hormones, genes, and cancer. New York: Oxford University Press, Inc.; 2003:12-37. ISBN:0-19-513576-8

27. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucl Acids Res* 2000;28(1):27-30.

28. Boonyaratanakornkit V, Edwards DP. Receptor Mechanisms Mediating Non-Genomic Actions of Sex Steroids. *Semin Reprod Med* 2007;25(03):139-53.

29. Fox EM, Andrade J, Shupnik MA. Novel actions of estrogen to promote proliferation: Integration of cytoplasmic and nuclear pathways. *Steroids* 2009;74(7):622-7.

30. Wierman ME. Sex steroid effects at target tissues: mechanisms of action. *Advan Physiol Edu* 2007;31(1):26-33.

31. Hurwitz JM, Santoro N. Inhibins, Activins, and Follistatin in the Aging Female and Male. *Semin Reprod Med* 2004;22(3):209-17.

32. Stenchever MA, Droegemueller W, Herbst AL, Mischell DR. Chapter 42 Menopause; Endocrinology, Consequences of Estrogen Deficiency, Effects of Hormonal Replacement Therapy, Treatment Regimens. In: Stenchever MA, ed. Comprehensive Gynecology. 4 ed. St.Louis, Mo: Mosby Inc.; 2001:1217-45. ISBN: 0-323-01402-x

33. Bray F, ed. Cancer in Norway 2008. Oslo: The Cancer Registry of Norway; 2009. ISBN: 978-82-90343-76-0

34. MacMahon B, Cole P, Brown J. Etiology of human breast cancer: a review. *J Natl Cancer Inst* 1973;50(1):21-42.

35. Key TJ, Appleby P, Barnes I, Reeves G. (The Endogenous Hormones and Breast Cancer Collaborative Group) Endogenous Sex Hormones and Breast Cancer in Postmenopausal Women: Reanalysis of Nine Prospective Studies. *J Natl Cancer Inst* 2002;94(8):606-16.

36. Kaaks R, Rinaldi S, Key TJ, et al. Postmenopausal serum androgens, oestrogens and breast cancer risk: the European prospective investigation into cancer and nutrition. *Endocr Relat Cancer* 2005;12(4):1071-82.

37. Kaaks R, Berrino F, Key T, et al. Serum Sex Steroids in Premenopausal Women and Breast Cancer Risk Within the European Prospective Investigation into Cancer and Nutrition (EPIC). *J Natl Cancer Inst* 2005;97(10):755-65.

38. Eliassen AH, Missmer SA, Tworoger SS, et al. Endogenous Steroid Hormone Concentrations and Risk of Breast Cancer Among Premenopausal Women. *J Natl Cancer Inst* 2006;98(19):1406-15.

39. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, volume 91, Combined Estrogen–Progestogen Contraceptives and Combined Estrogen–Progestogen Menopausal Therapy. Lyon: International Agency for Research on Cancer; 2005. ISBN 978 92 832 1291 1

40. Initial sequencing and analysis of the human genome. *Nature* 2001;409(6822):860-921.

41. Weed DL. Beyond black box epidemiology. *Am J Public Health* 1998;88(1):12-4.

42. Lund E, Dumeaux V. Systems Epidemiology in Cancer. *Cancer Epidemiol Biomarkers Prev* 2008;17(11):2954-7.

43. Spitz MR, Bondy ML. The evolving discipline of molecular epidemiology of cancer. *Carcinogenesis* 2010;31(1):127-34.

44. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature* 2000;406(6797):747-52.

45. Chatterjee SK, Zetter BR. Cancer biomarkers: knowing the present and predicting the future. *Future Oncology* 2005;1(1):37-50.

46. Dumeaux V, Borresen-Dale AL, Frantzen JO, Kumle M, Kristensen VN, Lund E. Gene expression analyses in breast cancer epidemiology: the Norwegian Women and Cancer postgenome cohort study. *Breast Cancer Res* 2008;10(1):R13.

47. Mohr S, Liew C-C. The peripheral-blood transcriptome: new insights into disease and risk assessment. *Trends in Molecular Medicine* 2007;13(10):422-32.

48. Liew C-C, Ma J, Tang H-C, Zheng R, Dempsey AA. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. *Journal of Laboratory and Clinical Medicine* 2006;147(3):126-32.

49. de Visser KE, Eichten A, Coussens LM. Paradoxical roles of the immune system during cancer development. *Nat Rev Cancer* 2006;6(1):24-37.

50. Bouman A, Heineman MJ, Faas MM. Sex hormones and the immune response in humans. *Hum Reprod Update* 2005;11(4):411-23.

51. Lund E, Dumeaux V, Braaten T, et al. Cohort Profile: The Norwegian Women and Cancer Study - NOWAC - Kvinner og kreft. *Int J Epidemiol* 2008;37(1):36-41.

52. Dumeaux V, Olsen KS, Nuel G, Paulssen RH, Børresen-Dale A-L, Lund E. Deciphering Normal Blood Gene Expression Variation - The NOWAC Postgenome Study. *PLoS Genet* 2010;6(3):e1000873.

53. Smyth GK. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Stat Appl Genet Mol Biol* 2004;3(1):Article 3.

54. Goeman JJ, van de Geer SA, de Kort F, van Houwelingen HC. A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics* 2004;20(1):93-9.

55. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protocols* 2008;4(1):44-57.

56. Huttenhower C, Haley EM, Hibbs MA, et al. Exploring the human genome with functional maps. *Genome Res* 2009;19(6):1093-106.

57. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Series B Stat Methodol* 1995;57(1):289-300.

58. Rothman KJ, Greenland S. Modern epidemiology. 2 ed. Philadelphia: Lippincott Williams & Wilkins; 1998. ISBN 0-316-75780-2

59. Sherman. Defining the menopausal transition. *Am J Med* 2005;118(12):3-7.

60. Burger H. The menopausal transition - Endocrinology. *J Sex Med* 2008;5(10):2266-73.

61. Lund E, Kumle M, Braaten T, et al. External validity in a population-based national prospective study--the Norwegian Women and Cancer Study (NOWAC). *Cancer Causes Control* 2003;14(10):1001-8.

62. Cobb JP, Mindrinos MN, Miller-Graziano C, et al. Application of genome-wide expression analysis to human health and disease. *Proc Natl Acad Sci USA* 2005;102(13):4801-6.

63. van Erk MJ, Blom WA, van Ommen B, Hendriks HF. High-protein and high-carbohydrate breakfasts differentially change the transcriptome of human blood cells. *Am J Clin Nutr* 2006;84(5):1233-41.

64. Karlovich C, Duchateau-Nguyen G, Johnson A, et al. A longitudinal study of gene expression in healthy individuals. *BMC Medical Genomics* 2009;2(1):33.

65. Rønning M, ed. Drug Consumption in Norway 2004-2008. Oslo: The Norwegian Institute of Public Health 2009. ISBN 978-82-8082-196-6 (printed version), ISBN 978-82-8082-198-0 (electronic version)

66. The Norwegian Prescription Database 2004–2008. Oslo: The Norwegian Institute of Public Health; 2009. ISBN: 978-82-8082-317-5 printed version, ISBN: 978-82-8082-318-2 electronic version

67. Nelson HD. Menopause. *The Lancet* 2008;371(9614):760-70.

68. MacLennan A, Broadbent J, Lester S, Moore V. Oral oestrogen and combined oestrogen/progestogen therapy versus placebo for hot flushes. Edited (no change to conclusions), published in Issue 1, 2009. *Cochrane Database Syst Rev* 2004;Issue 4.

69. Pinkerton JV, Stovall DW, Kightlinger RS. Advances in the treatment of menopausal symptoms. *Women's Health* 2009;5(4):361-84.

70. Lund E, Bakken K, Dumeaux V, Andersen V, Kumle M. Hormone replacement therapy and breast cancer in former users of oral contraceptives - The Norwegian Women and Cancer study. *Int J Cancer* 2007;121(3):645-8.

71. Ravdin PM, Cronin KA, Howlader N, et al. The Decrease in Breast-Cancer Incidence in 2003 in the United States. *N Engl J Med* 2007;356(16):1670-4.

72. Marshall SF, Clarke CA, Deapen D, et al. Recent breast cancer incidence trends according to hormone therapy use: the California Teachers Study cohort. *Breast Cancer Res* 2010;12(1):R4.

73. Kumle M. Declining breast cancer incidence and decreased HRT use. *The Lancet* 2008;372(9639):608-10.

74. Chlebowski RT, Kuller LH, Prentice RL, et al. Breast Cancer after Use of Estrogen plus Progestin in Postmenopausal Women. *N Engl J Med* 2009;360(6):573-87.

75. Pelucchi C, Levi F, La Vecchia C. The rise and fall in menopausal hormone therapy and breast cancer incidence. *The Breast* 2010;In Press, Corrected Proof.

76. Chlebowski RT, Anderson G, Pettinger M, et al. Estrogen Plus Progestin and Breast Cancer Detection by Means of Mammography and Breast Biopsy. *Arch Intern Med* 2008;168(4):370-7.

77. Buist DSM, Anderson ML, Reed SD, et al. Short-Term Hormone Therapy Suspension and Mammography Recall. *Ann Intern Med* 2009;150(11):752-65.

78. Walker MS, Hughes TA. Messenger RNA expression profiling using DNA microarray technology: Diagnostic tool, scientific analysis or un-interpretable data? (Review). *Int J Mol Med* 2008;21(1):13-7.

79. Lund E, Dumeaux V. Towards a more functional concept of causality in cancer research. *Int J Mol Epidemiol Genet* 2010;1(2):124-33.

80. Bakken K, Fournier A, Lund E, et al. Menopausal hormone therapy and breast cancer risk: Impact of different treatments. The European prospective investigation into cancer and nutrition (EPIC). Accepted in:. *Int J Cancer* 2010.

81. Fournier A, Berrino F, Clavel-Chapelon F. Unequal risks for breast cancer associated with different hormone replacement therapies: results from the E3N cohort study. *Breast Cancer Res Treat* 2008;107103-11.