UNIVERSITY OF TROMSØ UIT

FACULTY OF HEALTH SCIENCES
DEPARTMENT OF COMMUNITY MEDICINE

Blood gene expression, lifestyle and diet

The Norwegian Women and Cancer Post-genome Cohort

Karina Standahl Olsen

A dissertation for the degree of
Philosophiae Doctor

2013

Blood gene expression, lifestyle and diet

The Norwegian Women and Cancer Post-genome Cohort

Karina Standahl Olsen

Department of Community Medicine
Faculty of Health Sciences
University of Tromsø

Tromsø, Norway
2013

Acknowledgements

The work presented here was conducted at the Department of Community Medicine (ISM), Faculty of Health Sciences, University of Tromsø, from the autumn of 2006 to the spring of 2013. The work was sponsored by a university grant. Many people contributed more or less directly and at different stages to the completion of this thesis. I am grateful to all of you.

My main supervisor Eiliv Lund: Thank you for sharing your ideas and visions, and for making a (back then) young PhD student feel that her opinion and knowledge matters. Your guidance during these years, and your patience and understanding during hard times has been invaluable.

My co-supervisor Ruth H. Paulssen: Thank you for taking the time to talk about molecular biology, life, kids, and cats. I am also grateful for your comments to the thesis drafts.

Co-authors: your contributions and insightful comments have been highly valued, and your knowledge in your respective fields of research have taught me a lot.

Administrative staff at NOWAC and ISM, especially Bente, Merete, and Mona: thank you for fixing everything that needs to be fixed. Your magic makes life at ISM go around!

My office mates over the years: First, Kristin S.: I am so glad to have had you as my skillful and knowledgeable partner in crime when it comes to all things lab-related. Our long talks on all aspects of lab & life have meant a lot to me, you are a gem. Also, Marit, Oxana, Lotta and Sara: thank you for insights into life and science, and for listening to spontaneous outbursts of joy and sorrow as the ups and downs of life as a PhD student have unfolded.

Other colleagues, especially Kristin B., Guri, Toril, Tor Gisle, Bent-Martin, Kristin (snipis!), and many more: thank you for making walking up and down the hallways at ISM all day,

every day, a very nice activity, and for being the friendly faces one needs to see every now and then. The work-out group: thank you for the inspiration to postpone any potential musculoskeletal consequences of long hours by the pc.

It should be noted that I never had the intention of working with fatty acids: their biology is just too much to handle. However, my dear friend Caroline encouraged me to join in the research group of Svanhild Schønberg, to do my MSc together with her. Svanhild introduced me to the PUFAs, and she has been an inspiration. So has Caroline, who sparked my enthusiasm for microarrays. I am grateful to you both.

Friends outside work, especially my girls in Trondheim and Tromsdalen, as well as Reidun, Lene, Yngvild, and Ingeborg: you are all unique in your own special ways and you add great value to my life. Also, to my yoga teacher Torgunn: your classes are a weekly source of inspiration.

My family and family-in-law: without your undivided love and support my life would have been very different, and this project would never have come to a completion. In particular, thank you, *mamma*, for always encouraging me and being there for me.

Finally, to my dear husband Kenneth: I am deeply grateful for your patience and efforts during my years as a PhD student, and especially during the completion of my thesis. *Du e min klippe*. Mia: *mamma sitt gull*, thank you for putting everything into perspective, just by being here.

Karina
Tromsø, June 2013

Table of Contents

Acknowledgements i
Summary 3
Sammendrag 4
List of papers 5
List of figures and tables 6
Abbreviations 7
1 Introduction 8
1.1 Lifestyle and nutrition in the etiology of human disease 8
1.1.1 Historical aspects and potential for disease prevention 8
1.1.2 Tools for research on nutrition and lifestyle 10
1.1.3 Fish consumption 11
1.2 Fatty acids 13
1.2.1 Structure and metabolism 13
1.2.2 Sources and recommendations 15
1.2.3 Epidemiological findings 17
1.2.4 Molecular mechanisms 20
1.3 Vitamin D 22
1.3.1 Structure and metabolism 22
1.3.2 Sources and recommendations 23
1.3.3 Epidemiological findings 25
1.3.4 Molecular mechanisms 26
1.4 Gene expression analysis 28
1.4.1 Background 28
1.4.2 Challenges 30
1.5 Human beings as a model system 32
1.5.1 Blood as a target tissue 32
2 Aims 34
3 Material and methods 35
3.1 The Norwegian Women and Cancer Study 35
3.1.1 The NOWAC Post-genome Cohort 35
3.2 Gene expression profiling 36
3.3 Biomarker measurements: fatty acids and vitamin D 37
3.4 Statistical methods 37
4 Results – Summary of papers 40

4.1 Paper I 40
4.2 Paper II 41
4.3 Paper III 41
4.4 Paper IV 42
5 Discussion - Methodological considerations 43
5.1 Study design 43
5.2 External validity - selection bias 43
5.3 Internal validity - information bias 45
5.4 Gene expression analyses 45
5.5 Biomarkers of nutritional status 53
5.5.1 Fatty acid measurements 54
5.5.2 25(OH)D measurements 56
6 Discussion - Main results 58
6.1 Aspects of the NOWAC Post-genome Cohort 59
6.1.1 Use of human samples 59
6.1.2 Systems-level thinking 59
6.1.3 The agnostic approach 60
6.2 Associations between gene expression profiles and exposures 60
6.2.1 General tendencies in gene expression profiles of healthy women 61
6.2.2 Gene expression profiles associated with fatty acid ratios 62
6.2.3 Gene expression profiles associated with vitamin D status 64
6.2.4 Converging pathways 65
6.3 Pieces of a puzzle 66
7 Conclusions 67
8 Future aspects 68
8.1 Improved ways of exploring gene expression variation 68
8.2 Approaching the globolomic design 70
9 References 71

Summary

Lifestyle and diet are risk factors for cancer, cardiovascular and inflammatory diseases, but the molecular mechanisms involved have not been fully elucidated. High-throughput technologies like microarrays are being used to explore those mechanisms, but data from model systems (cell cultures and research animals) may not reflect human biology as well as previously thought.

The aim of this PhD thesis was to explore the feasibility of using blood samples from a population-based cohort to generate gene expression profiles that enable identification of molecular mechanisms related to several lifestyle and dietary factors.

The Norwegian Women and Cancer (NOWAC) study is a nationally representative, prospective cohort started in 1991 to investigate risk factors for breast cancer. The study includes more than 172 000 women who have answered questionnaires regarding lifestyle, diet, and health. In 2003-06, a subset of women in NOWAC were randomly drawn to constitute the NOWAC Post-genome Cohort, and approx. 50 000 blood samples eligible for gene expression analysis were collected, along with an additional blood sample and a questionnaire. For this thesis, full-genome gene expression microarrays, as well as plasma concentrations of fatty acids and vitamin D were analyzed in 500 samples from the Post-genome Cohort. For paper I, gene expression related to major lifestyle-related exposures (smoking, medication use), inter-individual variation (age, body mass index, fasting status), and technical factors (pre-analytical and analytical) was explored. Paper II and III took a nutrigenomics approach to explore gene expression related to vitamin D and fatty acid ratios. Due to the partially overlapping dietary source (fatty fish), the co-variability of vitamin D and fatty acids were analyzed in paper IV.

Technical variability influenced the gene expression profiles considerably, pointing to the need for rigorous pre-analytical and analytical control. In addition, inter-individual and exposure variables were associated with differential gene expression, both at the gene and pathway level. Vitamin D status was associated with modest gene expression differences. However, at the pathway level, several immuneregulatory processes were identified, supporting the emerging hypotheses of vitamin D as an immuneregulatory nutrient. The ratios of different polyunsaturated fatty acids of the n-6 and n-3 families were associated with both gene- and pathway-level differences, which pointed to cellular and molecular mechanisms influenced by dietary fat. The concentrations of vitamin D and the investigated fatty acid ratios were only weakly associated, indicating that the identified gene expression profiles likely arise from the unique influence of the two factors separately.

In conclusion, the thesis demonstrates that several exposures related to lifestyle and diet, as well as technical variability, are mirrored in blood gene expression profiles. This implies that gene expression analysis using blood samples may become an important tool for identification of molecular mechanisms related to many exposures, for characterization of pathological processes related to cancer and other diseases, and for prediction of disease risk.

Sammendrag

Livsstil og kosthold er risikofaktorer for kreft, kardiovaskulære og inflammatoriske sykdommer, men hvilke molekylære mekanismer som er involvert er ikke fullt ut beskrevet. Teknologi som microarray (mikromatriser) brukes i stor grad for å utforske disse mekanismene, men resultater fra modellsystemer som cellekulturer og forsøksdyr reflekterer ikke nødvendigvis menneskets biologi så godt som man tidligere har trodd.

Målet med denne doktorgradsavhandlingen var å utforske muligheten for å bruke blodprøver fra en populasjsonsbasert kohorte, for å analysere genuttrykksprofiler og dermed identifisere molekylære mekanismer relatert til ulike livsstils- og kostholdsfaktorer.

Kvinner og Kreft-studien er en nasjonalt representativ, prospektiv kohorte som ble startet i 1991, for å forske på risikofaktorer for brystkreft. Studien inkluderer i dag mer enn 172 000 kvinner, som har svart på spørreskjema angående livsstil, kosthold, og helse. I 2003-06 ble ca. 50 000 av deltakerne tilfeldig plukket ut for å delta i Kvinner og Kreft Post-genom kohorten. Fra disse kvinnene ble det samlet inn blodprøver med mulighet for genuttrykksanalyse, samt en standard blodprøve og et spørreskjema. Til denne avhandlingen ble det brukt materiale fra 500 kvinner i Post-genom kohorten. Fullgenom genuttryksanalyser ble uført, i tillegg til måling av konsentrasjon av fettsyrer og vitamin D i plasma. I artikkel I ble genuttrykk relatert til livsstilseksponeringer (røyking, medisinbruk), inter-individuell variasjon (alder, kropsmasseindeks, fastestatus), og tekniske faktorer (pre-analytiske og analytiske) utforsket. I artikkel II og III ble genuttrykk relatert til vitamin D og fettsyreratioer analysert, og disse artiklene kan sies å være en del av forskningsfeltet "nutrigenomics". På grunn av deres delvis overlappende kostholdskilde, fet fisk, ble samvariasjon mellom vitamin D og fettsyreratioer analysert i artikkel IV.

Teknisk variasjon påvirket genuttrykksprofilene i stor grad, noe som peker på behovet for tett kontroll av preanalytiske og analytiske prosedyrer. Inter-individuelle faktorer og ulike eksponeringer var assosiert med differensielt uttrykk både av enkeltgener og av signalspor. Vitamin D-status var assosiert med beskjedne forskjeller i genuttrykk. Likevel fant vi, på signalspornivå, flere immunologiske prosesser. Dette støtter hypotesen om at vitamin D kan være immunregulatorisk. Ratioen mellom flerumettede fettsyrer tilhørende n-6 eller n-3-familiene var assosiert med differensielt uttrykk av både enkeltgener og signalspor, og pekte ut prosesser som kan bli påvirket av ulike typer fett fra kosten. Konsentrasjonene av vitamin D og fettsyreratioer var kun svakt assosiert med hverandre, noe som indikerer at genuttrykksprofilene oppstår som følge av uavhengig påvirkning fra disse to kostholdsfaktorene.

Funnene i avhandlingen viser at ulike typer eksponeringer relatert til livsstil og helse, i tillegg til teknisk variasjon, gjenspeiles i genuttrykksprofiler i blod. Dette antyder at genuttrykksanalyser i blodprøver kan bli et viktig verktøy for å identifisere molekylære mekanismer relatert til ulike typer eksponeringer, for å karakterisere sykdomsprosesser, og for å forutsi risiko for sykdom.

List of papers

This thesis is based on the following papers, hereafter referred to by their roman numerals.

Paper I

Dumeaux V, Olsen KS, Nuel G, Paulssen RH, Borresen-Dale AL, Lund E (2010). **Deciphering normal blood gene expression variation--The NOWAC postgenome study.** *PLoS Genet* 6(3): e1000873. PubMed ID: 20300640

Paper II

Olsen KS, Rylander C, Brustad M, Aksnes L, Lund E (2013)**. Plasma 25 hydroxyvitamin D level and blood gene expression profiles: a cross-sectional study of the Norwegian Women and Cancer Post-genome Cohort.** *Eur J Clin Nutr*. Advance online publication 6 March 2013. PubMed ID: 23462941

Paper III

Olsen KS, Fenton C, Frøyland L, Waaseth M, Paulssen RH, Lund E (2013). **Plasma fatty acid ratios affect blood gene expression profiles – A cross-sectional study of the Norwegian Women and Cancer Post-genome Cohort.** *PLOS ONE* [accepted]

Paper IV

Olsen KS, Aksnes L, Frøyland L, Lund E, Rylander C. **Vitamin D status and PUFA ratios in a national representative cross-section of healthy middle-aged, Norwegian women** [manuscript]

List of figures and tables

Figure 1: Overview and examples of biomarkers 10

Figure 2: The structure of alpha-linolenic acid, 18:3n-3 13

Figure 3: Overview of PUFA metabolism and eicosanoid production 15

Figure 4: Major pathways of fatty acid production, transport, and metabolism 21

Figure 5: Overview of vitamin D metabolism 23

Figure 6: The structure of calcitriol ($1,25(OH)_2D$) 23

Figure 7: Central dogma of molecular biology 29

Figure 8: Basic principle of hybridization-based microarrays 30

Figure 9: Mechanisms by which altered supply of fatty acids could affect immune responses 62

Figure 10: The globolomic design of the NOWAC Post-genome Cohort, compared to a traditional prospective study design. 70

Table 1: Reports on aspects of nutrition and lifestyle 11

Table 2: Recommended and achieved intake of fat and PUFAs 16

Abbreviations

AA	Arachidonic acid
ABCG1	ATP-binding cassette sub-family G member 1
ABI	Applied Biosystems
AICR	American Institute of Cancer Research
ALA	Alpha-linolenic acid
BMI	Body mass index
COX	Cyclooxygenase
CV	Coefficient of variation
CVD	Cardio-vascular diseases
DBP	Vitamin D binding protein
DHA	Docosahexaenoic acid
DNA	Deoxyribonucleic acid
EPA	Eicosapentaenoic acid
FA	Fatty acid
FABP4	Fatty acid binding protein 4
FAO	Food and Agriculture Organization of the United Nations
GO	Gene Ontology
GSEA	Gene Set Enrichment Analysis
HPLC	High-performance liquid chromatography
IL	Interleukin
IOM	Institute of Medicine
KEGG	Kyoto Encyclopedia of Genes and Genomes
LA	Linoleic acid
LOX	Lipoxygenase
MAQC	Microarray Quality Control project
mRNA	Messenger ribonucleic acid
MSigDB	Molecular Signatures Database
NF-κB	Nuclear factor κB
NOWAC	Norwegian Women And Cancer study
PI3K	Phosphatidylinositol-3-kinase
PLC	Phospholipase C
PPAR	Peroxisome proliferator-activating receptor
PUFA	Polyunsaturated fatty acid
qRT-PCR	Quantitative real-time reverse transcription polymerase chain reaction
RNA	Ribonucleic acid
RXR	Retinoid x-receptor
TLR	Toll-like receptor
TNF	Tumor necrosis factor
UVB	Ultraviolet B
VDR	Vitamin D receptor
WCRF	World Cancer Research Fund
WHO	World Health Organization

1 Introduction

The human genome was characterized around the year 2000, marking the start of the era of genomic research. After the sequencing of the genome, large initiatives were launched to identify potentially disease-causing genomic variation. So far, results from those initiatives have been modest. An alternative focus for genomics research has been the study of gene activity, known as gene expression. With gene expression analysis using full-genome microarrays, the molecular mechanisms involved in the association between lifestyle, diet, health and disease can be approached with a genome-wide perspective. However, gene expression analysis has not been embraced in large-scale, population-based studies to the same extent as the study of gene variation.

To explore this research gap, an extensive collection of blood samples from a representative selection of the Norwegian female population was initiated in the Norwegian Women and Cancer (NOWAC) Post-genome Cohort. Unlike the samples collected by most other cohort studies, these blood samples were eligible for gene expression analysis. The present thesis takes advantage of this unique sample material, to explore potential associations between gene expression profiles and several exposures including lifestyle and dietary factors, in a cross-sectional study design. Due to the fact that Norwegians consume large amounts of fish compared to many other countries, two nutrients related to fish (polyunsaturated fatty acids (PUFAs) and vitamin D) are the focus of three of the papers included.

1.1 Lifestyle and nutrition in the etiology of human disease

1.1.1 Historical aspects and potential for disease prevention

Throughout human history, lifestyle and dietary patterns have been changing [1]. The transition from hunter-gatherer societies, via the peasant-agricultural, to the urban-industrial societies has led to a diet that is increasingly based on animal products, and that contains more energy, more fats from plants and animals, and more sugar. The consumption of fish today is highly variable according to geography [2] and economy [3]: rich countries consume

more fish than poorer countries. Also, different kinds of fish and seafood are consumed, and they provide varying amounts of nutrients [2]. The most recent steps of the dietary transition have coincided with a drop in human energy requirements, as the change of lifestyle occurring in industrialized societies during the last 60 years has led to a reduction in overall physical activity [1].

Dietary patterns and other lifestyle factors related to historical, geographical, and (socio-) economic changes, have led to increased life expectancy, but these changes also correlate with global disease patterns [1]. For example, during the last 50 years, European consumption of fat has increased by approx. 50%, and prevalence of obesity in adults has doubled [4]. Also, the incidence of lifestyle-related cancers and cardiovascular disease (CVD) increases in people that adopt a so-called Western lifestyle [1]. These examples point to the contribution of modifiable factors to disease etiology, and also highlight the potential for disease prevention. It has been estimated that approximately one quarter of cancer cases in high income countries can be prevented through adopting a healthy eating pattern, maintaining a healthy weight, and being physically active [1]. Based on an extensive review of current evidence, the World Cancer Research Fund (WCRF) and the American Institute for Cancer Research (AICR) formulated eight main recommendations for healthy living, with the aim of reducing cancer risk [1]. It has now been shown that following these recommendations may lower overall cancer risk by 18%, with numbers differing according to cancer site [5]. For example, breast cancer risk may be reduced by 16%, colorectal cancer by 27%. Furthermore, adherence to the WCRF/AICR recommendations was associated with significantly increased longevity, as well as reduced hazard of death caused by cancer, cardiovascular diseases (CVD), and respiratory diseases [6]. In addition to the factors related to diet, body fatness and physical activity, smoking is the major lifestyle-related risk factor for cancer. Tobacco use is estimated to cause 22% of cancer deaths worldwide, making it the greatest single avoidable risk factor [7]. Moreover, smoking is an even more important risk factor for CVD and respiratory diseases [8].

1.1.2 Tools for research on nutrition and lifestyle

Although the associations and biological mechanisms for the influence of tobacco smoke on disease has been extensively evaluated and reported, this is not the case with some of the other risk factors mentioned above, including diet, weight, and physical activity. Epidemiology may find associations between exposures and diseases, but there is a demand for biological plausibility for the associations to be regarded as potentially causal. Many dietary constituents are being evaluated as potential risk (or protective) factors for disease, but as a typical diet may provide more than 25 000 bioactive constituents, the characterization of associations and potential mechanisms is a complex task [1]. Important tools in the research for causal mechanisms related to nutrition and lifestyle factors include the use of biomarkers, and increasingly, the use of high-throughput molecular biology technologies. Biomarkers may be defined as any substance, structure or process that can be measured in the body, or its products, that influences or predicts the incidence of disease [9]. Generally, they are classified into biomarkers of exposures and biomarkers of effects (Figure 1). High-throughput technologies have evolved as an effort to elucidate the interaction of exposures and cellular and physiological homeostasis [1]. Gene expression profiles may be regarded as biomarkers of early effects (Figure 1) [10], and the technology will be further introduced in chapter 1.4.

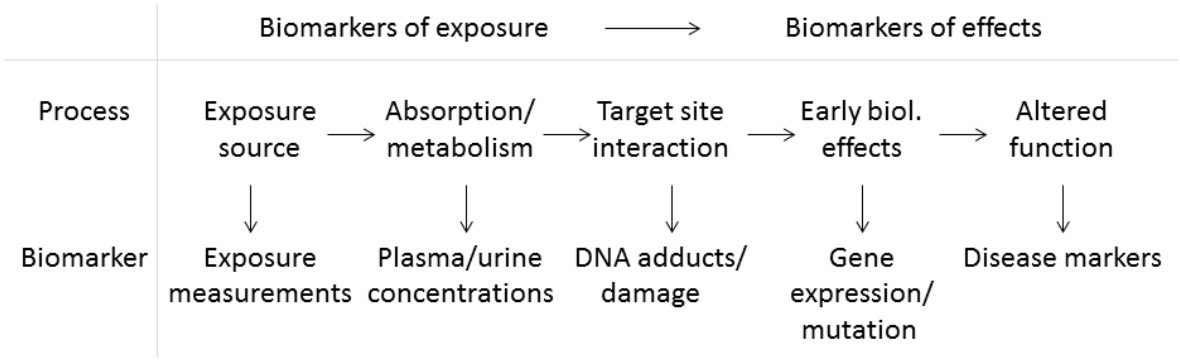

Figure 1: Overview and examples of biomarkers. Adapted from [9].

A main goal of medical research and epidemiology is disease prevention. Despite decades of research, more knowledge is needed at both the epidemiological and molecular level to describe and thereby be able to identify pathogenic processes in a pre-clinical setting. Only

then can interventions be made at the individual level, and preventive measures can be taken in the form of lifestyle and dietary advice to the public. The public health perspective is reflected by the fact that governments and institutions initiate the publication of large reports that describe the state of the available evidence, and attempt to draw conclusions based on reported findings. Several such reports are referred to in this thesis (Table 1).

Table 1: Reports on aspects of nutrition and lifestyle.

Title of report	Institution	Short name	Ref.
Food, nutrition, physical activity, and the prevention of cancer: a global perspective	World Cancer Research Fund (WCRF) and American Institute of Cancer Research (AICR)	WCRF/AICR report	[1]
Fats and fatty acids in human nutrition: report of an expert consultation	World Health Organization (WHO) and the Food and Agriculture Organization of the United Nations (FAO)	WHO/FAO report	[11]
A comprehensive assessment of fish and other seafood in the Norwegian diet	Norwegian Scientific Committee for Food Safety	-	[12]
Dietary Reference Intakes for Calcium and Vitamin D	Institute of Medicine (IOM), funded by National Institute of Health	IOM report	[13]

It should be mentioned that research on effects of nutrients may complicated by non-linear dose-response relationships. Nutrients may display a U-shaped risk curve, where both too low and too high levels are associated with adverse effects [14]. This complicates interpretation of reported findings. This property must also be taken into account when issuing dietary advice to the public: excessive intake, for example through dietary supplements, may be harmful. However, excessive intake will not be further discussed here, as the general tendency in the Norwegian population is deficiency of the nutrients in question.

1.1.3 Fish consumption

Fish is considered part of a healthy diet and provides a number of macro- and micronutrients that may be important for disease etiology. These nutrients include protein, PUFAs, retinol, vitamins D and E, selenium and iodine. Of these, PUFAs and vitamin D are the focus of the

present thesis. On average, Norwegians consume approx. 55-65 grams of fish per day, equivalent to approx. two fish meals per weak [12, 15]. This provides 40% of ingested vitamin D per day, and 9% of ingested PUFAs per day [15]. In a study of ten European countries, Norwegians consumed the second highest amounts of fish [2]. However, only 21% of women eat the recommended 200 grams of fatty fish per week (salmon, trout, mackerel, tuna, sardines, herring) [15]. Compared to other nationalities, Norwegians eat high amounts of very fatty fish, and high amounts of white fish products [2]. Consumption differs according to population strata, with potential adverse consequences for nutrient status in low-intake population strata [12]. Increased fish consumption is encouraged by Norwegian health authorities, but no specific amount is recommended [16].

The protective effect of fish consumption for CVD, inflammatory diseases, and possibly cancer is primarily linked to n-3 PUFAs which are present mainly in fatty fish. Also, vitamin D is present in higher amounts in fatty fish. Hence, it may not be reasonable to expect clear associations between total fish consumption and diseases. In the WCRF/AICR report, the available evidence for a potential association of cancer with consumption of fish was evaluated [1]. The report concluded that there is "limited-suggestive" evidence for a reduction in risk of colorectal cancer. An important limitation of the included studies was the broad definition of fish, with no stratification into lean and oily fish. Similarly, a recent systematic review, which included data from 106 case-control and cohort studies, concluded that the hypothesis that fish consumption reduces risk of prostate, breast and colorectal cancers is supported by little epidemiological data [17]. However, several challenges were pointed out by the authors, including the fact that many studies had not stratified between lean and fatty fish. Studies that did stratify, found inverse relationships between fatty fish and the risk of breast [18] and prostate cancer [19]. However, these association have not been found by all [20]. Fish is stored and processed in a number of different ways that may alter the properties of PUFAs, or introduce harmful substances such as high amounts of salt, or mutagenic compounds from frying. Storage and processing may modify the potential association of fish consumption and cancer, and consumption of pure filet was found to provide the most favorable PUFA

composition [21]. Other potential confounding factors and limited sample sizes must be taken into consideration when evaluating the literature.

1.2 Fatty acids

1.2.1 Structure and metabolism

Fatty acids (FAs) are macronutrients that are involved in multiple cellular and physiological mechanisms. FAs consist of a hydrocarbon chain with a methyl group (CH_3) at one end (denoted as the omega- or n-end) and a carboxyl group (COOH) at the other end (Figure 2). Chain length typically varies from 14 to 24 carbon atoms, and the presence and number of double bonds gives rise to the classification into saturated (SFA, no double bonds and therefore saturated with hydrogen atoms), monounsaturated (MUFA, one double bond), and PUFAs (multiple double bonds). FAs are named according to chain length, number of double bonds, and position of the first double bound counted from the n-end (Figure 2). The presence of double bonds alters the physical properties of the FA, by introducing kinks in the carbon chain, and rendering the FAs more prone to oxidation.

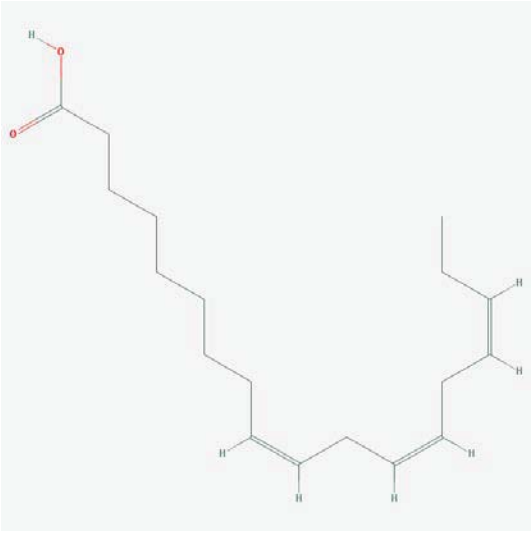

Figure 2: The structure of alpha-linolenic acid, 18:3n-3 (PubChem CID 5280934). Reprinted with permission [22].

FAs from the diet or from lipolysis in adipose tissue, enter the blood stream esterified in lipoproteins such as chylomicrons (dietary FAs) or very low density lipoproteins (VLDL,

endogenous FAs), or non-esterified bound to albumin. FAs are transported into cells throughout the body, where they may be stored as triglycerides, incorporated into the phospholipids of cellular membranes, modified by chain elongation and desaturation, oxidation and peroxidation, and lastly, they serve as precursors for immunoactive eicosanoids (including prostaglandins, thromboxanes, leukotrienes, resolvins and others). In addition, FAs are potent regulators of gene expression, as described below [23, 24].

The major FAs provided through the diet include those of the n-3, n-6 and n-9 families. Because mammals lack the enzyme that catalyses insertion of double bonds in the n-6 and n-3 position, two FAs are essential to humans and must be obtained through the diet: linoleic acid (LA, 18:2n-6) and alpha-linolenic acid (ALA, 18:3n-3). These serve as precursors for longer FAs with a higher degree of unsaturation, like the n-6 arachidonic acid (AA, 20:4n-6), and the n-3s eicosapentaenoic acid (EPA, 20:5n-3) and docosahexaenoic acid (DHA, 22:6n-3) (Figure 3). In humans, this metabolism is limited [25]. Importantly, the classes of n-6 and n-3 FAs are not inter-convertible [26], but they compete for the same set of metabolic enzymes during chain elongation and desaturation. Due to the high content of LA in the Western diet, the n-6 FAs dominate the enzymatic conversions, resulting in increased amounts of AA. On the other hand, ALA has a higher affinity for the Δ6 desaturase, but is generally present in lower amounts. ALA is the precursor for EPA and DHA, but these are generally provided through the diet. The dietary intake of specific PUFAs will lead to the distribution of those PUFAs into virtually every cell in the body [27]. Intracellularly, the PUFAs are further metabolized to yield the eicosanoids, which may have differing or even opposite biological effects (Figure 3). In general, eicosanoids derived from AA are regarded as pro-inflammatory, as further discussed in chapter 1.2.4.

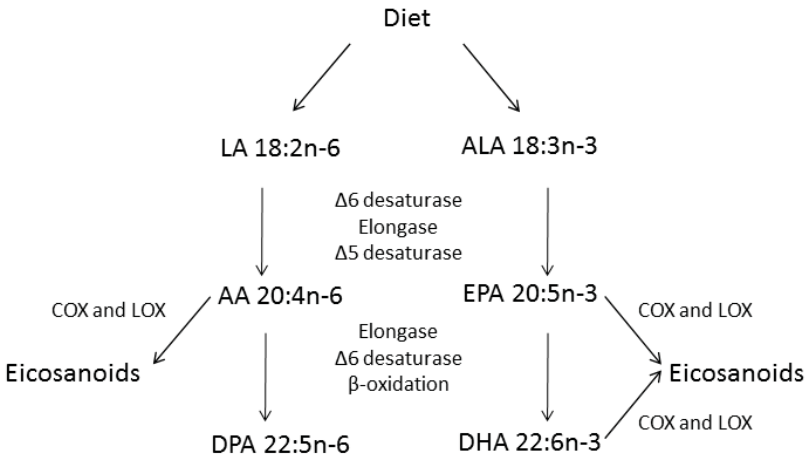

Figure 3: Overview of PUFA metabolism and eicosanoid production. Full names are stated in the list of abbreviations.

1.2.2 Sources and recommendations

Table 2 shows recommended and achieved intakes for total fat, various FA classes, and single FAs. In general, Norwegians consume higher amounts of fat than recommended, but lower amounts of PUFAs than recommended [15]. The consumption of marine PUFAs EPA and DHA was found to be just within the acceptable range, but the WHO/FAO report stated that the minimum value of 2 g/day for EPA/DHA may be increased in the future [11].

Table 2: Recommended and achieved intake of fat and PUFAs [a]

	Recommended per day		Achieved per day	
	% energy [b]	Amount (g)	% energy [b]	Amount (g)
Total fat	20-35 (25-35)[N]		30-34	67-74
PUFA [c]	6-11 (5-10)[N]		5,3-6.2	13
Total n-6	2.5-9			
Total n-3 [d]	0.5-2			
LA [e]	2.5-9			8.8
ALA	>0.5	1-2		1.2
EPA+DHA		0.25-2	0.35	0.84

a) Where values differ, values for women are given, and values for men are higher. Full names are stated in the list of abbreviations. Data from [11, 15, 28, 29].

b) % of total energy intake, minimum level or acceptable range.

c) LA+ALA+EPA+DHA.

d) ALA+ long chain n-3 FAs.

e) AA is not essential to adults whose diet provides >2.5 % energy of LA.

N) Recommendations in Norway.

N-6 PUFAs like LA are provided mainly from plant oils (corn, safflower, olive, peanut, and sunflower). The n-3 PUFA ALA is mainly provided in the form of plant oils from linseed, canola (modified rapeseed oil), and walnuts. Fatty fish and seafood are important sources of the long-chain n-3 PUFAs EPA and DHA. Smaller amounts of PUFAs are also present in meat, eggs and poultry, and processed foods that are common in the Western society are frequently rich in n-6 PUFAs. The most important sources of PUFAs in the Norwegian diet is butter, margarine, mayonnaise, dressings and bread [15]. Initiatives to increase n-3 consumption include dietary recommendations to increase consumption of (fatty) fish and oils from n-3 rich plants, such as linseed and rapeseed. In addition, Norwegian health authorities recommend the use of dietary supplements like n-3 oils/capsules or cod liver oil, and total sales for these supplements are 689 million NOK per year in Norway [30]. Among Norwegian women, 45 % take cod liver oil, but about half of the users take it only during the winter [31]. The standard dose of 5 ml cod liver oil (in the form of oil or capsules) provides 1.2 g n-3 PUFAs. However, a study found that only 58% of women taking cod liver oil took this amount [29]. Oils and capsules other than cod liver oil providing long chain PUFAs of varying kinds are also common in Norway. However, these have varying n-3 content, and do not contain vitamin D.

Recommendations have been put forward that the ratio of n-6/n-3 PUFAs should be "balanced", meaning 1:1, mainly based on the hypothesis that this was the dietary ratio during human evolution: our genes were evolutionary adapted to a balanced ratio, unlike the dietary ratio encountered in the Western diet today [32, 33]. The diet of industrialized societies is characterized by increased n-6 consumption and decreased n-3 consumption, resulting in a relative change in n-6/n-3 ratio over the last 150 years [32, 34, 35]. Some hypothesize that this discrepancy may influence the occurrence of lifestyle-related diseases. A 2002 WHO/FAO report reviewed the scientific evidence and indicated that a balanced ratio is essential to human health. However, a more recent WHO/FAO report stated no specific recommended level, given that intake of both n-6 and n-3 should be within the recommendations [11].

1.2.3 Epidemiological findings

Health effects from n-3 PUFAs were first described in 1961 after the observation of beneficial effects on cardiovascular disease [36]. Later, the low incidence of coronary heart disease in Greenland Inuits was linked to their high consumption of seafood [37]. Bang and Dyerberg summarized findings of reduced frequency of heart attacks in Norway during World War II and the tendency for long bleeding times in Greenland Inuits, and linked both observations to increased or high consumption of fish and decreased production of pro-inflammatory prostaglandins [38]. Since those pioneering studies, the potential effects of PUFAs have been explored for a wide variety of diseases and in relation to disease prevention. A brief presentation of general findings regarding potential effects of PUFAs on cancer, CVD and a selection of other diseases will be given here. Importantly, epidemiological findings have spurred a debate regarding PUFA effects. In a meta-analysis of randomized controlled trials and cohorts, Hooper *et al.* found no clear evidence of effects on general mortality, CVD events, or cancer [39]. However, the authors did not suggest changing dietary advice to increase fish or n-3 consumption. Their main critique towards findings from cohort studies pointed to the difficulty of adequately adjusting for lifestyle factors associated with fish/n-3 consumption, which generally is higher in people with a healthy lifestyle. Additional uncertainties in the available evidence is introduced by methodological issues such as short

follow-up time, the possibility of errors in estimations of intake, not considering individual PUFAs or n-6 and n-3 PUFAs separately, lack of consensus regarding adjustment for total energy intake, or biological factors like varying bioavailability of FAs and the influence of genetic polymorphisms [40-42]. The conclusions from Hooper *et al.* [39] were challenged in a report by a panel of experts appointed by the International Society for the Study of Fatty Acids and Lipids [43]. The authors argued that the combined weight of the scientific evidence from multiple study design types clearly demonstrate the benefit of fish oil PUFAs, especially for CVD (further discussed below).

- Cancer

Cancer causes 13% of all deaths worldwide, and is the leading cause of death in the developed world [44]. Although it is a diverse group of diseases, all cancers are characterized by uncontrolled cell proliferation. Tumourigenesis is viewed as a multistep process, reflecting the accumulation of genetic changes that enable the transformation of normal cells into malignant ones [45]. Importantly, tumours are no longer viewed as isolated entities of clonal cell proliferation, but rather as complex tissues composed of multiple cell types, that are actively interacting with their surroundings, known as the tumour microenvironment [46]. Via the blood stream or the lymphatic system the cancer might spread to distant parts of the body, forming metastases. Epidemiological studies exploring the influence of dietary PUFA intake on cancer incidence and/or death are numerous, however, firm conclusions have not been reached [17, 47, 48]. More consistent results have come from studies using biomarkers of exposure such as concentrations of single PUFAs in blood or plasma lipid fractions, rather than intake assessment. A protective effect was found for example for breast cancer, although the number of studies is limited [49, 50]. The WCRF/AICR report made no statements for PUFAs, because the report focused on food, not single dietary constituents. Clinical trials using defined amounts of specific PUFAs for cancer prevention or treatment (as co-treatment, or to alleviate symptoms of cachexia) are more rigorously designed, and may yield informative results [51]. Although a few trials are registered in the clinicaltrial.gov database, few results have been published [51]. More -and more carefully designed studies, are needed to explore the potential associations of single PUFAs or FA ratios on cancer incidence [42], such as the

VITAL study [52]. The epidemiological findings contrast the wealth of mechanistic evidence provided from cell culture studies that link PUFAs to several processes of carcinogenesis. Mechanisms include modulation of cell cycle control, inflammation, oxidative stress and cell membrane structure [42], further described in chapter 1.2.4.

- CVD

CVD is a leading cause of death in the developed world. Atherosclerosis is the result of slow, long-time deposition of lipids (cholesterol and triglyceride) in blood vessels, accompanied by vessel wall thickening and inflammation, as well as accumulation of macrophages engulfing large amounts of lipids (foam cells). The atherosclerotic plaque causes reduced blood flow at its site of origin, or may rupture and cause thrombosis and infarction (cell death due to lack of oxygen) at distal sites including the heart ("heart attacks"). It has been established that the marine n-3 PUFAs EPA and DHA reduce CVD risk factors, incidents, and deaths [43, 53-55]. Although this conclusion was challenged by the meta-analysis mentioned above [39], more recent publications have supported it [11]. On the other hand, the plant-derived ALA and n-6 PUFAs have not been assigned the same effects [53, 55-57]. Mechanisms of PUFAs in relation to CVD involve the reduction of plasma cholesterol and triglyceride levels, platelet aggregation and inflammation. The relative importance of the different mechanisms has not been established. In Norway, n-3 PUFAs are used as a prescription drug to treat hypertriglyceridemia, and as co-treatment for secondary prevention after heart attacks [30].

- Other diseases

PUFAs have been implicated in the etiology of several other diseases, primarily those involving an immunological aspect. Autoimmune diseases like rheumatoid arthritis may be reduced due to the modulation of AA-mediated inflammatory responses. Although some studies, including randomized controlled trials, have shown positive effects on chronic inflammatory diseases, the data has not been judged as adequate to inform clinical practice [58]. High levels of PUFAs or high fish consumption have been linked to increased risk of type 2 diabetes mellitus [59]. However, a systematic review/meta-analysis found no such associations, but concluded that the plant-derived ALA may be associated with modestly

lower risk of type 2 diabetes mellitus [60]. The WHO/FAO report concluded that there is possible evidence for a reduced risk of diabetes with increased PUFA intake [11]. Collectively, the conclusions are conflicting. Lastly, studies on PUFA exposure and risk of atopy/asthma in children and adolescents have shown promising results [61].

1.2.4 Molecular mechanisms

The specificity inferred by the chemical structure of PUFAs is a key aspect to their differential influence on physiological, cellular, and molecular mechanisms. The (relative) abundance of different FAs has been linked to disease prevention and pathogenesis as outlined above, but despite more than 30 years of research, the mechanisms of action are not fully understood. Hypotheses linking PUFAs to disease prevention and treatment is to a large extent related to their function as immunoactive substances [62]. Inflammation is essential to maintain health and homeostasis, but pathological inflammation may occur if regulatory control is lost, and it may contribute to the pathogenesis of chronic diseases and cancer. It has been firmly established that major inflammatory signalling pathways are modulated by PUFAs, but they do so in a variety of different ways [62, 63].

Dietary PUFAs are incorporated into the phospholipids of cellular membranes, where they serve as precursors for lipid mediators (the eicosanoids, Figures 3 and 4). Eicosanoids work in an auto- and paracrine manner to induce signal transduction via specific cell surface receptors. In general, lipid mediators derived from n-3 PUFAs are regarded as less pro-inflammatory or less biologically active compared to those derived from n-6 PUFAs, although this is a simplification [64]. The identification of the prostaglandins was the basis for the 1982 Nobel Prize for Physiology or Medicine, awarded to Bergström, Samuelsson and Vane [65]. In addition to eicosanoid production, PUFAs in the membrane phospholipids influence the physical properties of the membrane and lipid rafts, for example by increasing fluidity (Figure 4). This may affect signal transduction via membrane receptors, and the production of second messengers like diacylglycerol [64]. PUFAs interact with membrane-bound or intracellular receptors, some of which function directly as transcription factors. Transcription factors with PUFA and/or eicosanoid ligands include the peroxisome proliferator-activating

receptors (PPARs), retinoid x-receptor (RXR), Toll-like receptors (TLR), and others [58, 66]. Ligands for PPARs include dietary fats, prostaglandins, and oxidized lipids [67]. Upon ligand binding, PPARs heterodimerize with RXRs and bind to deoxyribonucleic acid (DNA) to control expression of target genes. On the other hand, TLRs reside in cellular membranes, and ligand binding causes TLRs to dimerize and recruit adaptor proteins to transduce signals intracellularly [68]. The dimerization and recruitment steps may be influenced by the properties of the membrane lipid bilayer [68]. Collectively, PUFAs may influence cellular processes like cycle regulation, apoptosis, endoplasmic reticulum stress, autophagy, lipid metabolism, oxidative stress, and calcium homeostasis [24, 69, 70].

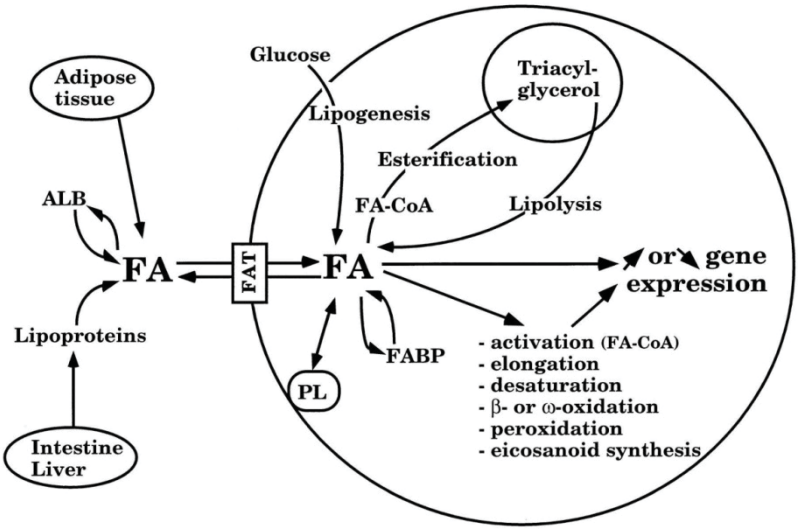

Figure 4: Major pathways of fatty acid production, transport, and metabolism. Full names are stated in the list of abbreviations. Reprinted with permission [71].

A multitude of mechanisms that modulate the activity of immune cells may be influenced by dietary PUFAs. Many of these, if not the majority of them, will –directly or indirectly- influence the level of gene expression in a number of inflammatory pathways. Via the mechanisms outlined above, PUFAs have been associated with expression levels of pro-inflammatory markers like TNFα, IL-1β and IL-6 in cell lines [62]. These findings have to a certain degree been verified in humans. For example, fish consumption was associated with decreased levels of biomarkers of endothelial dysfunction and low-grade inflammation in healthy adults [72]. The same tendency was shown for gene expression levels in an intervention study to reduce the n-6/n-3 ratio [73]. Two recent reviews examined the

association of inflammatory markers and n-3 PUFAs in individuals who were healthy or at risk of CVD: one concluded that PUFAs reduce inflammatory markers [74], whereas the other made no such conclusion [75]. In summary, molecular mechanisms affected by PUFAs may influence several aspects of immune cell function, with possible implications for disease prevention and progression.

1.3 Vitamin D

1.3.1 Structure and metabolism

The term "vitamin D" is used to describe a group of fat-soluble secosteroids [13, 76]. There are two main sources for vitamin D: the diet, and cutaneous synthesis after ultraviolet B (UVB) exposure (Figure 5). Dietary sources, or cutaneous conversion of previtamin D, provide the prohormone vitamin D_3 (cholecalciferol). Vitamin D_3 enters the circulation bound to vitamin D binding protein (DBP). Vitamin D_3 may be stored in adipose tissue, or taken up by the liver where it is hydroxylated to 25(OH)D. This metabolite is released to the circulation again, before being hydroxylated to the biologically active $1,25(OH)_2D$ (calcitriol, Figure 6) in the kidney. Conversion to the active metabolite in the kidney is regulated by parathyroid hormone and fibroblast-like growth factor-23, in response to serum calcium and phosphate levels [13]. Vitamin D_3 can also be converted to the active metabolite by cells of the immune system [77]. $1,25(OH)_2D$ binds to the vitamin D receptor (VDR), and molecular functions are presented in chapter 1.3.4. A final conversion in the kidney or intestine yields the inactive calcitoic acid, which is excreted in the bile.

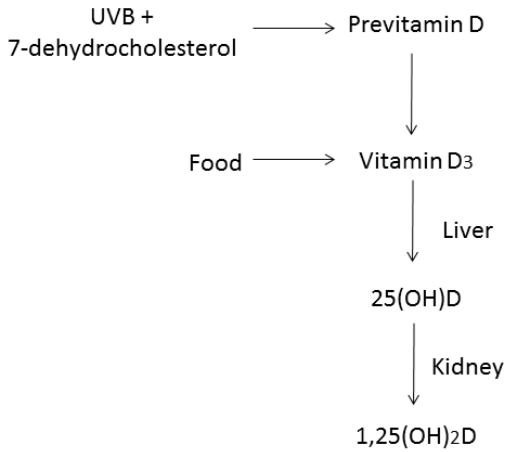

Figure 5: Overview of vitamin D metabolism. Full names are stated in the list of abbreviations.

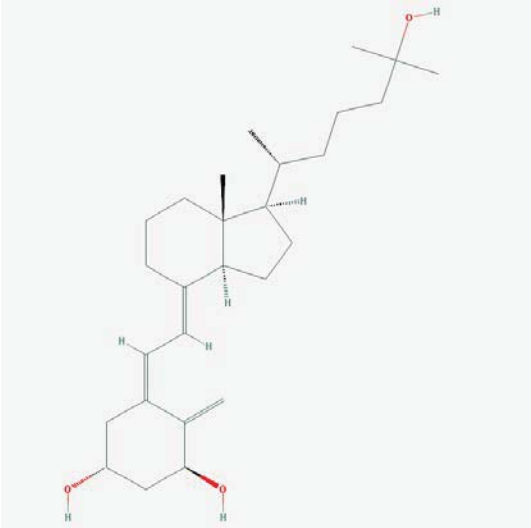

Figure 6: The structure of calcitriol ($1,25(OH)_2D$, PubChem CID 5280453). Reprinted with permission [22].

1.3.2 Sources and recommendations

The relative amount of circulating 25(OH)D that arises from cutaneous synthesis versus dietary intake has not been specified [13]. Natural dietary sources of vitamin D include fatty fish, fish oils and egg yolk. In addition, several foods are fortified with vitamin D, especially dairy products. The Norwegian recommendation for intake of vitamin D is 7.5 µg per day [15]. The recommended daily dose of cod liver oil (5 ml) provides 10 µg vitamin D, and other supplements, like multi-vitamin products, may provide significant amounts as well [12]. When excluding users of (any) dietary supplements (58% of women), mean daily dietary

intake in Norwegian women was estimated to 5 μg, and the mean intake was doubled when including users of dietary supplements [15].

When it comes to circulating 25(OH)D concentration, the contribution of cutaneous synthesis relative to dietary intake may vary substantially. For example, in the summer, the cutaneous synthesis is a significant contributor, but in the winter months it is much less so [78]. The degree of UVB exposure is influenced by a number of factors in addition to season, including weather conditions, latitude, skin pigmentation, use of clothing/sun screen, genetic factors, and age [13]. Importantly, in Norway, the sun is low or absent during wintertime, resulting in a prolonged period of no cutaneous vitamin D synthesis [79]. It follows from this that the dietary intake is all the more important during this period to maintain sufficient circulating 25(OH)D concentration [78]. UVB assessment requires tailored questionnaires or models in epidemiological studies [79], and these are not always used. The difficulty of UVB assessment complicates the establishment of dose-response relationships for the influence of both dietary intake and cutaneous synthesis on circulating 25(OH)D.

There is an ongoing debate regarding the target concentrations of circulating 25(OH)D, and the definitions of vitamin D status based on this concentration [80, 81]. The cut-offs directly influence the estimations of vitamin D status in a population, and subsequently may strongly influence dietary advice given to the public to prevent deficiency. At the individual level, cut-offs are used in the assessment of vitamin D status in patients, and may guide advice given by physicians to prevent or treat deficiency, often using supplements. Hence, definitions of vitamin D status based on circulating 25(OH)D will impact both public and personal health care. The IOM advocates a modest target level of circulating 25(OH)D. In their extensive report on dietary reference intakes for calcium and vitamin D, which serves as advice to the U.S. government, the IOM recommended a 25(OH)D concentration of 50 nmol/L, and defined vitamin D deficiency as concentrations below 20 nmol/L [13]. However, others use 50 nmol/L as a cut-off for vitamin D deficiency, and recommend a concentration of at least 75 nmol/L [82]. The daily requirement to reach target concentrations of 25(OH)D may vary

in different population strata. For example, the elderly may require higher intake. Also, because vitamin D is not readily mobilized from storage in adipose tissue, obese people may require higher doses compared to normal weight people [13].

1.3.3 Epidemiological findings

Low sun exposure and low vitamin D intake was causally associated with rickets more than a century ago [83]. In Northern Norway, a high prevalence of rickets was found in areas with low fish consumption as early as 1931 [84]. Since then, its association with numerous health effects has been explored. Findings remain controversial, and some advocate that the non-skeletal health effects of vitamin D may be more important than the skeletal effects [85]. The IOM report stated that available scientific evidence does not support a causal relationship between vitamin D and non-skeletal health outcomes [13]. Not many randomized controlled trials have been carried out, and the many observational studies are prone to uncertainties. These uncertainties are introduced by methodological variability (lack of standardized measurement methods, assessment of sun exposure, etc.), lack of defined dose-response relationships, and the fact that studies often co-administer vitamin D and calcium. Lastly, genetic variants in the VDR has been shown to modify the association of vitamin D and health outcomes [86]. A meta-analysis of observational studies in general populations found an inverse association between 25(OH)D and all-cause mortality [87], which was further supported by findings in a recent population-based cohort [88]. The current state of the evidence for cancer, CVD and a selection of other diseases will be briefly presented here.

- Cancer

The WCRF/AICR report on cancer prevention judged the evidence for foods containing vitamin D as "limited-suggestive" for a decreased risk of colorectal cancer [1]. This is in line with a more recent meta-analysis of intake and circulating 25(OH)D in relation to colorectal cancer [89]. The IOM report stated that evidence for colorectal cancer from both observational studies and randomized trials is available, although limited [13]. For other cancers there is a tendency for positive findings in smaller studies using larger doses of vitamin D supplements [90], and no associations found in larger studies using lower

doses [91]. A recent meta-analysis could find no evidence to draw conclusions regarding vitamin D supplements for cancer prevention [92].

- CVD

Observational studies have indicated a possible role for vitamin D in CVD, but these studies have often been of insufficient quality [93, 94]. The IOM report judged the evidence for vitamin D in relation to CVD as not convincing, and concluded that the role of vitamin D in CVD is at present unresolved [13]. Carefully designed randomized, controlled trials are needed to further explore the relationship between vitamin D and CVD [95].

- Other diseases

Paralleling its association with rickets in children, 25(OH)D has been used as a biomarker for osteomalacia and osteoporosis in adults. However, in a recent publication pointed out several uncertainties regarding the well-established link between bone disease and vitamin D [83]. First, low 25(OH)D has not been consistently found in people with bone mineralization defects, nor is osteomalacia found in as many osteoporosis patient as would be expected if the association was causal. Second, as pointed out previously, only a few large-scale randomized controlled trials of vitamin D supplementation without simultaneous calcium supplementation have been carried out. Meta-analyses of such studies show inconsistent results [13, 83, 96]. On the other hand, the evidence for moderate-dose vitamin D+calcium supplements to promote bone health has been judged as adequate [13, 83, 92]. The evidence for immune function, autoimmune disorders, diabetes, metabolic syndrome, physical performance including falls, and infections were evaluated at the same level as that for CVD: not convincing, and not adequate to conclude regarding the role of, or potential benefit from, vitamin D [13].

1.3.4 Molecular mechanisms

The lack of clear epidemiological findings contrast the amount of mechanistic data that describes the effects of vitamin D at the cellular level: based on gene expression profiling, it

has been estimated that $1,25(OH)_2D$ may influence expression levels of up to 5% of the genome [76]. Calcium and phosphate metabolism are the classical, well characterized mechanisms influenced by vitamin D. In addition, several non-classical mechanisms have emerged. These include cell growth and differentiation, apoptosis, cell adhesion, oxidative stress, DNA repair, and possibly autophagy [76, 97].

The biologically active $1,25(OH)_2D$ is mainly synthesized in the kidney, but may also be synthesized in many other tissues and cells, including immune cells. In target cells, activated $1,25(OH)_2D$ binds to the vitamin D receptor (VDR). The VDR is a transcription factor which upon ligand binding recruits its dimerization partner RXR and other co-activators, and modulates the expression of genes that contain the VDR responsive promoter element (VDRE) [76]. $1,25(OH)_2D$ may also repress gene transcription, via recruitment of co-repressors and inhibition of negative VDREs that are transcriptionally active in the absence of the ligand [76]. In addition to gene expression regulation, $1,25(OH)_2D$ has been shown to influence cellular function by non-genomic mechanisms. These mechanisms are initiated by $1,25(OH)_2D$ interaction with VDR located in caveolae lipid rafts of the plasma membrane. Downstream signaling occurs via signal transduction pathways including G-protein coupled receptors, phospholipase C (PLC), mitogen-activated protein (MAP-) kinase, phosphatidylinositol-3-kinase (PI3K) [76, 98]. These pathways may ultimately modulate gene expression levels. The specificity of genomic versus non-genomic actions is dictated by the stereochemical conformation of $1,25(OH)_2D$ [98].

The effect of vitamin D on cells of the immune system has been recognized for more than 20 years, but the mechanisms are still under investigation. Vitamin D may modulate the expression of several immune-regulatory compounds, such as chemokines, interleukins and adhesion molecules [99]. Cells of the adaptive immune system are generally inhibited by $1,25(OH)_2D$. For example, T cell proliferation, activity, and cytotoxicity may be decreased, but suppressive T regulatory cell numbers may be increased, which is important for self-tolerance [100]. Similarly, B cell proliferation and immunoglobulin production is decreased (which may be an indirect effect, as VDR expression in B cells has not been clearly

established) [77]. Cells of the innate immune system may also be inhibited by $1,25(OH)_2D$, and the induction of self-tolerance by dendritic cells may be favored [77, 101]. On the other hand, proliferation and differentiation of monocytes and macrophages may be increased, leading to a hypothesized role for vitamin D in defense against bacterial infection [77, 101].

The immune dampening effects of vitamin D have led to the investigation of potential roles in inflammatory, hypersensitive, and autoimmune diseases. Also, because inflammation is a prominent feature of other major diseases, there are possible implications of vitamin D for both cancer and CVD. Vitamin D influences the immune system, and thereby may modify inflammation as a cancer risk factor. Furthermore, direct effects on cancer cells have been clearly demonstrated [102]. There are tissue specific effects, but overall vitamin D seems to promote differentiation and reduce cancer cell growth rates by inducing cell cycle arrest or cell death [102]. In relation to CVD, one group identified that vitamin D inhibited foam cell formation and macrophage cholesterol uptake in diabetics, via reduced PPARγ signaling and reduced endoplasmic reticulum stress [103, 104]. In summary, despite the clear effects in cell line and animal studies, the physiological roles of vitamin D remains unclear [77, 98, 100].

1.4 Gene expression analysis

1.4.1 Background

Gene expression analysis is based on the central dogma of molecular biology, which was first stated by Francis Crick in 1958, and detailed by the same scientist in a Nature paper in 1970 [105]. The dogma describes the directionality of the flow of biological information from the sequence of purine and pyrimidine bases in the DNA, via transcription into mRNA, and translation into the amino acid sequence of proteins (Figure 7). The dogma entails the concept of the genetic code, which describes how triplets of nucleotides specify one amino acid. Also essential is the complementarity of the double-helical structure of DNA, specified by complementary base pairing. Collectively, these concepts form the basis for how we understand all living things: genetic information is encoded in the DNA in units (genes) and selectively used (expressed/transcribed to mRNA) as templates for production of proteins,

which in turn are the working units of the cell. The human genome contains approx. 20 500 genes [106].

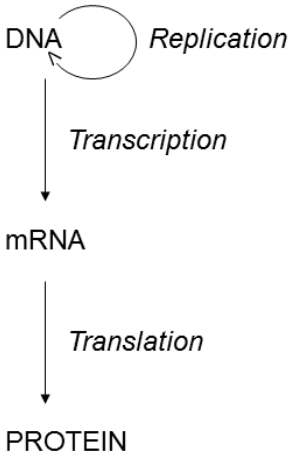

Figure 7: Central dogma of molecular biology. Full names are stated in the list of abbreviations.

Microarray technology is currently being used to investigate all levels of the biological information flow, from epigenetics, to DNA, RNA (protein-coding and non-coding/regulatory), as well as proteins. In principle, DNA microarrays are solid substrates (for example glass slides) with attached DNA oligonucleotide probes that correspond to genes (Figure 8). mRNA is extracted from biological samples, then purified, amplified, and labeled, before allowed to hybridize by complementary base pairing with the oligonucleotides on the microarray. After washing away unbound material, mRNA abundance is quantitated by image analysis.

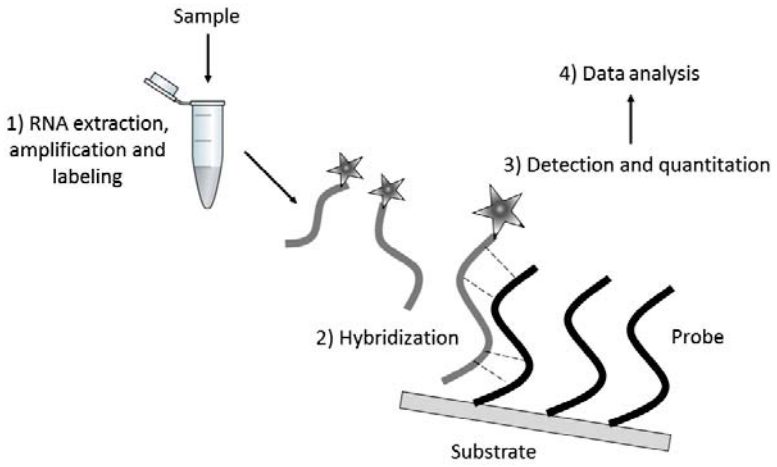

Figure 8: Basic principle of hybridization-based microarrays.

The complete collection of mRNA molecules present in a certain cell or tissue at a given time is referred to as the transcriptome, and assessment of the transcriptome is called transcriptomics. High-throughput ('omics) technologies, including microarrays, are increasingly being used to elucidate the complex interactions between environmental exposures and human health and disease. Gene expression microarrays are used for three main purposes: 1) identification of differentially expressed genes, 2) class discovery (grouping of samples based on gene expression profiles), and 3) class prediction (assignment of new samples to pre-defined groups). In the present thesis, microarrays were used to identify differentially expressed genes according to lifestyle and technical factors (paper I), as well as PUFA ratios (paper III) and vitamin D (paper II). Paper II and III thus fall within the field of nutrigenomics, where the use of genomic technology is merged with nutritional research.

1.4.2 Challenges

There are several challenges related to transcriptomics research: pre-analytical issues, the technology itself, data analysis, and lastly, interpretation of results. Problems reported from early microarray studies included limited reproducibility and comparability between platforms, which generated skepticism against use of the technology [107]. One of the most extensive evaluations of microarray variability stemming from intra-site, inter-site and inter-

platform differences was conducted by the MicroArray Quality Control (MAQC) project, initiated by the United Stated Food and Drug Administration. The project involved both the scientific community and commercial microarray providers, and came about to shed light on the reported problems. Results of the MAQC project were initially presented in a series of articles in Nature Biotechnology in 2006 (www.nature.com/nbt/focus/maqc). The project reported that the overall technical performance of microarrays "… *supports their continued use for gene expression profiling in basic and applied research*…" (quote from Shi *et al.* [108]). Although some of the assessment methods and results of the MAQC project have been challenged by others [109], the main conclusion has not.

One of the major challenges to providing reproducible microarray data involves the lack of consensus regarding data analysis methods. Data analysis is carried out in multiple consecutive steps, such as preprocessing and normalization, quality control, handling of flagged spots, data filtering (the removal of transcripts or microarrays due to quality control issues), imputation of missing values, identification of differentially expressed genes, and pathway-level analysis. Each step involves making choices regarding methods, cut-offs etc., which will undoubtedly influence the reported results [110]: selecting a different method may result in a different list of genes. It has been estimated that there are 10 million possible combinations of methods for analyzing a given microarray data set [110].

At the basis of the interpretation of microarray data was the simplistic idea of a linear coherence throughout the biological information flow (Figure 7): a gene is being expressed, and the abundance of the mRNA molecule corresponds to the amount of the translated protein, and lastly to the level of protein activity. The idea may be further extrapolated to the pathway level, and ultimately to physiological processes. Of course, this scheme far from reflects the complexity of the biological information flow, and the regulation of physiological processes through that flow. This will be further elaborated upon in the Discussion.

1.5 Human beings as a model system

There is an increasing understanding of the limitations of using animal models and cell lines to explore human pathogenic processes. Animal models have been extensively used in basic research, to perform experiments and test hypotheses that are not feasible or ethically acceptable to test in humans. However, a recent publication found that the degree of match between studies of transcriptomic inflammatory responses in humans and mice is close to random [111]. This discrepancy has also been shown for other aspects of immunology [112], and is highly relevant for the research on inflammatory processes involved in several major pathologies. However, costs for performing in-depth molecular analyses on human samples have dropped, and eligible biobanks of human sample material are being build. As a benefit of this development, the use of human sample material is increasing in biomedical research. By monitoring exposures using questionnaires and biomarkers, populations may be divided into comparison groups and thereby be regarded as "natural experiments" for the investigation of diverse arrays of exposures. This approach was employed in the present thesis.

1.5.1 Blood as a target tissue

Studies of blood cells are particularly useful for combining epidemiological and molecular aspects of biomedical research, because blood sample collection is relatively non-invasive and feasible in epidemiological cohorts. As the body's transport system, the blood interacts with all tissues, and blood cells are exposed to nutrients, metabolites, excreted factors and waste products. Hence, the blood has been suggested as a "surrogate tissue" for the study of diseases in other tissues [113]. Furthermore, the transcriptionally active blood cells are key immunologic regulators, and are in them self highly relevant for monitoring health status and studying inflammatory processes that are relevant for a range of pathologies including cancer and CVD.

Challenges related to gene expression profiling of blood cells are numerous. First, the blood is a complex tissue of fluctuating numbers of cell population subtypes, and cells at varying developmental stages [114]. The white blood cells (leukocytes) constitute only 1% of the blood volume, but they are the most transcriptionally active cell subtypes. Leukocytes include the

lymphocytes (20-50%, the T cells and B cells), monocytes (1-5%), and granulocytes (65%, the neutrophils, eosinophils and basophils). Platelets are cell fragments which may contain RNA from the bone marrow megakaryocytes that they originated from. Lastly, red blood cells constitute 44% of the blood volume, but they are regarded as transcriptionally inactive. The immature red blood cells (reticulocytes) are also regarded as transcriptionally inactive, but they do contain large amounts of mRNA coding for globin. This may constitute the majority of the total mRNA isolated from blood, and may pose another challenge to blood transcriptomics by masking less abundant transcripts. Globin reduction methods are available to increase the analytical sensitivity, but the methods may introduce bias [115]. Lastly, the stabilization of the RNA profile in a blood sample directly after blood collection is necessary, as the blood cells will react to the changed surroundings by altering their gene expression and thereby occlude the relationship between expression profiles and the biology/exposure of interest.

2 Aims

The overall aim of this thesis was to assess the feasibility of performing gene expression profiling in an epidemiological setting. In more detail, we aimed at exploring the specific influences of a range of exposures on gene expression profiles in a cross-sectional study of free-living, middle-aged Norwegian women. Based on this overall aim, three questions were addressed:

1) What factors related to inter-individual differences, major exposures, and technical variation affect gene expression profiles (paper I)?

2) How do dietary FAs affect gene expression profiles (paper III)? More specifically, what gene expression differences are related to high versus low plasma ratios of LA/ALA, AA/EPA, and total n-6/n-3?

3) How does vitamin D status affect gene expression profiles (paper II)? More specifically, what gene expression differences are related to sufficient versus deficient vitamin D status, measured as plasma 25(OH)D?

4) To what degree do the measured levels of vitamin D and PUFAs correlate (paper IV)? This needs to be clarified in order to discuss the gene expression profiles related to the two nutrients.

3 Material and methods

3.1 The Norwegian Women and Cancer Study

The Norwegian Women and Cancer Study (NOWAC, *Kvinner og Kreft*) was started in 1991, to investigate risk factors for breast cancer in a prospective cohort design. The study has been described in detail elsewhere [116, 117], and further information can be found online at http://site.uit.no/nowac. Women were invited to participate based on random sampling from the Norwegian Central Person Register, and enrollment has been carried out in three main steps (1991, 1995-97, 2003-07). Subjects participating for the second time were invited in 1998-2002, and those participating for the third time were invited in 2003-05. Today, the study includes more than 172 000 women aged 30-70 years at enrollment, who have answered questionnaires regarding lifestyle, health, medication use, diet, and other factors. The questionnaires used have varied in length over the years, ranging from two to eight pages. Furthermore, NOWAC holds an extensive biobank of more than 60 000 blood samples, as well as both cancerous (400) and healthy (400) breast tissue biopsies. Additional important features of the study includes repeated collection of exposure information (up to three questionnaires answered by each woman), and the opportunity to link to the Cancer Registry of Norway and the to the Cause of Death Registry. The required ethical approvals have been obtained from The Regional Committee for Medical Research Ethics and the Norwegian Data Inspectorate. All participants have given written informed consent.

3.1.1 The NOWAC Post-genome Cohort

The introduction of high-throughput 'omics technologies in the past two decades, has highlighted the need for epidemiological study designs and biobanks that could take advantage of the technological advances. The NOWAC Post-genome Cohort and its biobank was designed not only to meet the epidemiological goals of minimal error and bias (by optimizing the selection of study population and the study population size), but also to meet the requirements of the 'omics technologies. These requirements include adequate procedures for sample collection, shipment, storage and handling.

The NOWAC Post-genome Cohort comprises 48 943 women born between 1943 and 1957, who were randomly selected in series of 500 from the NOWAC study. Sample collection took place from 2003-06, and the overall response rate was 72%. Invited subjects received an information folder, a consent form, and a two-page questionnaire to be answered at the time of blood sampling. The questionnaire (Appendix) requested information about anthropometric factors, lifestyle, menopausal status, use of medication, use of dietary supplements, and specifications of the blood sampling procedure (date/time, posture, time of previous meal). In addition, the women received equipment necessary for donating two different blood samples: one buffered in citrate, and one buffered in an RNA-conserving solution. The women were asked to visit their general physician's office where blood sampling was performed. The blood samples were then shipped by mail to the NOWAC study center, using an enclosed, pre-paid envelope. Upon arrival, the citrate buffered blood sample was centrifuged, and the buffy coat and two plasma samples were stored at -80°C. The blood RNA collection tube was placed at -80°C without any pre-treatment, according to the manufacturer's instructions.

This thesis is based on four descriptive, cross-sectional studies of a series of 500 women in the NOWAC Post-genome Cohort, who donated blood samples in 2005. Of the 500 invited subjects, 89 % (444) women returned blood samples. All papers were restricted to postmenopausal women only, and general inclusion criteria were described in paper I. Furthermore, specific inclusion criteria were adapted according to the research questions of each paper.

3.2 Gene expression profiling

The PAXgene Blood RNA system (PreAnalytiX GmbH, Hembrechtikon, Switzerland) was used for blood sample collection, which conserves the RNA content of all cells present in the blood sample. The system provides an integrated work flow from blood sample collection to extraction of total RNA (above approx. 150 nucleotides in length [118]). The NanoDrop ND-

1000 spectrophotometer (ThermoFisher Scientific, Wilmington, Delaware, USA) and the Experion automated electrophoresis system (BioRad, Hercules, California, USA) was used for assessment of RNA quantity and quality. No protocol for reduction of globin mRNA transcripts was carried out. The Applied Biosystems (ABI) Expression Array System (Foster City, Louisiana, USA) was used for whole-genome gene expression analysis. Main steps of the microarray analysis workflow were outlined in Figure 8. The ABI Human Genome Survey Microarray v2.0 contains 32 878 probes for the interrogation of 29 098 transcripts.

3.3 Biomarker measurements: fatty acids and vitamin D

Plasma 25(OH)D concentrations were measured at Haukeland university hospital in Bergen, using high-performance liquid chromatography (HPLC) [119]. The following cut-offs were used to define vitamin D status: deficient: <37.5 nmol/l, sub-optimal: 37.5-50 nmol/l, sufficient: >50 nmol/l. For gene expression analysis, women with deficient vitamin D status were compared to those with sufficient status.

Concentrations of 34 FAs in citrate-buffered plasma samples were measured at the National Institute of Nutrition and Seafood Research in Bergen, using rapid gas chromatography [120]. The FA ratios (LA/ALA, AA/EPA, total n-6/n-3) were calculated from the measured concentrations, and the study sample was divided into deciles according to FA ratios. Women in the highest and the lowest FA ratio deciles were compared to identify gene expression differences.

3.4 Statistical methods

Statistical analyses in papers I, III and IV were performed using R with Bioconductor packages. In paper II, R and SPSS Statistic 19 were used. Details of the analyses are described in the respective papers. As the present thesis is not focused on biostatistics, only a brief description of the applied statistical methods will be given here. Gene expression analysis can be divided into analyses at the gene level, and analyses at the pathway or gene set level. To

analyze differences at the gene-level, papers I and II employed linear models (limma), and paper III employed t-tests. Adjustment for multiple testing is vital in gene expression analyses, and was carried out as detailed in the respective papers. Several factors, both technical and biological, could potentially confound the association of the measured exposures and the gene expression profiles as the output. Paper I and II included an adjustment for technical factors prior to analysis of differential gene expression (time between blood collection and freezing, RNA extraction date, array lot number). On the other hand, in paper III potential differences in these factors were evaluated in the comparison groups, and due to no significant differences, no such adjustment was carried out. Relevant biological confounders of which we had available information were evaluated in the respective papers. In paper I, smoking, hormone therapy, medication use, and an interaction variable (hormone therapy and medication use) were found to be significant, and included in the linear models. In paper II, BMI and marine FAs were included. In paper III, no adjustment for biological factors was done.

At the pathway level, papers I and II employed the Global Test [121] as a main analysis strategy for identifying potential differences between comparison groups, whereas paper III made use of the Gene Set Enrichment Analysis (GSEA) [122] and also an overlap analysis provided by Molecular Signatures Database (MSigDB) of the Broad Institute [122]. The Global Test is based on regression models which allows for inclusion of covariates, and the method is classified as self-contained [123]. GSEA is based on hypothesis testing using Fisher's exact test, and is classified as competitive [123]. Both tests allow a more biology driven approach to microarray data analysis, as they make use of previously accumulated biological knowledge [123]. The Global Test in papers I and II was used in a targeted way to investigate gene sets selected *a priori*. On the other hand, the GSEA employed in paper III used the major signaling databases Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO) and Panther as gene set input. Contrasting the Global Test and GSEA, which use the entire gene expression matrix as a starting point, the MSigDB overlap analysis uses the list of differentially expressed genes generated from the gene-level analysis (in case of

paper III, the t-tests). The method calculates the degree of overlap between the user-provided gene list and the reported gene sets in the MSigDB database.

In paper IV, the potential covariation between 25(OH)D and PUFAs was examined using Spearman's rank correlation and linear regression.

4 Results – Summary of papers

Papers I-III were based on analysis of gene expression patterns in the blood, using a cross-section of middle-aged women from the NOWAC Post-genome Cohort. The first paper describes general gene expression variability, whereas paper II and III take a nutrigenomics approach to describe gene expression profiles related to specific dietary compounds. The covariability of these compounds were assessed in paper IV.

4.1 Paper I

Deciphering normal blood gene expression variation - The NOWAC Post-genome study

In the analysis normal blood gene expression variation, three technical variables (time from blood sampling to freezing, RNA extraction date, microarray lot number) were found to explain 47% of the overall variation. Three biological variables (age, BMI and fasting status) and three exposure variables (smoking, medication use, hormone therapy use) collectively explained 8 % of the variation. More than 3000 probes were associated with smoking, as well as 17 out of 42 curated gene sets. Smoking-related processes included wounding, inflammation and metabolism of xenobiotics. BMI was also associated with differential expression of approx. 3000 genes, and processes related to adaptive immune responses were identified. Age was related to differential expression of 40 genes, but no biological processes were enriched. These finding illustrated that blood gene expression profiling is feasible in a free-living population, and that potential confounders including technical, inter-individual, and exposure variables must be evaluated in future studies.

4.2 Paper II

Plasma 25 hydroxyvitamin D level and blood gene expression profiles - A cross-sectional study of the Norwegian Women and Cancer Post-genome Cohort

The analysis of sufficient (25(OH)D>50 nmol/l) versus deficient (25(OH)D<37.5 nmol/l) vitamin D status revealed no significantly differentially expressed genes at our chosen thresholds for p-value and false discovery rate. However, using a targeted, pathway-level analysis, we tested 78 gene sets and identified that 26 of them were differentially expressed. The gene sets included pathways describing immunological processes, immune cell functions and major signaling pathways. Moreover, several gene sets based on *in vitro* studies of vitamin D function were differentially expressed, although the direction of gene expression change was not always consistent when comparing the *in vitro* results to our population-level data. Hence, care must be taken when extrapolating *in vitro* findings to *in vivo* function. The findings illustrated that vitamin D status is associated with modest gene expression differences at the population level, and that affected processes may influence disease pathogenesis.

4.3 Paper III

Plasma fatty acid ratios affect blood gene expression profiles - A cross-sectional study of the Norwegian Women and Cancer Post-genome Cohort

To investigate the potential transcriptomic effects of dietary FAs measured in plasma, we evaluated three FA ratios (LA/ALA, AA/EPA and total n-6/n-3). This approach could reveal if the amounts of n-6 FAs relative to n-3 FAs affected genes or pathways that may be linked to future onset of diseases. By comparing the highest versus the lowest FA ratio deciles, we identified that the LA/ALA ratio had the largest impact on gene expression profiles, followed by the total n-6/n-3 ratio, and the AA/EPA ratio. All ratios were associated with differential expression of genes related to immunological function, which was especially evident in the LA/ALA ratio. The overall tendency was for increased inflammatory activity in the highest FA ratio deciles. Genes and pathways related to lipid metabolism were associated with all FA

ratios, with many links to PPARγ-related pathways. Also, several autophagy marker genes were expressed at higher levels in the low ratio deciles. All FA ratios were related to specific miRNAs, and to groups of genes containing common promoter motifs that did not match any known transcription factors. Based on these results, we concluded that the different dietary FA ratios measured in plasma were mirrored in blood gene expression profiles, and that affected genes and pathways may influence disease onset.

4.4 Paper IV

Vitamin D status and PUFA ratios in a national representative cross-section of healthy middle-aged, Norwegian women

Due to their partially common dietary sources, and the increasing knowledge of congruent health outcomes and biological mechanisms, this study aimed at exploring the level of covariation between plasma levels of 25(OH)D and FA ratios in a national representative cross-section of healthy, Norwegian women. Measured plasma concentrations of 25(OH)D, marine FAs (EPA+DHA) and FA ratios were assessed using Spearman's rank correlation and linear regression. The results showed that marine FAs were significantly positively correlated to 25(OH)D concentrations, whereas there was a weak negative correlation of 25(OH)D with LA/ALA, AA/EPA and total n-6/n-3. The results showed that there is no linear relationship between vitamin D levels and FA ratios.

5 Discussion - Methodological considerations

5.1 Study design

This thesis is comprised of four descriptive studies using a cross-sectional study design. The cross-sectional design is used to describe the distribution of variables and their potential associations, but it may not be used to explore causality. On the contrary, a main function of descriptive studies is to generate new hypotheses, which may be further explored using other study designs.

Results from cross-sectional studies must be interpreted according to the inherent limitations of the study design. A main limitation is related to the fact that the exposure and the outcome is measured at the same point in time, so the temporal sequence of their occurrence cannot be elucidated. Furthermore, and like all epidemiological study designs, the cross-sectional design is prone to errors. Errors and biases in epidemiological research is often grouped into three main categories: selection bias, information bias and confounding [124, 125]. The nature and extent of these potential sources of error may vary according to study design, and issues particularly related to the cross-sectional design will be discussed here. The handling of these potential sources of error will influence the external and internal validity of the study.

5.2 External validity - selection bias

Selection bias may occur related to the choice of study population relative to the source population, and related to the participants of a study (response bias). Participants in NOWAC were randomly drawn from the National Population Register. Response rates have been relatively high [117], and the study population was found to be representative of the Norwegian female population in the relevant age groups [126]. Several aspects of the NOWAC study have been evaluated for external validity [126]. From the NOWAC registries, approx. 68 000 women were invited to comprise the NOWAC Post-genome Cohort (over-all response rate of the Post-genome Cohort was 72%, approx. 48 000 women). From the Post-

genome Cohort, a randomly chosen group of 500 women comprise the basis for the present thesis. A certain level of selection bias related to this group must be expected, mainly related to two issues. First, the women are third-time participants in NOWAC, and it is a well-known phenomenon that respondents have higher education and are at better health than non-respondents. However, a comparison of first- and third-time respondents in NOWAC who all donated blood samples, described in [127], showed no major differences between the two groups besides education level and use of hormone therapy. Second, selection bias may be introduced by the blood sampling procedure, which requires that women visit their general physician or other trained personnel for blood sampling. This active participation may lead to self-selection bias. Still, several aspects support the assumption that the study population is representative of the source population, including the results described above, the random selection of participants, and the high response rate in the study population of 500 women (89%). It should be mentioned that all papers included post-menopausal women only. In addition, exclusions were made for each paper, related to potential confounding factors (discussed below and in the respective papers). These exclusions limit the generalizability of the results.

A main goal of the present thesis was to explore the feasibility of performing molecular level analyses at an epidemiological scale. As a gene expression study started around 2005, there were few comparable studies to estimate effect sizes and thus, power calculations were not feasible. Accordingly, the sample size was chosen *ad hoc*. As a trade-off between generalizability and cost, a study population of 500 has proven to be fruitful. However, the size may seem limited in an epidemiological perspective, and may introduce random error affecting the external validity of the study and thereby the generalizability of the results. As a pilot study, external validity and generalizability were not the primary goals of this thesis.

5.3 Internal validity - information bias

The present thesis relies heavily on the use of biomarkers (Figure 1). Gene expression profiles measured in blood were used as biomarkers of early biologic effect in papers I-III, and

exposure was assessed in the form of biomarkers of nutritional status in paper II-IV. Measurement errors of both biomarkers and questionnaires may lead to information bias, which may cause misclassification of subjects into exposure or outcome categories. Due to the cross-sectional study design, the potential misclassification will mainly be non-differential, i.e. affect comparison groups equally, and it may introduce noise that may obscure modest associations. Multiple sources of error exist in the measurements of biomarkers, which may cause deviations of the measure from the true level of exposure. These sources will be discussed in the following sections.

Three main categories of sources of measurement error will be discussed for each biomarker: 1) pre-analytical issues (choice of laboratory methods and sample types, handling of samples), 2) analytical error of the measurement methods, and 3) errors due to between and within subject variation. This last category also comprises the concept of confounding: several factors that may vary between subjects could influence both the concentrations of the nutritional biomarkers (exposure) and the gene expression profiles (outcome). The influence of a third, confounding variable may cause erroneous conclusions about the associations between the exposure and the outcome. Confounding may be partly avoided through restriction (exclusion/inclusion criteria), matching (pairwise matching of subjects in comparison groups), stratification (analyzing associations in different strata of the confounding variable), or adjusted for in the statistical models [124]. In general, residual confounding may be present due to both unmeasured and unknown factors. Potential confounding factors were discussed in each paper and will be further elaborated upon here.

5.4 Gene expression analyses

Gene expression profiles were analyzed in whole blood samples, and were used as the outcome variable in papers I-III. By using expression profiles as a biomarker of early biological effects, we identified gene expression differences associated with a number of different exposures. However, several limitations related to gene expression technology must be considered. A detailed analysis and discussion of the factors that influence the variability of

the gene expression levels in the dataset was presented in paper I. The study identified three technical variables which accounted for 46,5 % of the overall variation in gene expression. These were time between blood sample collection and freezing of the sample (transportation time), RNA extraction date, and array lot number. A general discussion of the technology follows here.

5.4.1.1 Pre-analytical issues

Due to low invasiveness and practical advantages, blood sampling is widely used for measurements of biomarkers in epidemiological studies. For many applications including microarrays, the analysis of blood constituents is complicated by the biological variance of the blood as a tissue, but also by technical factors related to the procedures of blood sample collection, handling and storage, as well as specific down-stream processing like RNA isolation.

In our dataset, time between blood sampling and storage was among the factors that influenced gene expression variability the most, in spite of exclusion of samples that had been transported for more than three days (paper I). The use of an RNA-stabilizing agent during blood sample collection is thought to reduce the transcriptional impact of sample collection procedures and transportation, when RNA isolation is not possible directly after blood sampling. Expression profiles can also be stabilized by snap-freezing in liquid nitrogen, but this is not always feasible in the context of epidemiological sample collection. A few commercial blood RNA collection systems are available, but they may introduce differences in gene expression profiles [128, 129]. The PAXgene blood RNA system used in the NOWAC Post-genome Cohort has performed well in comparison to other blood sampling procedures, with regards to effects of pre-storage handling [130], storage over time [131], and RNA isolation output [129]. However, it may cause reduced sensitivity [132], and the protocol has not performed equally well in all studies [128]. The detectable gene expression variance introduced by different blood RNA collection methods must be kept in mind when interpreting studies that may not have used the same methods for blood RNA collection [129].

It has been suggested that reduction of globin transcripts in blood RNA samples may increase the sensitivity of gene expression results [133]. However, a pilot study comparing globin reduction methods showed not major advantage in our experimental setting, compared to no globin reduction [115]. In line with those findings, the main study was performed without globin reduction.

Variance associated with laboratory procedures is well known to influence microarray results [110], and RNA extraction date were among the variables influencing our gene expression data (paper I). This variable may reflect day-to-day variation, potential batch effects introduced by lot numbers of reagents, and also differences related to lab personnel. The influence of technical factors provides important clues for improvement of gene expression data quality. A high(-er) degree of automation of laboratory procedures would serve to reduce such variance, but batch effects of reagents and kits may still not be avoidable. Similar technical factors were found to influence the variance of a blood based gene expression test developed to predict CVD [134]. This illustrates the importance of raising awareness regarding the influence of laboratory factors, which has not yet been fully coped with. It also lays responsibility on assay developers (commercial providers) to improve production procedures, in order to minimize variance and increase reproducibility.

The difficulty of standardizing pre-analytical procedures poses a major obstacle for the implementation of gene expression profiling in the clinic, and initiatives have been launched to provide evidence-based guidelines for pre-analytical treatment of blood samples for gene expression analysis [135]. As long as they are recorded, laboratory factors can be monitored, potential differences between comparison groups can be discovered, and the factors may be adjusted for in the statistical analysis. Importantly, variance introduced by batches of reagents and microarrays must be reduced by randomizing the analysis of samples in different comparison groups.

5.4.1.2 Analytical error

The microarray used in this thesis, ABI Human Genome Survey Microarray v2.0, was included in the MAQC project described in the Introduction. Its performance was evaluated based on hybridizations of 58 microarrays performed in three different test sites (laboratories) [108]. Repeatability within test sites of quantitative expression signal values (CV= 5-15%) and qualitative present calls (concordance rate 80-95%) was comparable to other single-color microarray platforms. As expected, the variance increased when including data from multiple test sites, illustrating that measured signals and the number of detected genes may not be directly comparable between sites. Still, when comparing two samples (aliquots of the same two samples in all sites) analyzed with the same platform in different sites, the resulting list of differentially expressed genes overlapped with on average ~89% between test sites. Furthermore, when comparing results from different microarray platforms, lists of differentially expressed genes overlapped with ~74% across single-color platforms. This lead the authors to conclude that the platforms do reflect biology in spite of technological differences [108].

Our study (paper I) identified the microarray lot number as one of three technical contributors to variance in the dataset. The MAQC project did not investigate microarray lot numbers as a potential source of variance, but this factor has been pointed out as an important technical factor which may influence results [136]. The lot numbers may be adjusted for in the statistical analyses, as long as they are recorded. However, this variance may reduce the generalizability from one microarray project to the next.

Validation of microarray data is often carried out using an independent method of measurement, such as quantitative real-time reverse transcription polymerase chain reaction (qRT-PCR). This was not done in the papers included in this thesis, for several reasons. The repeated measurement of mRNA abundance in a sample by first using microarrays, then qRT-PCR, was reported to be of limited value [136]. On the other hand, for a validation of the biological findings to add significant strength to the presented conclusions, independent samples must be used [136], and analyses of different levels of the molecular biology information flow (Figure 7) should be carried out. However, key features of this thesis include

the population perspective, the observational/cross-sectional design, and focus on pathway-level analyses. Based on the observational and exploratory nature of the studies included here, biological validation of the findings were regarded as beyond the scope of the present work. The decision not to use qRT-PCR for validation was further supported by the finding that ABI microarray results were comparable to TaqMan qRT-PCR assays, albeit less sensitive [137]. Furthermore, we report only associations of gene expression levels and exposures, and make no claims of induction or repression of gene expression levels as a response to the exposures. It is in the nature of the cross-sectional design that no conclusions can be made regarding the temporal and causal aspects of the reported associations. The conclusions presented in the papers are adapted to the strengths and limitations of the gene expression data, and reflect the explorative and observational nature of the projects.

5.4.1.3 Analysis of microarray data

As the present thesis is not focused on statistical analysis of microarrays, only a brief discussion of the main steps of data analysis (preprocessing, identification of differentially expressed genes, and pathway-level analysis) will be included here. Initial data preprocessing for papers I-III was carried out in accordance with the microarray manufacturer's instructions. Normalization procedures has been suggested as a major source of variation of microarray data sets [110], but a study of ABI microarray data suggested a high level of concordance between normalization procedures for this platform, including quantile normalization which was used in papers I-III [138].

- Gene-level analysis

A main point that has been discussed in the literature regarding identification of differentially expressed genes, concerns the choice between the often used and well-known t-tests, versus employing more tailored, microarray-specific statistical methods [110, 139]. The methods used in the thesis reflect the lack of consensus. Importantly, the performance (false-positive rate and power) of both limma (paper I and II) and t-test (paper III) was rated as "very good" for large samples [139]. To produce gene lists, differentially expressed genes are ranked and selected based on chosen parameters. The choices made during data analysis will influence the

sensitivity and reproducibility of the gene lists [140]. The analyses chosen for all papers included here employed well-documented methods for controlling the number of false positives in multiple testing (type 1 errors, the false discovery rate [141, 142]), in addition to a measure of statistical significance (the p value).

Reports have been published of inconsistencies between probe-to-gene annotations of different releases of probe set identifiers by several microarray providers, including ABI [143]. These inconsistencies may cause only least-common subsets of probes to be eligible for analysis, especially when including data from multiple sources [143]. This problem was evident in the vitamin D paper, where gene sets used for GSEA were reduced in size with up to 75% due to lack of consistency in gene identifiers between different sources. Furthermore, ABI discontinued the update of their annotation files, depriving users of the opportunity to easily update annotations. Poor annotation causes loss of information, and hinders biological interpretation of the data. This constitutes a considerable limitation of the data presented in the thesis.

- Pathway-level analysis

To aid functional interpretation of microarray data, papers I-III include analyses on the level of groups of genes (gene sets or molecular pathways). Gene set testing may be carried out using a range of different methods, with different approaches and underlying assumptions. Still, the methods share the feature of allowing a biologically driven data analysis. The approach restricts the analysis to *a priori* knowledge, but provides a useful means to link exploratory nature of the technology to the wealth of molecular and biomedical data present in various databases. Importantly, functional insight may be reached despite of low overall gene expression differences, as was demonstrated in paper II. Related to this, there is an ongoing discussion within the 'omics field regarding the distinction between statistical and biological significance. As demonstrated in paper II, gene set/pathway-level analysis enables the identification of biologically relevant findings, even in the absence of strong and significant signals at the gene level.

The Global Test [121] (paper 1 and II), and GSEA [122] (paper III) do not necessarily lead to the same conclusions [123, 144]. The tests make use of differing definitions of the hypothesis to be tested, and differing sampling methods for calculations of p-values [123]. GSEA has been found to have low power in some instances, and may not be as intuitive when it comes to interpretation of the p-values. Based on the critique in Goeman *et al.* [123], one might consider replacing GSEA with other methods for gene set testing. On the other hand, the Global Test (paper I and II) closely resembles gene-level testing, making the interpretation more straight-forward. The Global Test has been found to be powerful, making it suitable for analyses where small effects are expected, as in the analysis of potential associations with vitamin D [121]. The Global Test has the advantage of the opportunity to test all genes as one gene set. This may be used as a preliminary test to identify whether the samples within a comparison group have very different gene expression profiles [121]. This overall test may be used to guide further analysis strategies, as in paper II. Furthermore, the result of this test in paper II pointed to inter-individual variation of gene expression profiles as a main contributor to the lack of clear associations between expression profiles and vitamin D status. Another advantage of the Global Test is the identification of what single genes contribute to a significant gene set result [121]. This functionality aids the biological interpretations.

Importantly, results from gene set analyses can only be generated based on the gene sets used as input. Gene sets can be derived from repositories like GO or KEGG, or they may be extracted from the literature. Both approaches were used in this thesis. When including the largest and most comprehensible databases available, the gene set approach reduces researcher bias in interpretation of lists of differentially expressed genes. Conversely, the inclusion of gene sets from published papers as in the Global Test (paper I and II), and in the MSigDB overlap analysis in paper III, gives the opportunity to identify studies with similar results. In our case of a cross-sectional study, similarities with more closely controlled cell culture experiments helps to narrow down the number of molecular mechanisms that should be considered. Still, as the input gene sets are chosen *a priori*, one might not identify all mechanisms at play. This may be the case in paper II, where 78 gene sets reported from the literature were chosen for a targeted, or more focused, analysis. This of course, has the

advantage of reducing the problem of multiple testing, and increasing the chance of arriving at statistically significant results.

5.4.1.4 Within and between subject variation

Variability of gene expression profiles over time within subjects (intra-individual variation) could not be studied in the present thesis, as only one blood sample was collected from each study participant. Whitney *et al.* showed that multiple samples from the same donor did not consistently cluster together [145], indicating a notable intra-individual variation of expression profiles. On the other hand, Eady *et al.* showed consistent levels of expression for the majority of genes in isolated PBMCs for a period of five weeks [146]. Importantly, in the study by Eady *et al.*, the blood sampling followed a closely controlled scheme, unlike the Whitney study, and also unlike the NOWAC Post-genome Cohort. Another study showed modest variation of expression profiles for a period of 1 and 3 months, however, the blood sampling procedures were not sufficiently described to be evaluated [147]. A general conclusion from these studies is that within-subject variation over time in healthy individuals seems to have smaller impact on gene expression profiles than between-subject variation. This lends support to our study design with the use of only one blood sample per participant.

Between-subject variation among healthy individuals have been shown to stem from age, gender, differing blood cell sub-populations, and the time of day that the blood sampling was done. Many factors that vary between subjects may not be standardized in a cross-sectional population-based study. The influence of several such factors were explored in paper I, and identified factors were treated as potential confounders in the subsequent papers. Two main factors that impose serious limitations to the present study will be discussed here: differing blood cell sub-populations, and fasting status. An important limitation of the study is related to the lack of blood cell counts and information regarding health status at the time of blood sampling. With the PAXgene system, whole blood is collected and processed, and the resulting RNA is derived from all circulating cell types. Hence, the resulting gene expression profiles may not be comparable to those of specific cell subtypes [148]. Also, the profiles may be subject to confounding due to fluctuations of these subtypes related to the immunological

status of the subject [145]. Because the relative proportions of the different cell subtypes vary between persons and over time, performing cell subtype counting may yield relative ratios that can be used for normalization purposes. Cell type proportions could also have been adjusted for by using expression markers as defined in [145] and demonstrated in [149], or by using information on disease status. Data on current health status was not collected in the NOWAC Post-genome Cohort, but by treating current use of medication as a potential confounder, one may hope to gain an impression of, and be able to adjust for, health differences. The fluctuations in cell subtype numbers may also be related to potential confounding factors whose level or influence is difficult to estimate and control for. This may include medical history, subclinical or untreated infections, stress, and nutrition.

A further limitation of the study is related to the non-fasting status of the majority of included subjects. Fasting status was identified as an inter-individual factor contributing to the overall variability of the dataset (paper I). Fasting subjects were excluded from paper II, and time since previous meal was treated as a potential confounding factor in paper III (fasting subjects not excluded). The potential overall transcriptomic effect of time since the last meal has not been explored in the literature, but it is certainly related to levels of biomarkers of nutritionals status. Hence, it may be regarded as a potential confounding factor. This is further discussed in the section regarding FA measurements.

In summary, the lack of standardized procedures for collection, storage, and analysis of blood gene expression profiles makes interpretation and comparisons between studies challenging [113].

5.5 Biomarkers of nutritional status

The biomarkers used in paper II (plasma 25(OH)D) and III (plasma FAs) may be classified as biomarkers of nutritional status, concentration biomarkers, or as biomarkers of internal dose. Collectively, the biomarkers are thought to reflect the exposure in the form of current intake status, comprised of dietary intake, absorption and metabolism of the specific nutrients.

Compared to estimations of dietary intake using questionnaires, food frequency questionnaires or other methods, the use of biomarkers of nutritional status provides an objective measurement not biased by memory, awareness or willingness to report details regarding dietary intake. Also, not all biomarkers can be estimated from a standard questionnaire, as is the case with the influence of UVB exposure on 25(OH)D concentration (further discussed in chapter 5.5.2). Thus, assessment of the potential association of the nutrient and disease or disease etiology is improved when employing biomarker measurements. However, several limitations should be considered. First, it is important to note that the use of biomarkers of nutritional status restricts the opportunity to make conclusions regarding *dietary intake* of a nutrient and its associations with the outcome variable. In papers II and III, the measured nutrients are not used as biomarkers to estimate dietary intake, but rather as independent biomarkers of exposure. As already mentioned, subjects included in the NOWAC Post-genome Cohort were not required to fast before blood sampling. Hence, the measured levels of biomarkers may be influenced by recent dietary intake. Potential (pre-) analytical errors and within/between subject variation (including confounding factors) will be discussed separately for FA measurements and 25(OH)D measurements.

5.5.1 Fatty acid measurements

5.5.1.1 Pre-analytical issues

For paper III, 34 unique FAs were measured in citrate-buffered blood plasma samples. FAs can be measured in various fractions of the blood (plasma, serum, red blood cell membranes) and in specific subclasses of lipids (triglycerides, cholesterol esters, phospholipids). It has been reported that FA composition of plasma was stable after four years when stored at -80°C [151]. In the NOWAC Post-genome Cohort, citrate-buffered plasma samples were stored at -80°C and analyzed within approx. one year [152]. The different lipid subclasses may reflect dietary intake and long-term exposure to varying degrees. In our study, total plasma FAs (not separated into lipid subclasses) were used, and studies have revealed that this biomarker may reflect dietary intake over the last few days [150, 153], or even more long-term

in a way similar to adipose tissue [154]. It has been suggested that changes in one blood lipid fraction (after dietary changes or supplementation) tend to be reproducible in the other blood lipid fractions [150]. Furthermore, measurements of the exogenously derived FAs are generally more closely associated with dietary intake, compared to FAs that can be endogenously synthesized [150, 153]. Still, differences may exist among the individual FAs, with regards to how well they reflect dietary intake [150, 155], but PUFA content of diets in cross-sectional studies tend to correlate well with FA composition of all blood lipid fractions, including total plasma FAs [150]. However, changes in the plasma total FAs may be diluted due to the lipid fractions that are more stable during dietary changes, like the non-esterified free FAs released from adipose tissue [150]. This study (paper III) was not undertaken to evaluate dietary FA intake, or the effect of estimated intake on gene expression profiles, but rather the effect of the FAs present in the plasma. The (non-) fasting status of the subjects, and the time since previous meal before blood sampling are discussed below.

5.5.1.2 Analytical error

The rapid gas chromatography method used in paper III to measure plasma FA concentrations was validated relative to conventional gas chromatography, and no statistically significant differences were found [120]. Analytical reproducibility could not be evaluated in the material used in the present study, as no repeated measurements were available. However, the coefficient of variation (CV=standard deviation divided by the mean) can be used to evaluate the extent of variability in relation to the mean of the population, and incorporates both analytical and biological variability. In the present study CV was approx. 25% for the n-6 PUFAs, and approx. 50% for n-3 PUFAs. These values are in line with, or higher than, what has been previously reported [150, 156]. Generally, reported CVs tend to be higher for the FAs present in lower concentrations [150].

5.5.1.3 Within and between subject variation

Overall reproducibility of analysis of plasma FAs in repeated samples from the same person has been demonstrated in both short-term (weeks) and long-term (years) studies [150]. However, most studies have used fasting blood samples, and our subjects were not required to

fast before blood sample collection. This poses a limitation to the measurement of FAs, and restricts our ability to make conclusions regarding long-term exposure to differing FA ratios, as discussed in paper III. The variation of concentrations of FAs in plasma of different people may be influenced by other factors than the diet and use of supplements. Some of these factors may also influence the gene expression profiles, and must be evaluated as potential confounders. Relevant confounders include potential diseases, lifestyle factors (e.g. smoking, exercise), and genetic polymorphisms that influence enzyme activity [150, 153]. Potential confounders for which we had information were assessed during inclusion/exclusion of subjects, or were evaluated in the comparison groups of paper III.

5.5.2 25(OH)D measurements

5.5.2.1 Pre-analytical issues

In paper II, plasma concentrations of 25(OH)D were measured using HPLC and UV detection [119]. Studies have shown that the stability of 25(OH)D in plasma is very high, and serum half-life of 25(OH)D is approximately 3 weeks [157]. The chosen method is considered one of the gold standard methods of 25(OH)D measurements [158, 159]. The method determines the sum of $25(OH)D_2$ and $25(OH)D_3$ [119]. There are several reasons why 25(OH)D concentration is used as a measure of vitamin D status, instead of any other vitamin D metabolites like the biologically active $1,25(OH)_2D$, separate measurements of $25(OH)D_2$ and $25(OH)D_3$, or free metabolites versus those bound to vitamin D binding protein [157]. Taken together, total 25(OH)D concentration is regarded as the best measure of vitamin D status, and reflects vitamin D obtained from both dietary intake and sun exposure over long periods of time [157].

5.5.2.2 Analytical error

Details and characteristics of the measurements (recovery, inter-assay variation, detection limit) were reported in paper II, and were in the range expected for the HPLC method [119]. It should be noted that discrepancies between analysis methods have been reported, so the

measured concentrations may not be directly comparable to those measured using other methods [159].

5.5.2.3 Within and between subject variation

Plasma concentration of 25(OH)D is known to vary with season, due to the contribution of UVB in sunlight to the formation of vitamin D in the skin. Seasonal changes of vitamin D status are well documented [13], and it was recently reported that season may coordinately influence both 25(OH)D concentration and the numbers and functions of T cell subpopulations [160] -and with that their gene expression profiles. Hence, season must be regarded as a potential confounder of the association between 25(OH)D and blood gene expression profiles. Because all blood samples used in the present study were collected in the spring, we considered adjustments according to season unnecessary. Also, due to the few months sample collection took place, we do not expect seasonal variation in T-cell subpopulations to contribute considerably to the gene expression profiles.

Cod liver oil is a significant source of vitamin A (retinol), and vitamin A may influence the cells of the immune system [161]. Various forms of the precursor vitamin are provided through the diet, and it is transported in the circulation in a protein-bound state, and then converted to the biologically active metabolite, retinoic acid, in different cells and tissues. Lastly, retinoic acid binds nuclear receptors to induce gene expression changes in target cells [77]. Importantly, the nuclear receptors involved in transducing signals from both FAs, $1,25(OH)_2D$ and retinoic acid often form hetero- or homodimers with each other [77]. Which partners form the receptor dimer determine the ligand specificity, and with that the biological consequences of ligand presence. These complex mechanisms of dimer formation, feedback mechanisms and modes of regulation have not been fully elucidated, but it has been suggested that $1,25(OH)_2D$ and retinoic acid may antagonize each other in competing for their common receptor binding partner, RXR [77, 162]. Unfortunately, we did not have data on vitamin A concentration available.

6 Discussion - Main results

The results presented in paper I-III show that both lifestyle and nutritional factors were associated with unique gene expression profiles in the blood of healthy, middle-aged Norwegian women. Other studies from the same sample material have been published, showing the influence of sex hormone levels [163] and perfluoroalkyl acids [164] on gene expression levels. Collectively, the results generated from the NOWAC Post-genome Cohort provide an extensive description of gene expression profiles in a general population, related to several exposures.

Conclusions made based on the results of any study must be adjusted to the strength of the presented data. When it comes to research on health impact of nutritional factors, meta-analyses and reviews tend to give conflicting results. Authors of such publications highlight several obstacles: the multifactorial nature of disease, difficulties in isolating effects of single nutrients, lack of data from randomized controlled trials, varying inclusion/exclusion criteria, and lack of standardization of intake assessment and/or biomarker measurements [13]. Methodological considerations for the present study were presented in the previous section. Importantly, the level of noise in the microarray data introduced by technical factors was high, and this may occlude low-level associations and hinder the potential use of gene expression profiles as biomarkers of the exposures in question.

Here, some general considerations regarding the approach of the NOWAC Post-genome Cohort will be discussed. Furthermore, both gene expression profiling as well as the cross-sectional study design may be regarded as methods for generating hypotheses for future studies. Based on our findings, some emerging hypotheses for how the investigated exposures may affect human biology will be highlighted.

6.1 Aspects of the NOWAC Post-genome Cohort

6.1.1 Use of human samples

In the NOWAC Post-genome Cohort, human sample material is used to study molecular effects of lifestyle and nutrition. Biological plausibility is central for the discussion of causality in epidemiological research, and it is listed by Hill as one of the criteria for causality [124, 165]. Data from cell lines and animal models has been extensively used as basis for evaluation of the biological plausibility of epidemiological findings. The accumulation of knowledge on human biology based on 'omics technologies has brought about an increased understanding of the limitations of relying on data from cells and model organisms to discuss causality in the multifactorial setting of human disease pathogenesis [111, 166]. An initiative to end the use of experimental animals has been launched at the pan-European level, in the form of an European Union directive stating that their goal is "(…) *full replacement of procedures on live animals for scientific and educational purposes as soon as it is scientifically possible to do so.*" (quote from [167]). In line with these trends, the NOWAC Post-genome Cohort was designed in an effort to enable study of molecular interactions in an in vivo, real-life setting, within a framework of careful epidemiological study design [152]. Sample material was collected to allow investigation using multiple 'omics technologies, for research on multiple levels of the biological information flow. As such, the Cohort coincides with the increased awareness of limitations of model organisms [166], and may contribute to increased understanding of the unique biology of humans. Time will show if the European Union directive will be followed by funding calls for relevant projects.

6.1.2 Systems-level thinking

In the clinical setting, the sensitivity and specificity of blood transcriptomics are assessed to evaluate the potential use of prognostic and diagnostic tests [113]. On the other hand, for research on molecular mechanisms and disease pathogenesis, validation of the proposed mechanisms is imperative for exploring causality. Such validation would include data on multiple levels of the biological information flow, and, ideally, congruent findings at the physiological, clinical and epidemiological levels. This perspective entails what is now known as systems biology, or even systems epidemiology [168]. The data presented here fall within

the category of research on molecular mechanisms and disease pathogenesis, and the inherent limitations that restricts the project's ability to determine causality have been discussed. As a pilot project for the NOWAC Post-genome Cohort, the exploration of gene expression profiles presented here provides a glimpse of the complex functional interaction between genes and the environment. The efforts to approach the systems-level thinking within the Post-genome Cohort will be further discussed in the chapter on future aspects.

6.1.3 The agnostic approach

The concept of non-hypothesis driven research has emerged as a consequence of the now wide-spread use of high-throughput 'omics technologies that enable screening of the full genome, transcriptome, or other -omes. The 'omics technologies may be used for what was previously regarded as "fishing trips", where the researcher would look for any associations without pre-defined biological hypotheses. The approach was met with initial skepticism, but is now increasingly being regarded as valid, with acronyms like explorative, agnostic, non-hypothesis driven, and discovery-based research coming to use. This development is closely linked to the large amounts of data generated by the 'omics technologies [169]. However, data analysis strategies for non-hypothesis driven research are scarce, and development is challenging because the languages spoken by scientists of different fields (molecular biology, epidemiology, statistics, etc.) are inherently different [169]. Contrasting the idea of non-hypothesis driven analysis, is the use of gene set/pathway-level analysis in the present thesis. Aspects of this analysis approach was discussed in chapter 5.4.1.3.

6.2 Associations between gene expression profiles and exposures

Our findings support the hypothesis that lifestyle and dietary compounds are associated with gene expression profiles in the blood. The influence of the investigated exposures may ultimately affect the onset and pathogenesis of several diseases including cancer and CVD. The biological processes described for the influence of the immune system on cancer development are numerous and sometimes conflicting. In their seminal paper of 2011 [46], Hanahan and Weinberg proposed that tumor-promoting inflammation and evasion of

immune destruction are emerging characteristics of cancer development. These processes may involve cells of both the innate and the adaptive immune system. For example, macrophages and mast cells are among those cell types suggested to contribute to induction and promotion of tumor cell development [46]. Furthermore, CVD is a chronic inflammatory disease which entails the activation of pro-inflammatory mechanisms and recruitment of immune cells and platelets to sites of vascular damage [170]. Known risk factors for CVD may influence immune cell reactivity, including smoking and a high fat diet. Our findings may serve to further elucidate the molecular mechanisms involved.

6.2.1 General tendencies in gene expression profiles of healthy women

The results in paper I confirmed the findings of other studies regarding the influence of technical variables on gene expression profiles [110]. These issues were discussed in chapter 5.4. Smoking, BMI, and fasting status were found to influence expression profiles considerably. On the other hand, age showed modest influence on our data, probably related to the narrow age distribution of the study population. Both BMI and smoking were associated with the differential expression of genes related to inflammation. Chronic low-grade inflammation has been suggested as an important link between obesity and increased risk for type 2 diabetes, metabolic syndrome, and atherosclerosis [171]. Identified core probes included granzyme B and Fc fragments, and these genes were highlighted as parts of the most significant gene sets in the functional clustering (Paper I, Table 4). These findings parallel the results presented in paper III, for expression levels of similar pathways in persons with FA ratios dominated by n-6 PUFAs. This might point to potentially converging mechanisms related to BMI and unfavorable FA ratios. Contrary to what might be expected from its link to cancer development, smoking status was not associated with major pathways related to cell viability, apoptosis or DNA repair in this dataset. Rather, in this healthy study population, pathways related to metabolism of xenobiotics (several CYP genes) and wound healing (fibronectin gene set) were identified (paper I, Table 2). These results were not in overall accordance with a recent study of blood cells exposed to cigarette smoke extract *in vitro*. Related to this discrepancy, and as mentioned previously (chapter 1.5), the limited degree of agreement between humans and model systems (cell cultures and animals) is an emerging

concept in biomedical research. Smoking is also a major risk factor for CVD, and data is accumulating in support of a causal link from smoking, via activation of the immune system, to CVD [172]. Blood gene expression profiling may prove a valuable tool to characterize this potential causal link, as was demonstrated by recent studies [173, 174].

6.2.2 Gene expression profiles associated with fatty acid ratios

Of the FA ratios that were analyzed, the plant-derived LA/ALA ratio had the largest effect on gene expression profiles (paper III, Table 3), and it was associated with the strongest influence on immune-related gene expression pathways. Our initial expectation was that the AA/EPA ratio would exert the strongest influence on immunological pathways compared to the other ratios, as these PUFAs serve as precursors for the immunoactive eicosanoids. Because the conversion of LA and ALA to AA and EPA is limited in humans [25], these findings may point to the importance of other regulatory mechanisms affected by PUFAs besides eicosanoid production. The precise mechanisms cannot be elucidated based on our data, but PUFAs may interact directly with nuclear receptors or in a more indirect way affect immune cell functions, as discussed below and summarized in Figure 9.

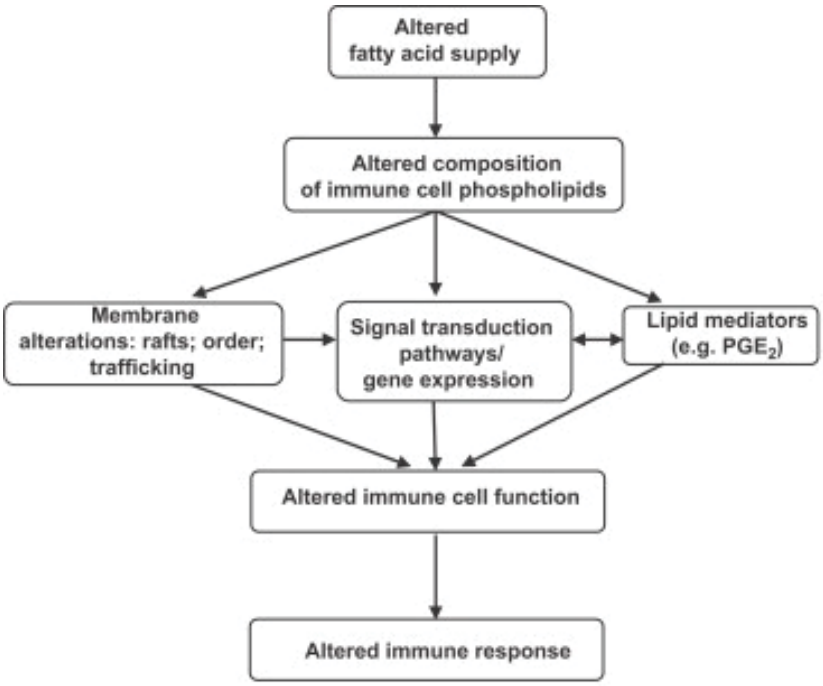

Figure 9: Mechanisms by which altered supply of fatty acids could affect immune responses. Abbreviations: PG, prostaglandins. Reprinted with permission [175].

Receptors that have been implicated in PUFA biology include PPARs and TLRs. PPARγ targets and binding partners were differentially expressed according to LA/ALA deciles (RXRA and fatty acid binding protein 4, FABP4, both up-regulated in the highest decile, paper III, Table 4 and Supplemental Table S2). Plasma levels of FABP4 are predictors of atherogenic dyslipidemia and metabolic syndrome, and may point to potential risk of developing the syndrome in the highest LA/ALA decile [176]. The PPARγ target ATP-binding cassette sub-family G member 1(ABCG1) was expressed at higher levels in the lowest AA/EPA and n-6/n-3 deciles (paper III, Table 5 and 6), indicating increased PPARγ activity in the lowest deciles compared to the highest deciles. PPARγ ligands include FAs and their metabolites, and activation is important for regulation of whole-body lipid metabolism and insulin sensitivity [67, 177]. The use of synthetic PPARγ ligands has been described as a gold standard for targeted treatment of metabolic disease [67]. Interestingly, an anti-inflammatory role was described for PPARγ in macrophages and in the immune system in general [67]. Along with the indication of higher PPARγ activity in the lowest FA ratios, autophagy markers were more highly expressed in the same groups. This was in line with the findings of a recent cell culture study after n-3 supplementation, where PPARγ activation by n-3 PUFAs was implicated in the induction of autophagy [70]. Autophagy is an intracellular degradation process which may provide ligands for PPARγ, thereby contributing to its anti-diabetic function [178]. The links between dietary FA metabolism, PPARs, immunology, autophagy, and diseases warrant further investigations. There have been reports on adverse effects of PPARγ activity. The use of synthetic PPARγ ligands for treatment of type 2 diabetes has been associated with increased risk for heart disease [67, 179], and PPARγ expression in monocytes was proposed as a link between smoking and the presence of atherosclerotic plaques [173]. Also, the induction of target genes may be highly tissue- or cell type specific. These somewhat contradictory results illustrate the complexity of PPARγ function, and complicates the interpretation of findings.

Another mechanisms that does not involve the eicosanoids is the putative influence of PUFAs on TLRs. TLR8 was expressed at a significantly higher level in the highest LA/ALA decile

compared to the lowest (paper III, Supplemental Table S2). This member of the TLR family is located in the intracellular membranes, and contrary to its plasma membrane-bound counterparts, TLR8 does not bind lipids, but nucleic acids [180]. Nucleic acid ligands were initially thought to derive only from pathogens (bacteria/virus), but it is now clear that endogenous nucleic acids derived from damaged cells may also activate TLRs [68]. Upon ligand binding, pro-inflammatory signaling is induced via nuclear factor κB (NF-κB) to produce cytokines and interferon β (IFN-β) [181]. However, much is still unknown about TLR8, compared to the other members of the TLR family. TLR8 was found up-regulated in active ulcerative colitis, perhaps linked to deregulation of the gut microbiota [182]. Also, TLR8 expression levels were associated with IL-1β expression, poor outcome, and increased inflammation after ischemic stroke [183]. In the latter case, the authors suggested the presence of debris after cellular damage as a possible cause for increased TLR8 expression. PUFAs may influence TLR signaling via modulation of membrane lipid rafts [184], which are cholesterol-rich microdomains important for co-localization of membrane proteins and receptor signaling. The mechanisms of n-3 PUFA influence in lipid rafts are not fully elucidated, but findings indicate reduced expression levels of raft-located TLRs, and disruption of the raft structure resulting in reduced protein clustering which is necessary for signal transduction [184]. In our dataset, the differential expression of genes related to cholesterol transport (ABCG1), oxidized lipids (OSBP2) and pathways for phospholipid transport may be related to differential TLR function dependent on FA ratios (paper III, Table 5, 6 and 7). Also, the significant overlap with pathways reflecting immune system response (paper III, Table 7), is in support of n-3 PUFA mediated modulation of TLR signaling. These mechanisms have previously been identified using *in vitro* and cell culture studies [185], and PUFA administration is emerging as a promising strategy for modulating lipid raft signaling [184].

6.2.3 Gene expression profiles associated with vitamin D status

In paper II, several pathways related to TLR signaling were identified (TLR signaling, innate immunity, ILR1 pathway, p38/MAPK signaling, Table 4). These findings point to the immuneregulatory function of vitamin D, which is emerging as a central hypothesis regarding benefits of sufficient vitamin D levels [77]. Also, TLR2 expression was associated with

vitamin D deficiency in gene set 65 (Biocarta Toll pathway). TLR2 resides in the plasma membrane and binds lipid ligands [180]. The potential activation of TLR2 in the vitamin D deficient group is supported by the identification of gene sets like cytokine production (gene set 52), as TLRs induce the production of these signaling molecules [180, 186]. TLR2 has been found expressed at higher levels in processes related to CVD (atherosclerosis, thrombosis), and is thought to contribute to the pro-inflammatory state in type 1 diabetes mellitus [187]. Also, TLR2 antagonists may have beneficial effects in myocardial infarction [187]. Collectively, there was a tendency for increased immune signaling in the vitamin D deficient group, in line with the current literature.

In the significant gene sets (paper II, Table 4), the direction of gene expression change in our samples were more in agreement with those reported from studies using human primary cell cultures and *in vivo* animal models, compared to those from studies using established cell cultures. This may of course be due to differences in study designs and other experimental factors, but it may also be viewed in light of the increased awareness of the discrepancies between model systems and human biology (chapter 1.5).

6.2.4 Converging pathways

Paper IV demonstrated that the FA ratios and vitamin D were only weakly correlated. Fatty fish and cod liver oil provides both n-3 PUFAs and vitamin D, but several sources including n-3 supplements and sun exposure only affects the levels of one of these nutritional factors. Our gene expression data showed no direct overlap between the profiles associated with the two exposures. However, the results may still suggest some converging pathways. Collectively, the data from paper II and III indicate that beneficial levels of both n-3 PUFAs and vitamin D may contribute to reducing inflammation and thereby reduce or inhibit chronic activation of related pathways. This can be exemplified by TLR-mediated inflammation, however, the differential expression of TLR was associated with the LA/ALA ratio (these PUFAs are mainly from plant sources), not AA/EPA. Favorable levels of both n-3 PUFAs (EPA and DHA) and vitamin D will be reached by consuming adequate amounts of fatty fish or fish oil. The potential influence of FAs on TLR-signaling could be investigated in a more focused design as

a candidate molecular mechanism related to fish. Furthermore, fish consumption is generally part of a healthy lifestyle, and reduced glucose exposure as part of a healthy eating pattern might serve to reduce TLR-mediated inflammation. Exposure to high glucose in diabetics induced TLR2 expression in several cell types and tissues [188, 189], and n-3 PUFAs are hypothesized to reduce TLR-mediated inflammation that is brought on by high glucose [189]. In addition, both AA/EPA and vitamin D were associated with gene sets and pathways related to major receptor signaling cascades (paper III, Supplemental Table S6, and paper II, Table 4). These included gene sets describing cell surface receptors and membrane bound receptors (for AA/EPA), as well as G-protein coupled signal transduction and nuclear receptor transcription (for vitamin D). More research is needed to explore these putative mechanisms in relation to disease prevention and progression, and as targets for therapeutic intervention.

6.3 Pieces of a puzzle

Determining causal relationships in complex biological systems has proven to be a challenging task for biomedical research. The influence of lifestyle and diet on human biology is multifactorial and intertwined, and the attempt to identify the impact of single exposures in this setting faces many obstacles. Following the sequencing of the human genome, technology development has allowed characterization of the biological response to diverse arrays of exposures, in the form of gene expression profiles. However, as the depth of analytical methods increases, so does our understanding of the inherent biological complexity. We are still not able to characterize in sufficient detail the levels of regulation that exist between an exposure in the form of, for example, inhaling cigarette smoke or intake of certain foods, via the impact of biochemical constituents on cells and organs, to their potential influence on disease development. Genetic differences and regulatory mechanisms that may influence the outcome are at play at all levels of the biological information flow (Figure 7), including methylation and other epigenetic mechanisms, miRNA and other non-protein coding transcripts, and multiple post-transcriptional and post-translational modifications. It follows from this that care must be taken in interpreting transcriptomics data, as mRNA abundance only makes up one piece of the puzzle.

7 Conclusions

Overall, the work presented herein demonstrates the feasibility of performing gene expression profiling in an epidemiological scale. Both technical variation, classical confounders, dietary FA ratios and vitamin D levels were reflected in blood gene expression profiles of middle-aged Norwegian women.

In more detail, the results provide a basis for the following conclusions:

1) Technical variability must be considered as part of rigorous study design in future projects, alongside considerations of the more classical confounding factors like BMI and smoking.

2) Gene expression differences according to vitamin D status were modest, but pathway-level analyses revealed associations with immunoregulatory processes. Results from *in vitro* studies were only partially reflected in our population-level analysis. Processes associated with vitamin D status may influence disease pathogenesis, and the results serve to shed light on the biological actions of vitamin D at the population level.

3) The three dietary FA ratios investigated (LA/ALA, AA/EPA, total n-6/n-3) had different impact of gene expression profiles. When comparing the highest versus the lowest of FA ratio deciles, differentially expressed genes and pathways related to all ratios may influence the pathogenesis of major diseases.

4) The concentrations of marine PUFAs, FA ratios, and vitamin D levels correlated to a limited degree. Hence, the gene expression results presented in paper II and III are likely to derive from the unique influence of the two factors respectively.

8 Future aspects

Despite decades of research, the multidimensional impacts of lifestyle and diet on health and disease remain to be fully elucidated. The same is true for the biological mechanisms at play during disease onset and progression. Improvements in analytical/measurement methods, study design and data analysis are warranted, in order to reach new insight into biological mechanisms and the potential for disease prevention.

8.1 Improved ways of exploring gene expression variation

Reducing variability related to laboratory procedures and other technical aspects of microarrays would help to decrease the influence of factors that are not related to the exposures of interest. Also, the establishment of consensus for data analysis would serve to increase the reproducibility and comparability of microarray data [110]. It should be noted that the ABI microarrays used in the present work have now gone out of production. In the NOWAC Post-genome Cohort, transcriptomics analyses using updated hybridization based bead arrays (Illumina) for assessment of gene expression profiles is ongoing, with two main goals. First, the potential use of blood gene expression profiles for prediction of breast cancer risk is being explored (Dumeaux *et al.*, manuscript submitted). In this setting, finding stable signals that are present in diverse populations is imperative for the identification of subsets of genes that hold predictive value. A blood-based test could be used in parallel with mammographic screening, to reduce the number of false positives and potential overtreatment, both of which are problematic issues related to such screening. Furthermore, performing a blood-based test may be more feasible than performing mammography, for example in remote areas of countries with few well-equipped hospitals and poor infrastructure. A second main goal of the blood transcriptomics currently being performed in the Post-genome Cohort is the characterization of the carcinogenic process. This goal comprises several aspects. First, collection of healthy breast tissue biopsies has been carried out to enable the use of healthy controls for gene expression comparison. Since the presence of a tumor affects adjacent tissue, the comparison to healthy controls provides an advantage

over studies using adjacent non-tumor tissue as controls. Second, potential co-variation between gene expression profiles in the tumor and in the blood of the same woman is being explored. This will provide additional insight into the *ex situ* physiological effects of the presence of a tumor. Lastly, by organizing case-control pairs by time before diagnosis, gene expression profiles that change up to the time of diagnosis can be identified. Time trajectories for gene expression profiles is defined as the compilation of prospective differences in gene expression for many independent case-control pairs, each representing one point in time (Lund, personal communication). Under the assumption that the carcinogenic process is reflected in the blood transcriptome, such trajectories may provide unique insight into the carcinogenic process, prior to its clinical manifestation [190].

Hybridization-based microarrays are now being replaced by next-generation deep sequencing technologies as the preferred tool for transcriptome-wide screening. Still, the hybridization based microarrays are increasingly being used for targeted investigations. Deep sequencing allows exploration of increasingly complex levels of biology, including alternative splicing of RNA, as well as discovery of new transcripts. However, the use of these technologies are faced with many of the same challenges as microarrays have been faced with, in addition to challenges specifically related to the deep sequencing technologies. "Sloppy science" [191] and over-communication of results, with congruent under-communication of study and technology limitations, may lead to many false positives, inflated expectations, and technology "hype". The resulting data could potentially be very promising for disease prevention, characterization and treatment, but the results may not be reproducible, and therefore futile [169, 192]. The ongoing technological development poses a challenge for epidemiological studies. The processing of large sample series takes time, and with time the technology changes. Embracing new products and methods may improve data quality and yield increasing analytical depth, but it also means that samples processed with previous versions of the technology may not be used for direct comparison to the newer data.

Lastly, in order to maximize the potential of new 'omics technologies, investments made in developing methods for all levels of data handling and analysis must match those investments

made for technology development [193]. Development of sound statistical methods and careful assessment of potentially new associations may yield biomedical insights which may ultimately translate into improved health outcomes [192].

8.2 Approaching the globolomic design

Systems biology and the globolomic design allows new analytic approaches to explore causality in epidemiology (Figure 10) [168]. Studies attempting to combine data from multiple 'omics analyses are now emerging. For example, one study explored the relationship between genetic variation (single nucleotide polymorphisms), epigenetics (DNA methylation), and gene expression [194]. Further combining these analyses with results from miRNA profiling, proteomics, metabolomics, and clinical data within a sound epidemiological study design may reveal new dimensions of the carcinogenic process [195]. Combining 'omics technologies within the fields of nutrigenomics and pharmacogenomics will allow exploration of nutrient-gene interactions and drug response on the basis of individual patients' genetic make-up. This may lead to the development of tailored nutritional supplementation for disease prevention, and targeted therapies [193]. Compilation of data from multiple population-based studies may be necessary to reach sufficient statistical power, in order to provide convincing evidence for multiple-level associations.

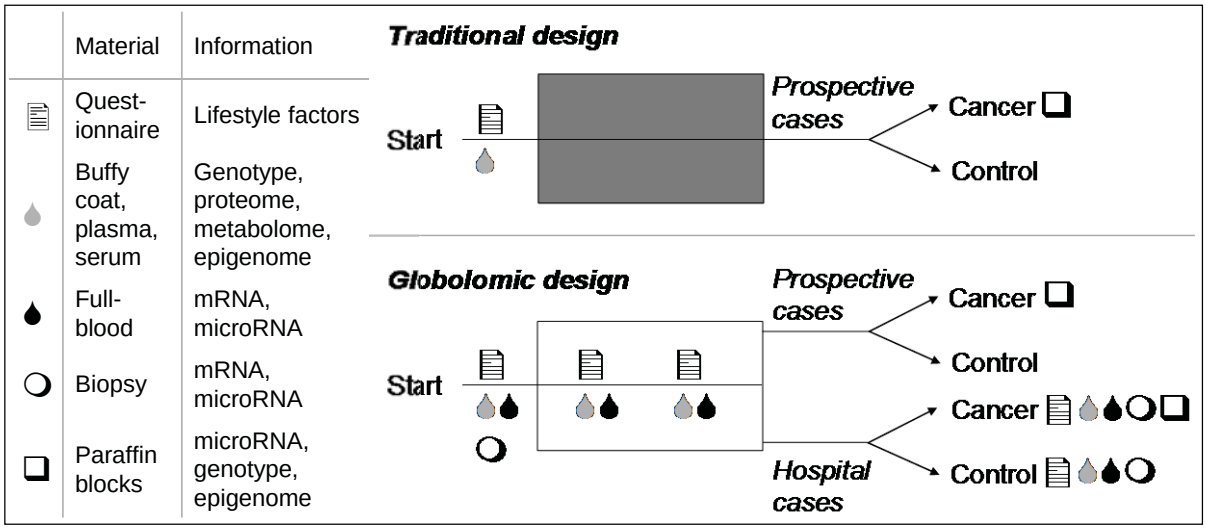

Figure 10: The globolomic design of the NOWAC Post-genome Cohort, compared to a traditional prospective study design.

9 References

1. WCRF and AICR, *Food, nutrition, physical activity, and the prevention of cancer: a global perspective*, 2007: Washington D.C.
2. Welch, A.A., et al., *Variability of fish consumption within the 10 European countries participating in the European Investigation into Cancer and Nutrition (EPIC) study.* Public Health Nutr, 2002. **5**(6B): p. 1273-85.
3. Ye, Y. and FAO, *Historical consumption and future demand for fish and fishery products. FAO Fisheries Circular No. 946*, 1999: Rome, Italy.
4. Molendi-Coste, O., V. Legry, and I.A. Leclercq, *Why and How Meet n-3 PUFA Dietary Recommendations?* Gastroenterol Res Pract, 2011. **2011**: p. 364040.
5. Romaguera, D., et al., *Is concordance with World Cancer Research Fund/American Institute for Cancer Research guidelines for cancer prevention related to subsequent risk of cancer? Results from the EPIC study.* Am J Clin Nutr, 2012. **96**(1): p. 150-63.
6. Vergnaud, A.C., et al., *Adherence to the World Cancer Research Fund/American Institute for Cancer Research guidelines and risk of death in Europe: results from the European Prospective Investigation into Nutrition and Cancer cohort study.* Am J Clin Nutr, 2013.
7. WHO, *Fact sheet no 297 - Cancer*, 2013.
8. WHO and IARC, *IARC Monographs on the evaluation of carsinogenic risks to humans - Tobacco smoke and involuntary smoking*, 2004: Lyon, France.
9. Farmer, P., *Introduction*, in *Biomarkers of carcinogen exposure and early effects*, P. Farmer and J. EMeny, Editors. 2006, ECNIS: Leicester, UK.
10. Vineis, P. and F. Perera, *Molecular Epidemiology and Biomarkers in Etiologic Cancer Research: The New in Light of the Old.* Cancer Epidemiology Biomarkers & Prevention, 2007. **16**(10): p. 1954-1965.
11. FAO and WHO, *Fats and fatty acids in human nutrition: report of an expert consultation*, 2010: Rome, Italy.
12. Alexander, J., et al., *A comprehensive assessment of fish and other seafood in the Norwegian diet*, 2006.
13. IOM, *Dietary Reference Intakes for Calcium and Vitamin D*, in *The National Academies Collection: Reports funded by National Institutes of Health*, A.C. Ross, et al., Editors. 2011, Institute of Medicine, National Academy Press: Washington (DC).
14. Hayes, D.P., *Adverse effects of nutritional inadequacy and excess: a hormetic model.* Am J Clin Nutr, 2008. **88**(2): p. 578S-581S.
15. Totland, T.H., et al., *Norkost 3 En landsomfattende kostholdsundersøkelse blant menn og kvinner i Norge i alderen 18-70 år, 2010-11. (Norwegian)*, 2012: Oslo, Norway.
16. Departementene, *Oppskrift for et sunnere kosthold. Handlingsplan for bedre kosthold i befolkningen (2007-2011) (In Norwegian)*, 2007: Olso, Norway.
17. Sala-Vila, A. and P.C. Calder, *Update on the Relationship of Fish Intake with Prostate, Breast, and Colorectal Cancers.* Critical reviews in food science and nutrition, 2011. **51**(9): p. 855-871.
18. Wakai, K., et al., *Dietary intakes of fat and fatty acids and risk of breast cancer: a prospective study in Japan.* Cancer Sci, 2005. **96**(9): p. 590-9.
19. Hedelin, M., et al., *Association of frequent consumption of fatty fish with prostate cancer risk is modified by COX-2 polymorphism.* International Journal of Cancer, 2007. **120**(2): p. 398-405.

20. Engeset, D., et al., *Fish consumption and breast cancer risk. The European Prospective Investigation into Cancer and Nutrition (EPIC).* Int J Cancer, 2006. **119**(1): p. 175-82.
21. Strobel, C., G. Jahreis, and K. Kuhnt, *Survey of n-3 and n-6 polyunsaturated fatty acids in fish and fish products.* Lipids in Health and Disease, 2012. **11**(1): p. 144.
22. Bolton, E., et al., *PubChem: Integrated platform of small molecules and biological activities. ,* in *Annual Reports in Computational Chemistry*2008, Americal Chemical Society: Washington, DC, USA.
23. Duplus, E. and C. Forest, *Is there a single mechanism for fatty acid regulation of gene transcription?* Biochemical Pharmacology, 2002. **64**(5-6): p. 893-901.
24. Bouwens, M., et al., *Fish-oil supplementation induces antiinflammatory gene expression profiles in human blood mononuclear cells.* Am J Clin Nutr, 2009. **90**(2): p. 415-424.
25. Plourde, M. and S.C. Cunnane, *Extremely limited synthesis of long chain polyunsaturates in adults: implications for their dietary essentiality and use as supplements.* Appl Physiol Nutr Metab, 2007. **32**(4): p. 619-34.
26. Spychalla, J.P., A.J. Kinney, and J. Browse, *Identification of an animal ω-3 fatty acid desaturase by heterologous expression in Arabidopsis.* Proceedings of the National Academy of Sciences, 1997. **94**(4): p. 1142-1147.
27. Jump, D.B., *The biochemistry of n-3 polyunsaturated fatty acids.* J Biol Chem, 2002. **277**(11): p. 8755-8.
28. Elmadfa, I. and M. Kornsteiner, *Fats and fatty acid requirements for adults.* Ann Nutr Metab, 2009. **55**(1-3): p. 56-75.
29. Johansson, L.R., et al., *Intake of very-long-chain n-3 fatty acids related to social status and lifestyle.* European journal of clinical nutrition, 1998. **52**(10): p. 716-21.
30. Retterstøl, K., *Omega-3 fettsyrer - mat og medisin (Norwegian).* Tidsskr Nor Laegeforen, 2012. **20**(132).
31. Brustad, M., T. Braaten, and E. Lund, *Predictors for cod-liver oil supplement use [mdash] the Norwegian Women and Cancer Study.* Eur J Clin Nutr, 2004. **58**(1): p. 128-136.
32. Simopoulos, A.P., *Evolutionary aspects of diet, the omega-6/omega-3 ratio and genetic variation: nutritional implications for chronic diseases.* Biomed Pharmacother, 2006. **60**(9): p. 502-7.
33. Strandvik, B., *The omega-6/omega-3 ratio is of importance!* Prostaglandins, Leukotrienes and Essential Fatty Acids, 2011. **85**(6): p. 405-406.
34. Simopoulos, A.P., *The importance of the omega-6/omega-3 fatty acid ratio in cardiovascular disease and other chronic diseases.* Experimental Biology and Medicine, 2008. **233**(6): p. 674-688.
35. Hibbeln, J.R., et al., *Healthy intakes of n−3 and n−6 fatty acids: estimations considering worldwide diversity.* The American Journal of Clinical Nutrition, 2006. **83**(6): p. S1483-1493S.
36. Notevarp, O., *Fettet vi spiser og hjertekarsykdommer (Norwegian),* 1961: Trondheim, Norway.
37. Dyerberg, J., H.O. Bang, and N. Hjorne, *Fatty acid composition of the plasma lipids in Greenland Eskimos.* The American Journal of Clinical Nutrition, 1975. **28**(9): p. 958-66.
38. Bang, H.O. and J. Dyerberg, *Personal reflections on the incidence of ischaemic heart disease in Oslo during the Second World War.* Acta Med Scand, 1981. **210**(4): p. 245-8.

39. Hooper, L., et al., *Risks and benefits of omega 3 fats for mortality, cardiovascular disease, and cancer: systematic review.* BMJ, 2006. **332**(7544): p. 752-60.
40. Sieri, S., et al., *Dietary fat and breast cancer risk in the European Prospective Investigation into Cancer and Nutrition.* Am J Clin Nutr, 2008. **88**(5): p. 1304-12.
41. Larsson, S.C., et al., *Dietary long-chain n-3 fatty acids for the prevention of cancer: a review of potential mechanisms.* Am J Clin Nutr, 2004. **79**(6): p. 935-945.
42. Signori, C., et al., *Chemoprevention of Breast Cancer by Fish Oil in Preclinical Models: Trials and Tribulations.* Cancer Research, 2011. **71**(19): p. 6091-6096.
43. von Schacky, C., et al. *Hooper Rebuttable*. 2006.
44. Ferlay, J., et al., *Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008.* Int J Cancer, 2010. **127**(12): p. 2893-917.
45. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.
46. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
47. Terry, P.D., T.E. Rohan, and A. Wolk, *Intakes of fish and marine fatty acids and the risks of cancers of the breast and prostate and of other hormone-related cancers: a review of the epidemiologic evidence.* Am J Clin Nutr, 2003. **77**(3): p. 532-543.
48. MacLean, C.H., et al., *Effects of omega-3 fatty acids on cancer risk: a systematic review.* JAMA : the journal of the American Medical Association, 2006. **295**(4): p. 403-15.
49. Shannon, J., et al., *Erythrocyte fatty acids and breast cancer risk: a case-control study in Shanghai, China.* The American Journal of Clinical Nutrition, 2007. **85**(4): p. 1090-1097.
50. Pauwels, E.K. and K. Kairemo, *Fatty acid facts, part II: role in the prevention of carcinogenesis, or, more fish on the dish?* Drug News Perspect, 2008. **21**(9): p. 504-10.
51. Berquin, I., et al., *ω-3 PUFAs: Interventional Trials for the Prevention and Treatment of Breast and Prostate Cancer*, in *Dietary Omega-3 Polyunsaturated Fatty Acids and Cancer*, G. Calviello and S. Serini, Editors. 2010, Springer Netherlands. p. 149-166.
52. Manson, J.E., et al., *The VITamin D and OmegA-3 TriaL (VITAL): Rationale and design of a large randomized controlled trial of vitamin D and marine omega-3 fatty acid supplements for the primary prevention of cancer and cardiovascular disease.* Contemporary Clinical Trials, 2012. **33**(1): p. 159-171.
53. Breslow, J.L., *n-3 Fatty acids and cardiovascular disease.* Am J Clin Nutr, 2006. **83**(6): p. S1477-1482.
54. Mozaffarian, D. and J.H. Wu, *(n-3) fatty acids and cardiovascular health: are effects of EPA and DHA shared or complementary?* J Nutr, 2012. **142**(3): p. 614S-625S.
55. Vedtofte, M.S., et al., *Dietary α-linolenic acid, linoleic acid, and n–3 long-chain PUFA and risk of ischemic heart disease.* The American Journal of Clinical Nutrition, 2011. **94**(4): p. 1097-1103.
56. Petrone, A.B., et al., *Omega-6 fatty acids and risk of heart failure in the Physicians' Health Study.* The American Journal of Clinical Nutrition, 2013. **97**(1): p. 66-71.
57. Wang, C., et al., *n-3 Fatty acids from fish or fish-oil supplements, but not alpha-linolenic acid, benefit cardiovascular disease outcomes in primary- and secondary-prevention studies: a systematic review.* The American Journal of Clinical Nutrition, 2006. **84**(1): p. 5-17.
58. Chapkin, R.S., et al., *Dietary docosahexaenoic and eicosapentaenoic acid: emerging mediators of inflammation.* Prostaglandins Leukot Essent Fatty Acids, 2009. **81**(2-3): p. 187-91.

59. Djousse, L., et al., *Dietary omega-3 fatty acids and fish consumption and risk of type 2 diabetes.* The American Journal of Clinical Nutrition, 2011. **93**(1): p. 143-50.
60. Wu, J.H., et al., *Omega-3 fatty acids and incident type 2 diabetes: a systematic review and meta-analysis.* Br J Nutr, 2012. **107 Suppl 2**: p. S214-27.
61. Kremmyda, L.-S., et al., *Atopy Risk in Infants and Children in Relation to Early Exposure to Fish, Oily Fish, or Long-Chain Omega-3 Fatty Acids: A Systematic Review.* Clinical Reviews in Allergy & Immunology, 2011. **41**(1): p. 36-66.
62. Wall, R., et al., *Fatty acids from fish: the anti-inflammatory potential of long-chain omega-3 fatty acids.* Nutr Rev, 2010. **68**(5): p. 280-9.
63. Calder, P.C., et al., *Inflammatory disease processes and interactions with nutrition.* Br J Nutr, 2009. **101 Suppl 1**: p. S1-45.
64. Calder, P.C., *Omega-3 fatty acids and inflammatory processes.* Nutrients, 2010. **2**(3): p. 355-74.
65. Nobel_Foundation. *The Nobel Prize in Physiology or Medicine 1982*. 2013; Available from: http://www.nobelprize.org/nobel_prizes/medicine/laureates/1982/.
66. Afman, L.A. and M. Müller, *Human nutrigenomics of gene regulation by dietary fatty acids.* Progress in Lipid Research, 2012. **51**(1): p. 63-70.
67. Ahmadian, M., et al., *PPAR[gamma] signaling and metabolism: the good, the bad and the future.* Nat Med, 2013. **99**(5): p. 557-566.
68. Fessler, M.B., L.L. Rudel, and J.M. Brown, *Toll-like receptor signaling links dietary fatty acids to the metabolic syndrome.* Curr Opin Lipidol, 2009. **20**(5): p. 379-85.
69. Jakobsen, C.H., et al., *DHA induces ER stress and growth arrest in human colon cancer cells: associations with cholesterol and calcium homeostasis.* Journal of lipid research, 2008. **49**(10): p. 2089-100.
70. Rovito, D., et al., *Omega-3 PUFA ethanolamides DHEA and EPEA induce autophagy through PPARgamma activation in MCF-7 breast cancer cells.* J Cell Physiol, 2013. **228**(6): p. 1314-22.
71. Duplus, E., M. Glorian, and C. Forest, *Fatty acid regulation of gene transcription.* J Biol Chem, 2000. **275**(40): p. 30749-52.
72. van Bussel, B.C., et al., *Fish Consumption in Healthy Adults Is Associated with Decreased Circulating Biomarkers of Endothelial Dysfunction and Inflammation during a 6-Year Follow-Up.* The Journal of Nutrition, 2011. **141**(9): p. 1719-25.
73. Weaver, K.L., et al., *Effect of dietary fatty acids on inflammatory gene expression in health humans.* Journal of Biological Chemistry, 2009. **284**(23): p. 15400-15407.
74. Yang, Y., et al., *Effects of n–3 PUFA supplementation on plasma soluble adhesion molecules: a meta-analysis of randomized controlled trials.* The American Journal of Clinical Nutrition, 2012. **95**(4): p. 972-980.
75. Myhrstad, M., et al., *Effect of marine n-3 fatty acids on circulating inflammatory markers in healthy subjects and subjects with cardiovascular risk factors.* Inflammation Research, 2011. **60**(4): p. 309-319.
76. Bouillon, R., et al., *Vitamin D and human health: lessons from vitamin D receptor null mice.* Endocr Rev, 2008. **29**(6): p. 726-76.
77. Mora, J.R., M. Iwata, and U.H. von Andrian, *Vitamin effects on the immune system: vitamins A and D take centre stage.* Nat Rev Immunol, 2008. **8**(9): p. 685-698.
78. Brustad, M., et al., *Vitamin D status of middle-aged women at 65-71 degrees N in relation to dietary intake and exposure to ultraviolet radiation.* Public Health Nutr, 2004. **7**(2): p. 327-35.
79. Edvardsen, K., O. Engelsen, and M. Brustad, *Duration of vitamin D synthesis from weather model data for use in prospective epidemiological studies.* Int J Biometeorol, 2009. **53**(5): p. 451-9.

80. Heaney, R.P. and M.F. Holick, *Why the IOM recommendations for vitamin D are deficient.* Journal of Bone and Mineral Research, 2011. **26**(3): p. 455-457.
81. Maxmen, A., *Nutrition advice: the vitamin D-lemma.* Nature, 2011. **475**(7354): p. 23-5.
82. Holick, M.F., et al., *Evaluation, Treatment, and Prevention of Vitamin D Deficiency: an Endocrine Society Clinical Practice Guideline.* Journal of Clinical Endocrinology & Metabolism, 2011.
83. Rosen, C.J. and C.L. Taylor, *Common misconceptions about vitamin D[mdash]implications for clinicians.* Nat Rev Endocrinol, 2013. **advance online publication**.
84. Kloster, J., *The distribution and frequency of rickets in one of the fishery districts of Finnmark and relation of diet to the disorder.* Acta Paediatr Scand 1931. **12**((suppl. 3)).
85. Bolland, M.J., et al., *Calcium and vitamin D supplements and health outcomes: a reanalysis of the Women's Health Initiative (WHI) limited-access data set.* The American Journal of Clinical Nutrition, 2011. **94**(4): p. 1144-9.
86. Levin, G.P., et al., *Genetic variants and associations of 25-hydroxyvitamin D concentrations with major clinical outcomes.* JAMA : the journal of the American Medical Association, 2012. **308**(18): p. 1898-905.
87. Zittermann, A., et al., *Vitamin D deficiency and mortality risk in the general population: a meta-analysis of prospective cohort studies.* The American Journal of Clinical Nutrition, 2012. **95**(1): p. 91-100.
88. Schöttker, B., et al., *Strong associations of 25-hydroxyvitamin D concentrations with all-cause, cardiovascular, cancer, and respiratory disease mortality in a large cohort study.* The American Journal of Clinical Nutrition, 2013. **97**(4): p. 782-793.
89. Ma, Y., et al., *Association Between Vitamin D and Risk of Colorectal Cancer: A Systematic Review of Prospective Studies.* Journal of Clinical Oncology, 2011. **29**(28): p. 3775-3782.
90. Lappe, J.M., et al., *Vitamin D and calcium supplementation reduces cancer risk: results of a randomized trial.* Am J Clin Nutr, 2007. **85**(6): p. 1586-91.
91. Chlebowski, R.T., et al., *Calcium plus vitamin D supplementation and the risk of breast cancer.* J Natl Cancer Inst, 2008. **100**(22): p. 1581-91.
92. Chung, M., et al., *Vitamin D with or without calcium supplementation for prevention of cancer and fractures: an updated meta-analysis for the U.S. Preventive Services Task Force.* Ann Intern Med, 2011. **155**(12): p. 827-38.
93. Wang, L., et al., *Circulating 25-Hydroxy-Vitamin D and Risk of Cardiovascular Disease: A Meta-Analysis of Prospective Studies.* Circulation: Cardiovascular Quality and Outcomes, 2012. **5**(6): p. 819-829.
94. Muldowney, S. and M. Kiely, *Vitamin D and cardiometabolic health: A review of the evidence.* Nutrition Research Reviews, 2011. **24**(1): p. 1-20.
95. Camargo, C.A., Jr., *Vitamin D and cardiovascular disease: time for large randomized trials.* J Am Coll Cardiol, 2011. **58**(14): p. 1442-4.
96. Bischoff-Ferrari, H., et al., *Benefit–risk assessment of vitamin D supplementation.* Osteoporosis International, 2010. **21**(7): p. 1121-1132.
97. Wu, S. and J. Sun, *Vitamin D, vitamin D receptor, and macroautophagy in inflammation and infection.* Discov Med, 2011. **11**(59): p. 325-35.
98. Haussler, M.R., et al., *Vitamin D receptor (VDR)-mediated actions of 1alpha,25(OH)2vitamin D3: genomic and non-genomic mechanisms.* Best practice & research. Clinical endocrinology & metabolism, 2011. **25**(4): p. 543-59.

99. Sun, X. and M.B. Zemel, *Calcitriol and calcium regulate cytokine production and adipocyte-macrophage cross-talk.* Journal of Nutritional Biochemistry, 2008. **19**(6): p. 392-399.
100. Cantorna, M.T., *Why do T cells express the vitamin D receptor?* Annals of the New York Academy of Sciences, 2011. **1217**(1): p. 77-82.
101. Hart, P.H., S. Gorman, and J.J. Finlay-Jones, *Modulation of the immune system by UV radiation: more than just the effects of vitamin D?* Nat Rev Immunol, 2011. **11**(9): p. 584-596.
102. Welsh, J., *Cellular and molecular effects of vitamin D on carcinogenesis.* Archives of biochemistry and biophysics, 2011.
103. Oh, J., et al., *1,25(OH)2 vitamin D inhibits foam cell formation and suppresses macrophage cholesterol uptake in patients with type 2 diabetes mellitus.* Circulation, 2009. **120**(8): p. 687-698.
104. Riek, A.E., J. Oh, and C. Bernal-Mizrachi, *Vitamin D regulates macrophage cholesterol metabolism in diabetes.* Journal of Steroid Biochemistry and Molecular Biology, 2010. **121**(1-2): p. 430-433.
105. Crick, F., *Central dogma of molecular biology.* Nature, 1970. **227**(5258): p. 561-3.
106. NIH. *An Overview of the Human Genome Project*. 2012; Available from: http://www.genome.gov/12011238.
107. Marshall, E., *Getting the noise out of gene arrays.* Science, 2004. **306**(5696): p. 630-1.
108. Shi, L., et al., *The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements.* Nat Biotechnol, 2006. **24**(9): p. 1151-61.
109. Chen, J.J., et al., *Reproducibility of microarray data: a further analysis of microarray quality control (MAQC) data.* BMC Bioinformatics, 2007. **8**: p. 412.
110. Shi, L., et al., *Reproducible and reliable microarray results through quality control: good laboratory proficiency and appropriate data analysis practices are essential.* Curr Opin Biotechnol, 2008. **19**(1): p. 10-8.
111. Seok, J., et al., *Genomic responses in mouse models poorly mimic human inflammatory diseases.* Proceedings of the National Academy of Sciences, 2013. **110**(9): p. 3507-3512.
112. Mestas, J. and C.C.W. Hughes, *Of Mice and Not Men: Differences between Mouse and Human Immunology.* The Journal of Immunology, 2004. **172**(5): p. 2731-2738.
113. Mohr, S. and C.-C. Liew, *The peripheral-blood transcriptome: new insights into disease and risk assessment.* Trends in Molecular Medicine, 2007. **13**(10): p. 422-432.
114. Fan, H. and P.S. Hedge, *The transcriptome in blood: challenges and solutions for robust expression profiling.* Current Molecular medicine, 2005. **5**: p. 3-10.
115. Dumeaux, V., E. Lund, and A.L. Borresen-Dale, *Comparison of globin RNA processing methods for genome-wide transcriptome analysis from whole blood.* Biomarkers in Medicine, 2008. **2**: p. 11 - 21.
116. Lund, E., et al., *The Norwegian Women and Cancer study, NOWAC External validity in a population-based national prospective study - the Norwegian Women and Cancer Study (NOWAC).* Cancer Causes Control, 2003. **14**: p. 1001 - 1008.
117. Lund, E., et al., *Cohort profile: The Norwegian Women and Cancer Study - NOWAC - Kvinner og kreft.* Int J Epidemiol, 2008. **37**: p. 36 - 41.
118. Kruhoffer, M., et al., *Isolation of microarray-grade total RNA, microRNA, and DNA from a single PAXgene blood RNA tube.* J Mol Diagn, 2007. **9**(4): p. 452-8.

119. Aksnes, L., *Simultaneous determination of retinol, alpha-tocopherol, and 25-hydroxyvitamin D in human serum by high-performance liquid chromatography.* J Pediatr Gastroenterol Nutr, 1994. **18**(3): p. 339-43.
120. Araujo, P., et al., *Evaluation of a rapid method for the quantitative analysis of fatty acids in various matrices.* J Chromatogr A, 2008. **1212**(1-2): p. 106-13.
121. Goeman, J.J., et al., *A global test for groups of genes: testing association with a clinical outcome.* Bioinformatics, 2004. **20**(1): p. 93-9.
122. Subramanian, A., et al., *Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(43): p. 15545-15550.
123. Goeman, J.J. and P. Buhlmann, *Analyzing gene expression data in terms of gene sets: methodological issues.* Bioinformatics, 2007. **23**(8): p. 980-7.
124. Grimes, D.A. and K.F. Schulz, *Bias and causal associations in observational research.* Lancet, 2002. **359**(9302): p. 248-52.
125. Bhopal, R., *Concepts of epidemiology*. Second ed2008, New York, USA: Oxford University Press.
126. Lund, E., et al., *External validity in a population-based national prospective study--the Norwegian Women and Cancer Study (NOWAC).* Cancer Causes Control, 2003. **14**(10): p. 1001-8.
127. Waaseth, M., *Hormone therapy use, sex hormone concentrations and gene expression - The Norwegian Women and Cancer study (NOWAC)*, in *Department of Community Medicine, Faculty of Health Sciences*2010, University of Tromsø: Tromsø, Norway.
128. Asare, A., et al., *Differential gene expression profiles are dependent upon method of peripheral blood collection and RNA isolation.* BMC Genomics, 2008. **9**(1): p. 474.
129. Menke, A., et al., *Peripheral blood gene expression: it all boils down to the RNA collection tubes.* BMC research notes, 2012. **5**(1): p. 1.
130. Rainen, L., et al., *Stabilization of mRNA Expression in Whole Blood Samples.* Clin Chem, 2002. **48**(11): p. 1883-1890.
131. Debey-Pascher, S., et al., *RNA-Stabilized Whole Blood Samples but Not Peripheral Blood Mononuclear Cells Can Be Stored for Prolonged Time Periods Prior to Transcriptome Analysis.* The Journal of Molecular Diagnostics, 2011. **13**(4): p. 452-460.
132. Feezor, R.J., et al., *Whole blood and leukocyte RNA isolation for gene expression analyses.* Physiol Genomics, 2004. **19**(3): p. 247-54.
133. Vartanian, K., et al., *Gene expression profiling of whole blood: Comparison of target preparation methods for accurate and reproducible microarray analysis.* BMC Genomics, 2009. **10**(1): p. 2.
134. Elashoff, M.R., et al., *Identification of Factors Contributing to Variability in a Blood-Based Gene Expression Test.* PLoS ONE, 2012. **7**(7): p. e40068.
135. Pazzagli, M., et al., *SPIDIA-RNA: First external quality assessment for the pre-analytical phase of blood samples used for RNA based analyses.* Methods, 2013. **59**(1): p. 20-31.
136. Allison, D.B., et al., *Microarray data analysis: from disarray to consolidation and consensus.* Nat Rev Genet, 2006. **7**(1): p. 55-65.
137. Canales, R.D., et al., *Evaluation of DNA microarray results with quantitative gene expression platforms.* Nat Biotechnol, 2006. **24**(9): p. 1115-22.
138. Barbacioru, C., et al., *Effect of various normalization methods on Applied Biosystems expression array system data.* BMC Bioinformatics, 2006. **7**(1): p. 533.

139. Jeanmougin, M., et al., *Should we abandon the t-Test in the analysis of gene expression microarray data: A comparison of variance modeling strategies.* PLoS ONE, 2010. **5**(9): p. 1-9.
140. Shi, L., et al., *The balance of reproducibility, sensitivity, and specificity of lists of differentially expressed genes in microarray studies.* BMC Bioinformatics, 2008. **9 Suppl 9**: p. S10.
141. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.* Journal of the Royal Statistical Society. Series B (Methodological), 1995. **57**(1): p. 289-300.
142. Mehta, T.S., et al., *Epistemological issues in omics and high-dimensional biology: give the people what they want.* Physiol Genomics, 2006. **28**(1): p. 24-32.
143. Noth, S., A. Benecke, and s.e. group, *Avoiding inconsistencies over time and tracking difficulties in Applied Biosystems AB1700TM/PantherTM probe-to-gene annotations.* BMC Bioinformatics, 2005. **6**(1): p. 307.
144. Liu, Q., et al., *Comparative evaluation of gene-set analysis methods.* BMC Bioinformatics, 2007. **8**(1): p. 431.
145. Whitney, A.R., et al., *Individuality and variation in gene expression patterns in human blood.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(4): p. 1896-1901.
146. Eady, J.J., et al., *Variation in gene expression profiles of peripheral blood mononuclear cells from healthy volunteers.* Physiol Genomics, 2005. **22**(3): p. 402-11.
147. Karlovich, C., et al., *A longitudinal study of gene expression in healthy individuals.* BMC Med Genomics, 2009. **2**: p. 33.
148. Joehanes, R., et al., *Gene Expression Analysis of Whole Blood, Peripheral Blood Mononuclear Cells, and Lymphoblastoid Cell Lines from the Framingham Heart Study.* Physiological Genomics, 2011.
149. Inouye, M., et al., *An Immune Response Network Associated with Blood Lipid Levels.* PLoS Genet, 2010. **6**(9): p. e1001113.
150. Hodson, L., C.M. Skeaff, and B.A. Fielding, *Fatty acid composition of adipose tissue and blood in humans and its use as a biomarker of dietary intake.* Progress in Lipid Research, 2008. **47**(5): p. 348-80.
151. Hodson, L., et al., *Stability of plasma and erythrocyte fatty acid composition during cold storage.* Clin Chim Acta, 2002. **321**(1-2): p. 63-7.
152. Dumeaux, V., et al., *Gene expression analyses in breast cancer epidemiology: the Norwegian Women and Cancer postgenome cohort study.* Breast Cancer Research, 2008. **10**(1): p. R13.
153. Arab, L., *Biomarkers of Fat and Fatty Acid Intake.* J. Nutr., 2003. **133**(3): p. 925S-932.
154. Baylin, A., et al., *Fasting Whole Blood as a Biomarker of Essential Fatty Acid Intake in Epidemiologic Studies: Comparison with Adipose Tissue and Plasma.* Am. J. Epidemiol., 2005. **162**(4): p. 373-381.
155. Astorg, P., et al., *Plasma n-6 and n-3 polyunsaturated fatty acids as biomarkers of their dietary intakes: a cross-sectional study within a cohort of middle-aged French men and women.* Eur J Clin Nutr, 2007.
156. Arab, L. and J. Akbar, *Biomarkers and the measurement of fatty acids.* Public Health Nutr, 2002. **5**(6A): p. 865-71.
157. Zerwekh, J.E., *Blood biomarkers of vitamin D status.* The American Journal of Clinical Nutrition, 2008. **87**(4): p. 1087S-1091S.
158. Hollis, B.W., *Measuring 25-hydroxyvitamin D in a clinical environment: challenges and needs.* The American Journal of Clinical Nutrition, 2008. **88**(2): p. 507S-510S.

159. Binkley, N., et al., *Assay variation confounds the diagnosis of hypovitaminosis D: a call for standardization.* The Journal of clinical endocrinology and metabolism, 2004. **89**(7): p. 3152-7.
160. Khoo, A.-L., et al., *Seasonal Variation in Vitamin D₃ Levels Is Paralleled by Changes in the Peripheral Blood Human T Cell Compartment.* PLoS ONE, 2012. **7**(1): p. e29250.
161. Ross, A.C., *Vitamin A and retinoic acid in T cell–related immunity.* The American Journal of Clinical Nutrition, 2012. **96**(5): p. 1166S-1172S.
162. Bastie, J.-N., et al., *1α,25-Dihydroxyvitamin D3 Transrepresses Retinoic Acid Transcriptional Activity via Vitamin D Receptor in Myeloid Cells.* Molecular Endocrinology, 2004. **18**(11): p. 2685-2699.
163. Waaseth, M., et al., *Sex hormones and gene expression signatures in peripheral blood from postmenopausal women - the NOWAC postgenome study.* BMC Medical Genomics, 2011. **4**: p. 29.
164. Rylander, C., et al., *Using blood gene signatures for assessing effects of exposure to perfluoroalkyl acids (PFAAs) in humans: the NOWAC postgenome study.* International Journal of Molecular Epidemiology and Genetics, 2011. **2**(3): p. 207-16.
165. Hill, A.B., *The Environment and Disease: Association or Causation?* Proc R Soc Med, 1965. **58**: p. 295-300.
166. *Of men, not mice.* Nat Med, 2013. **19**(4): p. 379-379.
167. EU, *Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes.* Official Journal of the European Union, 2010. **L 276**: p. 33-79.
168. Lund, E. and V. Dumeaux, *Systems epidemiology in cancer.* Cancer Epidemiol Biomarkers Prev, 2008. **17**(11): p. 2954-7.
169. Ransohoff, D.F., *Rules of evidence for cancer molecular-marker discovery and validation.* Nat Rev Cancer, 2004. **4**(4): p. 309-14.
170. Kanzler, I., et al., *Anti-inflammatory therapeutic approaches to reduce acute atherosclerotic complications.* Curr Pharm Biotechnol, 2012. **13**(1): p. 37-45.
171. Calder, P.C., et al., *Dietary factors and low-grade inflammation in relation to overweight and obesity.* British Journal of Nutrition, 2011. **106**(SUPPL. 3): p. S5-S78.
172. Brody, J.S., *Transcriptome alterations induced by cigarette smoke.* Int J Cancer, 2012. **131**(12): p. 2754-62.
173. Verdugo, R.A., et al., *Graphical Modeling of Gene Expression in Monocytes Suggests Molecular Mechanisms Explaining Increased Atherosclerosis in Smokers.* PLoS ONE, 2013. **8**(1): p. e50888.
174. Wright, W.R., et al., *Inflammatory Transcriptome Profiling of Human Monocytes Exposed Acutely to Cigarette Smoke.* PLoS ONE, 2012. **7**(2): p. e30120.
175. Calder, P.C., *Immunomodulation by omega-3 fatty acids.* Prostaglandins, Leukotrienes and Essential Fatty Acids, 2007. **77**(5-6): p. 327-335.
176. Cabré, A., et al., *Plasma fatty acid binding protein 4 is associated with atherogenic dyslipidemia in diabetes.* Journal of Lipid Research, 2008. **49**(8): p. 1746-1751.
177. Poulsen, L., M. Siersbaek, and S. Mandrup, *PPARs: fatty acid sensors controlling metabolism.* Semin Cell Dev Biol, 2012. **23**(6): p. 631-9.
178. Singh, R. and A.M. Cuervo, *Autophagy in the cellular energetic balance.* Cell Metab, 2011. **13**(5): p. 495-504.
179. Graham, D.J., et al., *Risk of acute myocardial infarction, stroke, heart failure, and death in elderly Medicare patients treated with rosiglitazone or pioglitazone.* JAMA : the journal of the American Medical Association, 2010. **304**(4): p. 411-8.

180. Barton, G.M. and J.C. Kagan, *A cell biological view of Toll-like receptor function: regulation through compartmentalization.* Nat Rev Immunol, 2009. **9**(8): p. 535-42.
181. Cervantes, J.L., et al., *TLR8: the forgotten relative revindicated.* Cell Mol Immunol, 2012. **9**(6): p. 434-438.
182. Sanchez-Munoz, F., et al., *Transcript levels of Toll-Like receptors 5, 8 and 9 correlate with inflammatory activity in Ulcerative Colitis.* BMC Gastroenterology, 2011. **11**(1): p. 138.
183. Brea, D., et al., *Toll-like receptors 7 and 8 expression is associated with poor outcome and greater inflammatory response in acute ischemic stroke.* Clin Immunol, 2011. **139**(2): p. 193-8.
184. Fessler, M.B. and J.S. Parks, *Intracellular Lipid Flux and Membrane Microdomains as Organizing Principles in Inflammatory Cell Signaling.* The Journal of Immunology, 2011. **187**(4): p. 1529-1535.
185. Shaikh, S.R., et al., *Docosahexaenoic acid modifies the clustering and size of lipid rafts and the lateral organization and surface expression of MHC class I of EL4 cells.* J Nutr, 2009. **139**(9): p. 1632-9.
186. Glass, C.K. and K. Saijo, *Nuclear receptor transrepression pathways that regulate inflammation in macrophages and T cells.* Nat Rev Immunol, 2010. **10**(5): p. 365-376.
187. Spirig, R., J. Tsui, and S. Shaw, *The Emerging Role of TLR and Innate Immunity in Cardiovascular Disease.* Cardiology Research and Practice, 2012. **2012**: p. 12.
188. Dasu, M.R., et al., *High glucose induces toll-like receptor expression in human monocytes: mechanism of activation.* Diabetes, 2008. **57**(11): p. 3090-8.
189. Dasu, M.R., S. Ramirez, and R.R. Isseroff, *Toll-like receptors and diabetes: a therapeutic perspective.* Clinical Science, 2012. **122**(5-6): p. 203-214.
190. Lund, E. and S. Plancade, *Transcriptional output in a prospective design conditionally on follow-up and exposure: the multistage model of cancer.* Int J Mol Epidemiol Genet, 2012. **3**(2): p. 107-14.
191. *Error prone.* Nature, 2012. **487**(7408): p. 406-406.
192. Ioannidis, J.P.A., *This I believe in genetics: discovery can be a nuisance, replication is science, implementation matters.* Frontiers in Genetics, 2013. **4**.
193. Costa, V., et al., *Uncovering the complexity of transcriptomes with RNA-Seq.* J Biomed Biotechnol, 2010. **2010**: p. 853916.
194. van Eijk, K., et al., *Genetic analysis of DNA methylation and gene expression levels in whole blood of healthy human subjects.* BMC Genomics, 2012. **13**(1): p. 636.
195. Hawkins, R.D., G.C. Hon, and B. Ren, *Next-generation genomics: an integrative approach.* Nat Rev Genet, 2010. **11**(7): p. 476-486.

PAPER I

Dumeaux V, Olsen KS, Nuel G, Paulssen RH, Borresen-Dale AL, Lund E (2010)

Deciphering normal blood gene expression variation

-The NOWAC postgenome study

PLoS Genet 6(3): e1000873

PubMed ID: 20300640

PAPER II

Olsen KS, Rylander C, Brustad M, Aksnes L, Lund E (2013)

Plasma 25 hydroxyvitamin D level and blood gene expression profiles: a cross-sectional study of the Norwegian Women and Cancer Post-genome Cohort

Eur J Clin Nutr. Advance online publication 6 March 2013

PubMed ID: 23462941

PAPER III

Olsen KS, Fenton C, Frøyland L, Waaseth M, Paulssen RH, Lund E (2013)

Plasma fatty acid ratios affect blood gene expression profiles

– A cross-sectional study of the Norwegian Women and Cancer Post-genome Cohort

PLOS ONE [accepted]

PAPER IV

Olsen KS, Aksnes L, Frøyland L, Lund E, Rylander C

Vitamin D status and PUFA ratios in a national representative cross-section of healthy middle-aged, Norwegian women -The Norwegian Women and Cancer Post-genome Cohort

[manuscript]

APPENDIX

Invitation letter

Information letter

Questionnaire

KVINNER OG KREFT

Du sendte i 2004 et utfylt spørreskjema til Institutt for samfunnsmedisin som del av den landsdekkende undersøkelsen "Kvinner og kreft". I følgeskrivet til spørreskjemaet informerte vi om at en del kvinner senere ville bli forespurt om de var villig til å avgi blodprøve. Du svarte ja til å sende en ny blodprøve i tillegg til den du allerede har gitt oss. Den nye prøven vil være spesielt egnet til å studere endringer i blodet over tid. Blodprøvene vil bli avidentifisert ved ankomst Institutt for samfunnsmedisin.

Formålet med blodprøven vil være:

- Måle nivå av vitaminer, mineraler og andre stoffer i blodet som kan settes i forbindelse med kostholdet.
- I fremtiden kunne studere de såkalte genetiske markører dvs. egenskaper i arvestoffet som kan disponere for kreft.
- Teste nye ideer eller hypoteser som oppstår i fremtiden.

Det er frivillig om du vil delta. Du kan trekke deg uten begrunnelse, og du kan be om at opplysninger du har gitt blir slettet, uten at dette vil få konsekvenser for deg.
Blodprøven vil kun bli benyttet til forskning og ingen resultater vil bli utlevert til deg eller noen andre. Blodprøven vil bli lagret i 30 år.

Ansvarlig for undersøkelsen er professor Eiliv Lund. Undersøkelsen er tilrådd av Regional komité for medisinsk forskningsetikk, Nord-Norge (REK NORD), og Datatilsynet har gitt konsesjon for oppbevaring av opplysninger.
Fremtidige forskningsprosjekter som vil benytte de lagrete blodprøvene vil forelegges Regional komité for medisinsk forskningsetikk, Nord-Norge (REK NORD).

Du kan finne mer informasjon om "Kvinner og kreft"og om forskningsresultatene på våre nettsider: www.ism.uit.no/kk/

Med vennlig hilsen

Eiliv Lund
professor dr.med.

Bente A. Augdal
prosjektmedarbeider

✂

Ønsker du ikke å delta og vil slippe påminning pr. brev ber vi deg fylle ut svar-slippen og returnere denne sammen med utstyret tilbake til oss (forseglet utstyr **må** ikke åpnes).

Jeg ønsker **ikke** å delta i blodprøvetakingen. .....

Underskrift

INFORMASJON
TIL DEG SOM ØNSKER Å DELTA

Hvis du ønsker å delta, må du ta kontakt med ditt legekontor, bedriftshelsetjeneste eller annen kyndig person og avtale tid for blodprøvetaking. **Det er viktig for prøvens holdbarhet at den tas mandag, tirsdag eller onsdag, slik at den kan nå oss via post innen fredag.**
Vedlagt utstyr og informasjon om prøvetakingen leveres til den som tar prøven.

Spørreskjema fylles ut prøvetakingsdagen og returneres til oss sammen med blodglassene.
Du vil ikke bli belastet med noen utgifter i forbindelse med prøvetakingen.

Utstyr:
- To prøveglass (1 stk rød kork, 1 stk blå kork)
- Nål til prøvetaking (kun et stikk i armen)
- Ett spørreskjema (til utfylling prøvedagen)
- Returkonvolutt for prøvene og spørreskjema

TIL PRØVETAKEREN

Vi ber om hjelp med prøvetaking av 2 blodglass, som skal benyttes til forskning i den nasjonale studien av brystkreft ”Kvinner og kreft”.

Deltakeren har mottatt det utstyr og de glass du behøver for å kunne hjelpe oss til å utføre denne delen av studien.

- Glassene merkes med ID-nr. til deltakeren.
- Fyll først det **røde** og deretter det **blå** prøveglasset med vanlig venepunksjon. Vær tålmodig, det røde glasset fylles sakte. Vend rørene forsiktig 8 – 10 ganger.
- Blodprøvene skal ikke sentrifugeres.
- Glassene legges i transporthylstrene og pakkes i returkonvolutten sammen med spørreskjemaet som deltakeren har fylt ut, konvolutten sendes oss snarest mulig.

Deltakeren skal ikke belastes med noen utgifter i forbindelse med blodprøvetakingen.
Betaling tilsvarende takst 701a (se baksiden) refunderes ved at det fylles ut en giro med kontonummer, og at denne sendes sammen med glassene tilbake til oss.

Takk for hjelpen!

Ønsker du mer informasjon kan du kontakte Bente A. Augdal telefon 77 64 66 38 eller Merethe Kumle telefon 77 64 48 84.

Prosjektet støttes av Norges forskningsråd.

Den norske lægeforening

Professor Eiliv Lund
Det medisinske fakultet
Institutt for samfunnsmedisin
Universitetet i Tromsø
9037 Tromsø

AM/3594/98/560.0 Oslo, 1.6.2003

Taking og sending av blodprøver i forb.m. undersøkelsen "Kvinner og kreft"

Vi viser til henvendelsen fra Det medisinske fakultet.

Legeforeningen anser at alle leger bør være positive til å delta i undersøkelsen "Kvinner og kreft" som jo kan få stor helsemessig og faglig betydning.

Vi vil også oppfordre til at leger aksepterer en betaling svarende til takst 701a for taking og innsending av blodprøvene, og håper selvsagt dermed at undersøkelsen får den nødvendige oppslutning. Legeforeningen forutsetter at kvinnen er klar over at hun skal informere om hva ærendet gjelder ved første henvendelse til legen.

Med vennlig hilsen

Terje Vigen
fung. generalsekretær

Øyvind Sæbø
forhandlingssjef

KVINNER OG KREFT

2005
KONFIDENSIELT

ID-nr:

LAB-kobling.

Følgende opplysninger fylles ut i forbindelse med blodprøvetaking.

DETTE SKJEMA **MÅ** FØLGE BLODPRØVEN!

Skjemaet skal leses optisk. Vennligst bruk blå eller sort penn. Du kan ikke bruke komma, bruk blokkbokstaver.

Jeg har lest informasjonen om blodprøveundersøkelsen og samtykker i å delta i denne: Ja: ☐

PRØVETAKINGSDAGEN

Fyll inn tidspunkt når blodprøven er tatt: Dato: dag ☐☐ mnd ☐☐

Klokkeslett: ☐☐ ☐☐

STILLING NÅR BLODPRØVEN BLE TATT

Sittende ☐
Liggende ☐

MENSTRUASJONSFORHOLD

Har du menstruasjon?

Ja ☐
Nei ☐
Uregelmessig ☐
Er gravid ☐

Hvis ja:
Angi dato for første dag i siste menstruasjon: dag ☐☐ mnd ☐☐

Når spiste du siste måltid før blodprøven ble tatt: Dato: dag ☐☐ mnd ☐☐

Klokkeslett: ☐☐ ☐☐

RØYKEVANER SISTE UKEN

Har du røkt i løpet av siste uke?

Ja ☐
Nei ☐

Hvis ja: Hvor mange sigaretter røkte du?

Antall i går: ☐☐
Antall i dag: ☐☐

VEKT OG HØYDE

Hvor mye veier du i dag? kg ☐☐☐

Hvor høy er du? cm ☐☐☐

Er disse målene tatt på legekontoret i dag?

Ja ☐
Nei ☐

MEDISINER I LØPET AV SISTE UKE

Har du brukt P-piller i løpet av siste uke?

Ja ☐
Nei ☐

Hvis ja:
Angi dato for siste tablett dag ☐☐ mnd ☐☐

Preparatnavn: .....
☐☐☐ (ikke skriv her)

Har du i løpet av siste uke brukt hormontabletter/-plaster (østrogen, gestagen) for overgangsalderen?

Ja ☐
Nei ☐

Hvis ja:
Angi dato for siste tablett dag ☐☐ mnd ☐☐

Preparatnavn: .....
☐☐☐ (ikke skriv her)

Preparatnavn: .....
☐☐☐ (ikke skriv her)

+

Har du brukt andre medisiner i løpet av siste uke?

Ja ☐
Nei ☐

Hvis ja:
Angi dato for siste tablett dag ☐☐ mnd ☐☐

Preparatnavn: .....
☐☐☐☐☐☐☐ (ikke skriv her)

dag ☐☐ mnd ☐☐

Preparat navn: .....
☐☐☐☐☐☐☐ (ikke skriv her)

+

dag ☐☐ mnd ☐☐

Preparat navn: .....
☐☐☐☐☐☐☐ (ikke skriv her)

BRUK AV KOSTTILSKUDD I LØPET AV SISTE UKE

Har du brukt tran (flytende) i løpet av siste uke? +

Ja ☐
Nei ☐

Hvis ja:
Angi dato du sist tok tran dag ☐☐ mnd ☐☐

Hvor mye tran tok du da?

☐ 1 ts ☐ 1/2 ss ☐ 1+ ss

Har du brukt trankapsler/Omega-3/fiskeolje i løpet av siste uke?

Ja ☐
Nei ☐

Hvis ja:
Angi dato du sist tok trankapsel/ Omega-3/fiskeolje dag ☐☐ mnd ☐☐

Hvor mange tok du da? +

☐ 1 ☐ 2 ☐ 3+

Navn på preparatet du tok sist:

.....
☐☐☐ (ikke skriv her)

Har du brukt soya i løpet av siste uke?

Ja ☐
Nei ☐

Preparatnavn: .....
☐☐☐ (ikke skriv her)

Preparatnavn: .....
☐☐☐ (ikke skriv her)

Har du brukt andre kosttilskudd (vitaminer/mineraler) i løpet av siste uke?

Ja ☐
Nei ☐

Hvis ja:
Angi dato for siste tablett dag ☐☐ mnd ☐☐

Preparatnavn: .....
☐☐☐ (ikke skriv her)

Preparatnavn: .....
☐☐☐ (ikke skriv her)

Takk for hjelpen!

ISBN xxx-xx-xxxx-xxx-x