Prostate Cancer Aetiology

Epidemiological studies of the IGF- and One-Carbon Metabolism Pathways

Mattias Johansson

Department of Surgical and Perioperative Sciences
Urology and Andrology, Umeå University, Umeå
Department of Medical Epidemiology and Biostatistics
Karolinska Institutet, Stockholm

Sweden 2008
To my parents
## THESIS AT A GLANCE

<table>
<thead>
<tr>
<th>Paper</th>
<th>Study population</th>
<th>Pathway</th>
<th>Exposure variables</th>
<th>Main findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>CAncer Prostate in Sweden (CAPS)</td>
<td>Insulin-like Growth Factor (IGF)</td>
<td>Genetic variation of the <em>IGF1</em> gene</td>
<td>Association between 3' genetic variant and increased prostate cancer risk</td>
</tr>
<tr>
<td>II</td>
<td>CAncer Prostate in Sweden (CAPS), Northern Sweden Health and Disease Cohort (NSHDC)</td>
<td>Insulin-like Growth Factor (IGF)</td>
<td>3' genetic variation of the <em>IGF1</em> gene</td>
<td>Association between 3' genetic variants and elevated plasma IGF1</td>
</tr>
<tr>
<td>III</td>
<td>CAncer Prostate in Sweden (CAPS)</td>
<td>Insulin-like Growth Factor (IGF)</td>
<td>Genetic variation of the <em>IGFBP1</em>, <em>IGFBP3</em> and <em>IGFALS</em> genes</td>
<td>Null association between genetic variants and prostate cancer risk and survival</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Plasma levels of total and intact IGFBP3</td>
<td>Positive association between intact IGFBP3 and risk of prostate cancer-specific death</td>
</tr>
<tr>
<td>IV</td>
<td>European Prospective Investigation into nutrition and Cancer (EPIC)</td>
<td>One-carbon metabolism</td>
<td>Plasma/serum levels of folate and vitamin B12</td>
<td>Null association with overall prostate cancer risk</td>
</tr>
<tr>
<td>V</td>
<td>Northern Sweden Health and Disease Cohort (NSHDC)</td>
<td>One-carbon metabolism</td>
<td>Plasma levels of ten B-vitamins and metabolites</td>
<td>Positive associations between choline, vitamin B2 and B12 and prostate cancer risk</td>
</tr>
</tbody>
</table>
Prostate cancer is the most common malignancy in Sweden and more than 9,000 men are diagnosed every year. The incidence of prostate cancer varies substantially worldwide, yet little is known about its causes. The aim of this thesis was to investigate the involvement of the insulin-like growth factor- and the one-carbon metabolism pathways in prostate cancer aetiology, studying both circulating biomarkers and genetic variation. Papers included in this thesis were conducted within the case-control study CAncer Prostate in Sweden (CAPS), and the two prospective studies European Prospective Investigation into nutrition and Cancer (EPIC), and Northern Sweden Health and Disease Cohort (NSHDC).

In paper I, we investigated the relation between genetic variants of the *IGF1* gene and prostate cancer risk within the CAPS study. A common haplotype in the 3′ region of the *IGF1* gene was found associated with increased prostate cancer risk.

In paper II, we investigated if the variants of the *IGF1* gene that were associated with prostate cancer risk in paper I, are also associated with circulating levels of IGF1. Plasma levels of IGF1 were analysed in controls from the CAPS study and three haplotype tagging SNPs (htSNPs) were genotyped in subjects from the NSHDC study in which circulating IGF1 had previously been analysed. The genetic variants associated with increased prostate cancer risk in paper I, were also found associated with elevated levels of circulating IGF1. We concluded that
variation in the 3’ region of the *IGF1* gene affects prostate cancer risk by influencing circulating levels of IGF1.

In paper III, we investigated if variants of the *IGFBP1, IGFBP3* and *IGFALS* genes are associated with i) prostate cancer risk, ii) circulating levels of total and intact IGFBP3, and iii) prostate cancer-specific survival probability. In addition, we investigated if circulating levels of total and intact IGFBP3 are associated with prostate cancer-specific survival probability. No clear association between genetic variation and overall prostate cancer risk or survival was observed, but we found a strong association between elevated levels of intact IGFBP3 and increased risk of prostate cancer-specific death. We could, however, not exclude that this association was confounded by treatment or the tumour itself.

In paper IV, we investigated if circulating levels of folate and vitamin B12 are associated with prostate cancer risk within the EPIC study. We observed no associations between levels of folate, vitamin B12 and overall prostate cancer risk, but elevated levels of vitamin B12 were associated with increased risk of advanced stage disease.

In paper V, we investigated if circulating levels of ten B-vitamins and related metabolites within the one-carbon metabolism pathway are associated with prostate cancer risk. Circulating levels of folate, vitamin B12, homocysteine, choline, betaine, methionine, vitamin B2, vitamin B6, cysteine, and methylmalonic acid (MMA) were analysed within the NSHDC study. Overall positive associations with prostate cancer risk were observed for levels of choline, vitamin B2 and vitamin B12, and inverse associations were observed for levels of homocysteine and MMA. We also observed a risk modification by smoking status on the association between vitamin B12 and risk; in non-smokers vitamin B12 was positively associated with risk, whereas the association between vitamin B12 and risk was inverse or null in ever/current-smokers.

In summary, our results suggest that genetic variation of the *IGF1* gene modifies prostate cancer risk by affecting circulating levels of IGF1. The association between circulating levels of intact IGFBP3 and prostate cancer-specific survival is intriguing, but further studies are needed to conclude if this association is causal or confounded. We also observed associations between several factors of one-carbon metabolism and prostate cancer risk, but these associations were statistically weak and require confirmation in other prospective studies.
TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS ................................. 12

BACKGROUND .......................................................... 13

Prostate cancer epidemiology ....................................... 13
  Incidence and mortality ........................................ 14
  Risk factors of prostate cancer ................................ 15

Genetics .................................................................... 19
  The human genome ............................................... 19
  Genetic variation .................................................. 20
  Genetic association studies .................................... 21
  Genetics of prostate cancer ................................... 24

Candidate pathways .................................................. 29
  The Insulin-like growth factor (IGF) pathway .......... 29
  The One-carbon metabolism pathway .................... 31

AIM OF THESIS ......................................................... 35

MATERIAL AND METHODS ....................................... 36

Study populations ................................................... 36
  CAncer Prostate in Sweden (CAPS) ......................... 36
  Northern Sweden Health and Disease Cohort (NSHDC) ... 38
  European Prospective Investigation into nutrition and Cancer (EPIC) ... 40

Biochemical analyses (Papers II, III, IV and V) .......... 43

Genotyping methods (Papers I – III) ......................... 43

Statistical methods ................................................ 44
LIST OF ORIGINAL PUBLICATIONS


All publications were reproduced with permission from the publishers.
Prostate cancer epidemiology

Prostate cancer is the most frequently diagnosed malignancy in men of the western world today. In 2006, the incidence rate for prostate cancer in Europe was almost as high as for breast cancer. Sweden has one of the highest incidence rates of prostate cancer in Europe with 9,300 cases diagnosed in 2006, compared to 7,200 diagnosed breast cancer cases the same year.

Figure 1. Age-standardized incidence- and mortality rates (per 100,000) in 2001.
Incidence and mortality

Prostate cancer incidence rates display a remarkable variation across the world. In the western countries, prostate cancer is the most frequently diagnosed cancer in men, whilst the disease is relatively rare in many south East Asian countries, see Figure 1. Mortality rates also vary substantially worldwide, but not to the extent of the incidence. In Sweden, about 2,500 prostate cancer-specific deaths are reported annually.

More than in any other cancer disease, incidence rates of prostate cancer must be interpreted in light of the diagnostic procedures, i.e. PSA testing (see Discussion, Study design – strengths and limitations). In Sweden, concurrently with the increasing use of the PSA test, the incidence has increased dramatically since 1995, see Figure 2. In the US, where opportunistic PSA testing is very common, the life-time risk of being diagnosed with prostate cancer is now 17 %, compared to 10 % in Sweden.

Perhaps not surprisingly, the distribution of prostate tumour characteristics has migrated to lower stages due to the frequent use of the PSA test. In 1986 in the US, less than one third of all diagnosed cases had an organ-confined tumour (T-stage 1 or 2), whereas today more than two thirds of all diagnosed cases have organ-confined tumours. This
shift is apparent also in Sweden where the number of cases diagnosed with T1c tumours, i.e. non-palpable tumours with no other symptoms than elevated PSA levels, more than quadrupled between 1996 and 2005⁵. If this change in tumour characteristics, with cases diagnosed in earlier stages with potentially curable tumours, will translate into improved survival rates is still not clear. However, it is becoming generally accepted that the increased use of PSA testing has led to a substantial over-treatment of many men that would have remained undiagnosed before the PSA era⁴.

Risk factors of prostate cancer
Prostate cancer is a disease of the elderly. In fact, in the western countries prostate cancer has the steepest age-dependant incline of any cancer. Even though prostate cancer is very common, little is known about its causes⁶. Apart from age, the only firmly established risk factors to date are family history, ethnicity and specific genetic variants (see Background, Genetics).

Ethnicity/Race
Ethnicity has consistently been associated with prostate cancer, and African-American men experience substantially higher risk compared to white Americans with standardized incidence rate ratios of about 1.7⁷. Moreover, African-Americans have 60 times higher incidence rates than men in Changhai, China. These observations are still poorly understood and are probably attributable to many different factors, including diagnostic practice. Migrant studies of Asian men moving to the US indicate that, even though the incidence increases, it does not reach the same levels as for the Caucasian- or African Americans⁸,⁹. This suggests that ethinical and regional differences in prostate cancer incidence are determined not only by diagnostic practice, but also by genetic differences and environmental factors.

Life-style related factors
The incidence patterns of prostate cancer across the world provide some interesting clues on the aetiology of the disease. The low incidence of south East Asia compared to the western countries raises the hypothesis that ecological differences, such as diet and obesity, are related to prostate cancer development.
Diet
Dietary and nutritional factors prevalent in south East Asia as compared to the western countries have been widely investigated in relation to prostate cancer. Ecological differences implicate intake of soy products and vegetables (common in south East Asia), and fat and meat intake as well as dairy products (common in the western countries), in prostate cancer aetiology.

There is some evidence that pulses (legumes) including soy and soy products protect against prostate cancer. These products contain compounds that may have cancer-protective effects because of their influence in oestrogen metabolism. Furthermore, phytoestrogens in pulses and soy may inhibit growth of the prostate.

Sub-optimal Vitamin D status has been linked to increased risk of prostate cancer. Calcium and diet high in calcium such as milk and dairy products have been consistently associated with increased risk of prostate cancer by results from ecological-, cohort-, and case-control studies. High calcium levels increase prostate cell proliferation by affecting vitamin D metabolism.

In contrast to some other cancer sites, vegetables have not been shown to protect against prostate cancer. The only consistent finding in relation to prostate cancer has been an inverse association for lycopene, found in processed tomatoes.

Fat intake has been frequently studied in relation to prostate cancer. Results from these studies have been mixed with a tendency to support a risk increasing effect associated with monounsaturated, animal, and saturated fats. More consistently, high dietary intake of red meat has been implicated in prostate cancer aetiology, possibly due to heterocyclic amines induced due to high temperature cooking. Several studies have also investigated if intake of fatty fish, rich on omega-3, protects against prostate cancer, but results supporting this hypothesis have been inconsistent.

Selenium and foods containing selenium have been inversely associated with prostate cancer by results from case-control and cohort studies, as well as randomized clinical trials. Suggested mechanisms for this relation include prevention of clonal expansion of tumours, promotion
of apoptosis and modulation of DNA repair mechanisms\textsuperscript{17}.

Many other dietary factors and nutrients, such as vitamin E, beta-carotene, alpha-tocopherol (vitamin D), zinc and multi-vitamins have been investigated in relation to prostate cancer, but the results remain inconclusive\textsuperscript{10,18–20}. Overall, no dietary factors have been convincingly linked to prostate cancer, but calcium probably increases the risk, whilst lycopene and selenium probably decrease the risk of prostate cancer\textsuperscript{10}.

\textit{Obesity and diabetes mellitus}

Obesity and type 2 diabetes mellitus (DM), conditions caused by genetic and lifestyle factors, have been frequently studied in relation to prostate cancer risk. Obesity increases risk of several cancers, including cancer of the colorectum, kidney, pancreas, endometrium and postmenopausal breast cancer\textsuperscript{21}. Similarly, DM has been shown to increase the risk of these cancers, as well as overall cancer risk\textsuperscript{22,23}. In contrast, DM has consistently been associated with a 10-15 \% decreased risk of prostate cancer\textsuperscript{24}. Such an inverse association is supported by recent genome-wide association studies showing that a variant of the \textit{TCF2} gene decreases the risk of DM, and also increases the risk of prostate cancer\textsuperscript{25}. Studies on obesity and overall prostate cancer risk implicate no important role for obesity as a risk factor\textsuperscript{21,26}. Recent studies, however, have shown a decreased risk of low-risk prostate cancer at diagnosis, but no, or an increased risk of high-risk disease\textsuperscript{27–29}. The latter is also supported by obesity being a risk factor of fatal prostate cancer\textsuperscript{30}. These observations; from studies of DM, obesity, and also of other related factors, indicate a complex association with prostate cancer, such that the aetiology behind early stage disease may differ from that of advanced disease and of prostate cancer death.

\textit{Other risk factors}

\textit{Inflammation}

Inflammation is frequently found in prostate cancer specimens and chronic inflammation (prostatitis) has since long been implicated in the disease. Inflammation stimulates angiogenesis, enhances cell proliferation and might damage DNA through radical oxygen species. Evidence supporting this hypothesis is now accumulating, both from epidemiological and molecular studies\textsuperscript{31}. In support for these observations, long-term intake of non-steroidal anti-inflammatory drugs (NSAIDs) has been
linked to reduced risk of prostate cancer\textsuperscript{32}.

Sexually transmitted diseases (STDs) have also been associated with increased prostate cancer risk. Implicated STDs include human immunodeficiency virus (HIV), syphilis, and recurrent gonorrhoea infections\textsuperscript{33,34}. Mechanisms behind these associations are not clear, but they might be due to induction of chronic inflammation of the prostate, thereby leading to cancer development.

\textit{Androgens}

Androgens have been implicated in prostate carcinogenesis, but prospective studies investigating circulating sex hormones have overall not supported this relationship. In a recent pooled study including most prospective studies on serum sex hormones, no association with prostate cancer risk was observed overall\textsuperscript{35}.

\textit{Benign prostatic hyperplasia}

Benign prostatic hyperplasia (BPH) is a histological condition leading to benign prostatic enlargement (BPE, increased size of the prostate gland), a frequent condition in middle-aged and elderly men. BPE often causes obstructive and irritative voiding symptoms referred to as lower urinary tract symptoms (LUTS)\textsuperscript{36}. BPH may cause serum PSA levels to rise but is currently not considered a precursor of prostate cancer. Related symptoms, however, e.g. LUTS, may lead to repeated PSA tests in order to rule out prostate cancer, and this follow-up may in turn lead to the discovery of small tumours.
Genetics

The human genome

The genome contains the blueprints of the human species, including instructions on how the body builds the cells. The human genome is constructed by deoxyribonucleic acid (DNA), consisting of about 3 billion DNA base pairs that are distributed on 24 chromosomes, see Figures 3 and 4. Men and women share the first 22 chromosomes but differ on the sex chromosome; men have one X- and one Y chromosome, whereas women have two X chromosomes. The genetic information of the human genome is coded by four bases: thymine (T); cytosine (C); adenine (A); and Guanine (G). In the DNA molecule, A bonds exclusively to T, and C bonds exclusively to G.

![DNA structure](http://commons.wikimedia.org/wiki/Main_Page)

Figure 3. The deoxyribonucleic acid (DNA) alpha helix molecule is a long polymer chain constructed of four nucleotides, thymine (T), cytosine (C), adenine (A) and guanine (G). These nucleotides are linked by a backbone made of sugars and phosphate groups. Figure adapted from wikimedia commons (http://commons.wikimedia.org/wiki/Main_Page).

It has been estimated that the human genome includes approximately 20,500 genes. Every gene contains regulatory regions called exons that are separated by non-coding regions called introns. Historically, introns were labelled junk DNA, but it is now understood that the introns contain important regulatory functions. In the making of a protein, the
gene is first transcribed into pre-mRNA, followed by a splicing step in which the introns are removed, creating mRNA. Thereafter, the remaining mRNA is translated according to the genetic code in which the nucleotides, in sets of three, encode an amino acid. Genes also contain a non-coding promoter region which provides information to the transcription machinery on where the transcription should start. The non-coding region before the gene (upstream) is called 5’ untranslated region (5’ UTR) and the non-coding region after the genes (downstream) is called 3’ UTR.

Figure 4. Within the cells, DNA is organised in chromosomes of which the human species have 24. Figure adapted from wikimedia commons (http://commons.wikimedia.org/wiki/Main_Page).

Genetic variation
Across any given population, there is a substantial amount of variation in the genome. These genetic variants - labelled polymorphisms if they occur in at least one of every hundred chromosome - make people look, behave and function differently. It has been estimated that the human genome contains approximately 10 million polymorphisms\(^3\). The most common type of polymorphism in the genome is the single nucleotide polymorphism (SNP), accounting for nearly 80 % of all genetic variants. A SNP is a single base substitution, see Figure 5. Other types of common polymorphisms include insertions, deletions, and recently it has been recognized that large regions of the genome sometimes get relocated, duplicated or deleted. These alterations are called copy number variants (CNVs) if they span more than one kilobase (kb), and they are believed to account for a significant proportion of phenotypic
variation\textsuperscript{39}.

Figure 5. A single nucleotide polymorphism (SNP). Figure adapted from wikimedia commons (http://commons.wikimedia.org/wiki/Main_Page).

Genetic association studies
The aetiology of many human disorders has a genetic component. Some disorders are caused by mutations in single genes that give rise to diseases such as cystic fibrosis and maturity onset diabetes of the young. These diseases are referred to as Mendelian diseases and their susceptibility loci are usually identified with the use of linkage studies where families, in which a specific disease is particularly prevalent, are screened with a large number of genetic markers across the genome\textsuperscript{40}. To date, more than 1,500 Mendelian disease loci have been identified\textsuperscript{41}. However, by nature, most Mendelian diseases are rare, and they may therefore be argued to be less important on the population level. Most common diseases are usually regarded as complex diseases in which many low-penetrant genetic variants, with odds ratios between 1 and 1.5, contribute to the susceptibility of the disease\textsuperscript{42}. Because complex diseases are common, they are thought to be caused by common genetic variants. This is the common disease – common variant hypothesis (CD-CV)\textsuperscript{43}. Studies aiming to find risk variants of complex diseases by linkage studies have not been successful\textsuperscript{44}. In general, it can be argued that monogenetic rare diseases are preferably studied in family based studies, e.g. linkage studies, whilst the genetics of complex diseases are more effectively studied in case-control studies\textsuperscript{44,45}.
Linkage disequilibrium and haplotype tagging

Even though the total number of genetic variants throughout the genome is overwhelming, there is a high degree of dependence between them. This is because loci from the same chromosome are physically linked and tend to be inherited together. However, during meiosis (one phase during cell division), chromosomes sometime recombine between two loci causing the linkage to break down, see Figure 6. The amount of linkage is measured in linkage disequilibrium (LD ≤ 1). If the LD between two SNPs is equal to 1, there is no evidence of recombination between them, and a maximum of three haplotypes are observed as defined by those SNPs. In population based studies, a haplotype is usually defined by a set of closely linked markers along a restricted region of a chromosome.

**Figure 6.** Figure showing how pair-wise LD of two SNPs is created and broken down by recombination events between the SNPs.

Linkage disequilibrium and haplotypes have become important tools in genetic association studies by making it possible to measure genetic markers indirectly by their allelic association with other genotyped variants. This procedure is usually referred to as tagging and the overall
aim is to attain information on all genetic variation within a region in a cost efficient manner, i.e. not spend money on redundant information. Many different tagging approaches have been proposed in the literature, but most of them can be divided into two groups: pair-wise tagging in which a subset of tagging SNPs (tSNPs) are selected aiming to explain other single SNPs (Figure 7), and haplotype tagging in which a subset of haplotype tagging SNPs (htSNPs) are selected aiming to explain all common haplotypes within a region (Figure 8).46

![Figure 7. Pair-wise tagging – Red SNPs are selected as tagging SNPs (tSNPs) and genotyped in a study population as proxies for the orange SNPs by pair-wise correlation. Blue SNPs are not genotyped nor explained by the tSNPs and the information from these would therefore be missed in an association study. Figure adapted from Kruglyak47.]

LD varies substantially across the genome but generally decreases when the distance between polymorphisms increases. It has also been recognized that LD patterns are not random, but often arrange in blocks of high LD48. Even though the number of SNPs within a haplotype block can be substantial, the haplotype diversity across the block tends to be minimal. The difference in haplotype variation and the number of haplotype tagging SNPs needed to explain the haplotype variation between a region of low LD and region of high LD is illustrated in Figure 8.

In order to select the tagging- or haplotype tagging SNPs for a region, one first has to assess the LD structure of that particular region. This can be done by genotyping a large number of SNPs within the region of interest in a number of representative individuals. In 2002, however, the HapMap project was initiated aiming to construct a haplotype map of the human genome49. To date, the HapMap project has genotyped 6.8 million SNPs across the human genome in a number of subjects of four populations. This data is publicly available on the internet and can be downloaded in order to assess the LD structure of any given region of the genome.
Genetics of prostate cancer

Is prostate cancer a genetic disease?

There is an overwhelming body of evidence in the literature suggesting that heredity plays an important role in prostate cancer aetiology. Epidemiological studies supporting this conclusion include a wide variety of designs, such as twin studies, case-control studies, cohort studies and family-based segregation analyses.
Case-control studies
A meta-analysis of eleven case-control studies investigating prostate cancer risk and family history of the disease estimated a pooled odds ratio of 2.5 (95% CI: 2.2-2.8) for subjects having at least one first-degree relative with prostate cancer compared to those who had none. This association was more pronounced in subjects of younger age (age < 65: OR = 4.3) than in older subjects (age $\geq$ 65: OR = 2.4). Steinberg et al. also showed a trend effect of having two (OR = 4.9) or more (OR = 10.9) first-degree relatives with prostate cancer compared to those who had none. These studies support a genetic component of prostate cancer, however, because case-control studies typically rely on self-reported data obtained after the diagnosis, they may be subject to bias as it is likely that cases are more aware of family members with the disease than the controls.

Cohort studies
A study design less sensitive to bias of various types is the cohort study in which subjects are recruited and followed until diagnosis or death. Another appealing feature of cohort studies, not available in case-control studies, is the possibility to calculate relative risks (RRs) or standardized incidence rates (SIRs), instead of odds ratios. Cohort studies have generally supported the results from the case-control studies, displaying RRs between 1.8 and 3.2, and SIRs between 1.6 and 1.7. In Sweden, Grönberg et al. reported a SIR of 1.7 in sons of men with prostate cancer. A consistent observation across these case-control- and cohort studies was that men whose brothers are affected by prostate cancer have higher risk than men whose fathers are affected. Some studies have therefore suggested that this may be due to an x-linked or recessive mode of inheritance. However, this observation may also be subject to bias because men whose brothers have prostate cancer may be expected to undergo screening to a further extent than men whose fathers have prostate cancer.

Twin studies
Twin studies provide a powerful approach to investigate if familial clustering of a trait is truly genetic, or if it is due to shared environmental factors. In twin studies, the concordance rate in monozygotic (MZ) twin pairs is compared to that of dizygotic (DZ) twin pairs. A higher
concordance rate among the MZ twin pairs, suggests a genetic component of the trait since MZ twins share 100 % of their genes, whereas DZ twins only share 50 % of their genes. Obviously, the validity of the twin studies relies on the assumption that both the MZ and the DZ twin pairs encounter similar environmental factors. In 2000, Lichtenstein et al. published a comprehensive register-based study conducted within the Nordic twin registers where the heredity of common cancers was investigated. The result supported a genetic component of several common cancers including breast-, colorectal- and prostate cancer. In particular prostate cancer displayed a strong result with concordance rates of 21% for MZ twins and 6.4 % for DZ twins. This resulted in an estimated hereditability of 42 % (95 % confidence interval (CI): 29-50 %) for prostate cancer, compared to 35 % and 27 % for colorectal and breast cancer, respectively. This study, together with two other studies, strongly suggests that a considerable part of prostate cancer risk is due to inherited genetic variants. 

Segregation analyses

Segregation analysis can be used to statistically determine if the familial clustering of a trait follows a particular genetic pattern, e.g. a dominant mode of inheritance, or an environmental pattern. In breast cancer, segregation studies have showed support of an autosomal dominant mode of inheritance with high penetrance. These studies provided the foundation for the discoveries of the breast cancer genes (BRCA1 and BRCA1) identified by linkage analysis. Similarly, in prostate cancer, segregation analysis have supported a rare (0.3-2.4 %) autosomal dominant allele, also with a high life-time penetrance (63-97 %). In Sweden, Grönberg et al. reported that an allele with a frequency of 1.7 % and a penetrance of 63 % gave the best fitting model. Overall, the result from these studies should be interpreted with caution because they were conducted in pedigrees that only included first degree relatives of the proband. In such small pedigrees with only fathers and sons, it is generally hard to distinguish between different modes of inheritance. Indeed, more recently, a more complex study using a method in which they simultaneously estimated the number of susceptibility loci, estimated the allele frequency and also allowed for covariates, supported the role of several loci, thus indicating a more complex pattern of inheritance.
Finding prostate cancer genes

**Linkage studies**
In the backwaters of studies of various designs supporting a genetic role in prostate cancer aetiology, a large number of studies have been conducted in order to find the specific polymorphisms modifying prostate cancer risk. Linkage studies can be used to identify regions across the genome where a susceptibility locus is likely to be found. Families, in which a specific disease is particularly prevalent, are screened with a large number of genetic markers across the genome\(^40\). In prostate cancer, linkage studies provided early hope of finding prostate cancer genes, such as the highly penetrant \textit{BRCA} genes for breast cancer\(^66\). Over the years, many studies have reported suggestive linkage on a number of loci across the genome, but most of these findings have not been replicated\(^50\). Generally, linkage studies are most powerful in finding regions with rare and highly penetrant genetic variants. Given the disappointing outcome of linkage studies, it seems likely that the genetic background of prostate cancer is more complex involving many low penetrant and interacting genes. To date, the only linkage study leading to the discovery of a convincing risk locus was performed in 323 prostate cancer families in Iceland\(^67\). In this study, Amundadottir \textit{et al.} found suggestive linkage (\textit{LOD score}=2.11) at chromosome 8q24 and identified the common rs1447295 SNP as the strongest associated variant. Subsequent studies have confirmed this SNP to be the first identified prostate cancer susceptibility locus with odds ratios ranging from 1.3 to 1.8 depending on the investigated population.

**Candidate gene studies**
Both animal- and cell line experiments, as well as linkage studies, have implicated specific genes and pathways in prostate cancer aetiology. In studies investigating these genes, i.e. candidate gene studies, polymorphisms are genotyped in a number of prostate cancer cases and healthy controls in which the allele frequencies are compared, most often by logistic regression. Associations are typically measured in odds ratios, reflecting the odds of having prostate cancer for subjects carrying the rare allele, compared with the odds of having the disease for subjects carrying the common allele. Overall, candidate gene studies have been inherently unsuccessful in finding prostate cancer genes. Over the last ten years, hundreds of studies have investigated single polymorphisms, as
well as tagging polymorphisms, in relation to prostate cancer risk. Even though many biologically plausible associations have been reported, none have been repeatedly confirmed. Frequently investigated pathways include the androgen biosynthesis and metabolism pathways, the growth factors and non-androgenic hormone pathways, the carcinogen metabolism pathway, the DNA repair pathway and the inflammation/angiogenesis/cytokines pathways.

**Genome-wide association studies (GWAS)**

Recent advances in high-throughput genotyping technologies have made it economically feasible to analyse hundreds of thousands of SNPs across the genome in large case-control series. Genome-wide studies (GWA) often adopt a multi-stage design, in which the first stage include genotyping of SNPs across the whole genome (>300,000) in a sufficiently powered case-control population. In the second stage, the most significant SNPs from the first stage are investigated in another study population. Sometimes a third stage is completed where the SNPs that are significantly associated with risk in both stage one and two are again investigated in an additional study population. This approach has been adopted by a number of genome-wide scans investigating prostate cancer over the last year. These studies have led to compelling evidence that genetic variants on chromosomes: 2p15, 3p12, 6q25, 7p15, 7q21, 8q24, 9q33, 10q11, 10q26, 11q13, 17q12, 17q24, 19q13, Xp11 and Xp11.22 predispose to prostate cancer susceptibility. Some of these loci include biologically plausible genes, such as the *MSMB* gene associated with recurrence after radical prostatectomy, the *LMTK2* gene which has been found to be mutated in prostate cancer cells and the *TCF2* gene which is associated with diabetes (see **Background, Prostate cancer epidemiology**). Other loci include regions with no genes, i.e. gene deserts. This observation, together with the modest risk modification associated with these variants, underlines the complex nature of the genetic background of prostate cancer. However, the first stage of most GWA studies have included about 2,000 cases, and they had thus limited power to find less common susceptibility loci. Most findings, to date, have therefore included SNPs with relatively common rare alleles. It is likely that combining these GWA studies over the next few years will lead to the discovery of additional, less common risk SNPs.
Candidate pathways

This thesis includes investigations of the insulin-like growth factor- and one-carbon metabolism pathways.

The Insulin-like growth factor (IGF) pathway
Insulin-like growth factors affect cell proliferation, differentiation, and apoptosis, and have therefore been implicated in cancer development of several sites, including breast, colorectal, lung and prostate.\(^{73}\)

Proposed mechanisms
Insulin-like growth factor 1 (IGF1) is polypeptide hormone similar to insulin in molecular structure.\(^{74}\) Circulating IGF1 is primarily produced in the liver, stimulated by growth hormone (GH) which is produced by the pituitary gland under the influence of somatostatin and growth-hormone-releasing hormone (GHRH), see Figure 9.

![Figure 9. Regulation of insulin-like growth factor 1. Figure adapted from Pollak et al.\(^{74}\).](image)

It has also been recognized that IGF1 can be synthesized in other organs.\(^{74}\) In the circulation, bioavailability of IGF1 is modulated by six binding proteins (IGFBP1-6), of which IGFBP3 binds almost 90 % of
circulating IGF1. When binding IGF1, IGFBP3 forms a ternary complex with acid labile subunit (IGFALS) which has a key role in regulating IGF1 release towards tissue \(^75,76\). Most of the remaining circulating IGF1 is bound to the other five binding proteins, or is circulating free. The binding proteins can prolong the circulating half-life of IGF1 or compete with IGF-receptors for ligands. How the binding proteins affect the biological action of IGF1 on the cellular level is still poorly understood, but studies on cell lines and mice have consistently shown that IGFBP3 decreases cell proliferation and induces apoptosis independently of IGF1\(^74,77\).

At the cellular level, the IGF1-receptor (IGF1R) stimulates cell proliferation, as well as inhibits programmed cell death (apoptosis)\(^74\). In studies of cell-lines and animal-models, increased \(IGF1R\) expression has been associated with neoplastic progression, particularly in prostate cancer.

**Epidemiological evidence to date**

Nutritional factors, especially the availability of energy and amino acids, are important determinants of circulating IGF1 levels\(^78\). However, heritability studies have shown that in well-nourished populations, a large part (40-60 %) of variation in circulating IGF1 is (co-)determined by genetic factors\(^79\). In epidemiological studies, elevated levels of IGF1 have been associated with several malignancies including breast, colorectal, lung and prostate by results from some – but not all – retrospective and prospective case-control studies\(^73\). In prostate cancer, high levels of circulating IGF1 have principally been considered as an established risk factor. However, as indicated in a systematic review by Renehan et al., this association is not conclusive\(^73\). In this review, elevated IGF1 levels were overall significantly associated with increased prostate cancer risk, but the included studies were few, and their results inconsistent. Recently, the Endogenous Hormones and Prostate Cancer Collaborative Group investigated the relation between circulating IGF1 and prostate cancer risk in 3,300 prospective cases and 4,450 controls, nested in twelve different cohorts. In this study (submitted), high circulating IGF1 levels were strongly associated (\(p_{\text{trend}} = 4 \times 10^{-7}\)) with a modestly increased risk of prostate cancer. The association between circulating levels of IGFBP3 and prostate cancer risk seem less consistent, and strength and
direction of associations have been mixed in different studies. Most studies investigating the IGF1 gene in relation to prostate cancer have analysed a microsatellite polymorphism (variable length CA repeat sequence) in the promoter region of the gene, but the result have been inconclusive. Only one study has investigated tagging polymorphisms in the IGF1 gene in relation to prostate cancer, and this study found some evidence for an association between two promoter SNPs and prostate cancer risk. Several polymorphisms have been associated with both circulating IGF1 and IGFBP3 levels. In particular the rare allele of the rs2854744 SNP in the IGFBP3 gene has been related to elevated circulating IGFBP3 in several studies.

Overall, elevated circulating levels of IGF1 are associated with increased risk of prostate cancer by results from prospective studies. Several genetic variants have been associated with circulating levels of IGF1 and IGFBP3, but it remains unclear if these polymorphisms translate into modified prostate cancer risk.

The One-carbon metabolism pathway
The one-carbon metabolism pathway, also referred to as folate metabolism, describes the metabolism of methyl groups (one-carbon groups). Factors of one-carbon metabolism, in particular folate – a B-vitamin found in dark green, leafy vegetables – have lately attracted increasing attention, both in the media and in medical research. In pregnant women, high levels of folate prevent neural tube defects in the developing fetus, which has led some countries, including the United States, to implement mandatory folic acid fortification of foods. To date, there is little and inconsistent evidence on the relation between factors of one-carbon metabolism and cancer.

Proposed mechanisms
A diagram of the one-carbon metabolism pathway is presented in Figure 10. Methyl groups are supplied by dietary serine, choline (via betaine) and methionine. Vitamin B6 acts as a cofactor in the transmethylation reaction of tetrahydrafolate (THF) to 5,10-methylene THF with the methyl groups from serine. 5,10-methylene THF is then reduced to 5-methylene THF in an irreversible reaction mediated by methylene-tetrahydrafolate reductase (MTHFR), for which vitamin B2 acts as a cofactor. 5-methylene THF is the predominant form of folate in the
circulation. Methyl groups are then donated from 5-methylene THF for the conversion of homocysteine to methionine, a reaction in which vitamin B12 acts as a cofactor. Homocysteine can also be metabolized to cysteine through reactions catalyzed by vitamin B6, a pathway that is important primarily in the postprandial state. The methionine derivative, S-adenosylmethionine (SAM), is the universal methyl donor for the methylation of a vast variety of molecules, including DNA.

**Figure 10. Diagram of the one-carbon metabolism pathway.**

Patterns of genomic hypomethylation and gene-specific promoter hypermethylation, with consequent silencing of gene expression in the latter, are often observed in tumour tissue. In prostate cancer, hypermethylation of CpG islands in the promoter of the Glutathione S-transferase π gene (GSTP1) have been observed in more than 90% of tumours, making this the most frequently reported epigenetic change in prostate cancer, and suggesting that DNA hypermethylation may be particularly important in prostate cancer development. High levels of folate or vitamin B12 might therefore promote DNA stability via genomic methylation, but they might also increase prostate cancer risk by inducing hypermethylation, though evidence for this relation is, to date, sparse. In addition to its central role in methylation, folate in the form of 5,10-methyltetrahydrofolate may have a protective role in cancer development by promoting the synthesis of thymidylate from uracil, minimizing the misincorporation of uracil into DNA. Excessive uracil
Prostate Cancer Aetiology - Background

can lead to double-strand breaks and possibly to cancer development\(^93\).

Two mechanisms are generally discussed by which one-carbon metabolism could influence cancer development, DNA methylation and DNA synthesis and repair. It has therefore been hypothesized that folic acid possesses a dual modulatory effect in cancer\(^90\). According to this hypothesis, folic acid might prevent neoplastic transformation in healthy tissue, but promote growth in a possibly undiagnosed pre-neoplastic or neoplastic lesion. One can consider that a sufficient pool of available methyl groups would, according to both these mechanisms, be desirable for both healthy and neoplastic tissue.

Epidemiological evidence

Despite initial epidemiological reports suggesting a possible cancer-preventive role for folate, particularly in colorectal cancer, prospective and intervention studies have generally been less promising\(^94\). Perhaps most surprising were the findings of the Aspirin-Folate Polyp Prevention Trial, a phase 3 secondary prevention study, in which a statistically significant higher frequency of subjects with three or more recurrent adenomas, as well as an increased risk of prostate cancer, were found in the folic acid supplementation arm of the study compared to the placebo arm\(^89\). In a breast-, prostate-, lung- and colorectal cancer screening study, subjects with high supplemental intake of folate had increased risk of postmenopausal breast cancer\(^95\). These unexpected results require confirmation, but may support the hypothesis of a dual modulatory role of folate in cancer\(^90\). There is little evidence on the relation between folate and prostate cancer risk, but some studies have indicated a risk increasing effect associated with high folate status\(^89,95-98\). In Europe, there is an ongoing debate regarding a mandatory folate fortification program, but partly because of the uncertain impact on cancer development, no consensus has yet been achieved.

The \textit{MTHFR} gene harbours a missense SNP, usually referred to as \textit{MTHFR} 677 C→T (rs1801133), whose rare allele causes a thermolabile variant of the MTHFR protein. The rare T allele of this SNP has been reported to decrease the enzyme activity of the MTHFR protein to 45 % of the more common C allele, and consequently to lower levels of circulating folate and higher levels of homocysteine, see Figure 10\(^99\). The \textit{MTHFR} 677 C→T SNP has therefore been implicated in various
diseases as a proxy for low folate levels which would not be sensitive to reversed causality along the lines of Mendelian randomization (see Discussion, The IGF pathway and prostate cancer)\textsuperscript{100}. We recently investigated this SNP in relation to prostate cancer within the CAPS study (see Material and Methods, Study populations), but found no strong evidence for an association with overall prostate cancer risk\textsuperscript{101}. Other studies of the \textit{MTHFR} 677 C→T in relation to prostate cancer have been small and inconclusive\textsuperscript{102-106}.

Other factors of one-carbon metabolism studied in relation to prostate cancer development include vitamin B12 and homocysteine\textsuperscript{96-98}. In 2004, an investigation from the NSHDC study (see Material and Methods, Study populations) reported a strong positive association between circulating levels of vitamin B12 and prostate cancer risk\textsuperscript{96}. Few other factors of one-carbon metabolism have been studied in relation to prostate cancer. Overall, it remains unclear if factors of one-carbon metabolism are important in prostate cancer development.
AIM OF THESIS

Overall aim

The overall aim of this thesis was to investigate if factors of the IGF- and one-carbon metabolism pathways play a role in prostate cancer aetiology.

Specific aims

The IGF pathway (Papers I – III)

• To investigate genetic variation of the IGF1 gene in relation to prostate cancer risk (Paper I)

• To investigate genetic variation of the IGF1 gene in relation to plasma levels of IGF1 (Paper II)

• To investigate genetic variation of the IGFBP1, IGFBP3 and IGFALS genes in relation to prostate cancer risk (Paper III)

• To investigate genetic variation of the IGFBP1, IGFBP3 and IGFALS genes in relation to plasma levels of total- and intact IGFBP3 (Paper III)

• To investigate genetic variation of the IGFBP1, IGFBP3 and IGFALS genes in relation to prostate cancer-specific survival probability (Paper III)

• To investigate plasma levels of total- and intact IGFBP3 in relation to prostate cancer-specific survival probability (Paper III)

The One-carbon metabolism pathway (Papers IV – V)

• To investigate circulating markers of one-carbon metabolism including folate, vitamin B12, homocysteine, choline, betaine, methionine, vitamin B2, vitamin B6, cysteine, and methylmalonic acid (MMA) in relation to prostate cancer risk (Papers IV – V)
MATERIAL AND METHODS

Study populations

CAncer Prostate in Sweden (CAPS)
The CAncer Prostate in Sweden (CAPS) study is a population based case-control study. Newly diagnosed prostate cancer cases 35-79 years of age in northern Sweden, and 35-65 years of age in southern Sweden, were invited to participate in the study, see Figure 11.

Figure 11. Recruitment area of the CAncer Prostate in Sweden (CAPS) study.
The study base included 67% of the Swedish population of in total 9 million citizens. Incident prostate cancer cases were identified between March 2001 and October 2003 through four out of six regional cancer registers in Sweden. The Swedish cancer register includes almost 100% of all diagnosed prostate cancer cases. Invitation letters were sent to the attending physician to ask for permission to invite their patients to the study. The physician then mailed a letter of information to the patient with information about the study. Recruitment to the CAPS study was completed in two separate rounds, CAPS 1 and CAPS 2. In CAPS 1, recruitment took place between March 2001 and September 2002, and subjects completed a comprehensive self-administered questionnaire including diet, family history of prostate cancer, physical activity and smoking. In CAPS 2, recruitment took place between October 2002 and October 2003, and subjects completed a more concise questionnaire including family history of prostate cancer, prostatitis, and use of non-steroid inflammatory drugs and aspirin use.

In total, 3,648 cases were invited to the study of which 3,155 (86%) agreed to participate. Blood samples were obtained from 2,965 cases. Clinical characteristics were obtained from the National Prostate Cancer Register including local tumour stage (T-stage), lymph node stage (N-stage), metastasis at bone scan (M-stage), tumour differentiation assessed by Gleason score and serum prostate specific antigen (PSA) level at time of diagnosis. Cases who reported at least one relative with prostate cancer were followed up by a second self-administered questionnaire and a telephone interview with a research nurse. Reported prostate cancer diagnosis in first-, second-, and third-degree relatives were thereafter verified and cases were classified as a hereditary prostate cancer (HPC) if fulfilling the Carter Criteria: three or more relatives diagnosed with prostate cancer in any nuclear family; prostate cancer in three successive generations in either of the probands’ paternal or maternal lineages; two first-degree relatives affected with prostate cancer at 55 years of age or younger. If a patient had a first-degree relative affected with prostate cancer he was classified as a familial prostate cancer (FPC). In total, CAPS consists of 2,862 cases with no family history, 206 FPC- and 87 HPC cases.

Controls were randomly selected from the Swedish population register within groups of men matching the case distribution for age (groups of
5-year intervals) and region (Figure 11). A total of 3,153 controls were invited to the study and 1,896 out of these men (60 %) agreed to participate and donated a blood sample.

Each study participant was identified through his unique national registration number. Using this registration number, complete follow-up for prostate cancer-specific mortality was achieved up until March 1st, 2007 through linkage to the Swedish Cause of Death Register. For individuals deceased after December 31st, 2003, cause of death was established through review of death certificates by an experienced oncologist. We defined prostate cancer-specific death as those who had prostate cancer classified as the underlying cause of death. The average follow-up time was 3.8 years (range 0.3 to 5.9 years). A total of 475 (17 % of cases included in paper III) individuals died during follow-up and of those, 324 had prostate cancer classified as their underlying cause of death.

All CAPS participants were asked to donate four 10 mL blood samples at their local health care centre. The blood samples were then sent by overnight mail to the Medical Biobank in Umeå where they were separated into serum, plasma, leukocytes and erythrocytes and stored at -70°C until analysis. Genomic DNA was extracted from leukocytes using standard techniques.

Written informed consent was obtained from all participants and the research ethical committee at the Karolinska Institutet and Umeå University Hospital approved the study.

Northern Sweden Health and Disease Cohort (NSHDC)
In the mid 1980’s it was recognized that the incidence rate of cardiovascular disease was notably high in the Västerbotten County of northern Sweden (approximately 254,000 inhabitants) compared to the rest of the country (Figure 12). To address this potentially serious public health problem, the Västerbotten Intervention Project (VIP) was initiated in 1985 in order to advocate a healthy diet and lifestyle. All residents in the Västerbotten County were invited to participate in the project by attending a health check-up at 40, 50 and 60 years of age. At the health check-up which was held at the local health care centre, participants were asked to complete a self-administered questionnaire including various demographic factors such as education, smoking
hhabits, physical activity and diet. In addition, height and weight were measured and participants were asked to donate a blood sample of 20 mL for future research. By the end of 2007, a total of 78,250 individuals had been recruited to the cohort. The VIP cohort is now one of three sub-cohorts included in Northern Health and Disease Cohort (NSHDC).

**Figure 12. Map showing the Västerbotten county in northern Sweden.**

Identification of incident prostate cancer cases in the NSHDC cohort was done by linkage with the Northern Sweden part of the National Prostate Cancer Register\(^\text{113}\). In January 2006, there were 641 prospectively collected cases in the cohort and of these, 568 had plasma- and DNA-samples available for analysis.

Two control subjects per case (1.9 on average) were selected randomly from all male cohort members who were alive and free of cancer (except non-melanoma skin cancer) at the time of the diagnosis of the index case. The controls were matched to the index case for age at recruitment (± 6 months) and date of recruitment (± 2 months).

Data on tumour characteristics of prostate cancer cases were obtained
from the National Prostate Cancer Register (NPCR)\textsuperscript{113}. Data included local tumour stage, lymph node metastasis at bone scan, serum prostate specific antigen level (PSA) at diagnosis and tumour differentiation assessed by WHO grading for cases diagnosed before 2000, and Gleason score for cases diagnosed after 2000. For cases diagnosed before 2000, tumour specimens were re-analysed by a pathologist and classified according to the Gleason system. In addition, we assessed total core length and total core length occupied by cancer in order to calculate the fraction of malignant tissue in the biopsies.

Blood samples were generally collected in the morning after 5 minutes of rest. After blood draw, the 20 mL blood sample was aliquoted into 10 sub-samples consisting of 6 plasma-, two leukocyte- (buffy coat) and two erythrocyte-samples, each containing 1.5 mL. Samples were thereafter placed in freezers within one hour after the blood draw at either -70 °C or -20 °C to be stored at a maximum of one week before shipment to the Northern Sweden Medical Biobank for long-term storage at -70 °C.

All NSHDC participants signed an informed consent form at the time of recruitment. The studies included in this thesis were approved by the Research Ethical Committee of Umeå University Hospital.

European Prospective Investigation into nutrition and Cancer (EPIC)
In 1992, the International Agency for Research on Cancer (IARC) initiated a comprehensive collaboration across Europe labelled “European Prospective Investigation into nutrition and Cancer” (EPIC)\textsuperscript{114,115}. The principal aim of the study was to investigate the aetiology of human cancer of various sites in relation to diet and other lifestyle factors, further motivated by difficulties in measuring dietary factors and the need for very large prospective studies in order to establish association between these factors and disease. Originally, there were seven EPIC countries included in the collaboration: France; Germany; Greece; Italy; The Netherlands; Spain; and the UK. These core countries were subsequently followed by three Nordic countries, Sweden, Denmark and Norway comprising in total 10 countries and 23 individual centres across Europe by May 2002, see Figure 13.

Men and women between 35 and 70 years of age were invited to participate in the study by mail or in person via the individual study
In total 519,978 subjects were recruited between 1992 and 2000 of which 70.5% were female.

Each individual who agreed to participate was sent two questionnaires on dietary- and non-dietary variables, respectively. Generally, the questionnaires were completed at home and the participants were then invited to a study centre where anthropometric measurements and a blood sample were retrieved. In total 385,747 participants donated a blood sample.

Follow-up procedures aiming to identify cancer cases in the cohort differed between the participating countries. Wherever possible – in Denmark, Italy, The Netherlands, Norway, Spain, Sweden and the UK – population cancer registers were used, whereas the other countries – France, Germany and Greece – relied on a combination of methods including health insurance records, cancer and pathology registers, and on active follow-up through study subjects and their next-of-kin.

In Paper IV of this thesis, seven sub-cohorts of the EPIC study
contributed with prostate cancer cases. After exclusion of cases with no available blood sample and those who had missing information on the date of blood donation or who had a history of another cancer (except non-melanoma skin cancer) at the time of blood donation, 547 prostate cancer cases remained (apart from NSHDC). Italy contributed with 61 cases, Germany 182, Greece 9, the Netherlands 24, Spain 94 and the UK 177 cases. For each case, one male control (two in NSHDC) was chosen at random from appropriate risk sets consisting of all cohort members alive and free of cancer (except non-melanoma skin cancer) at the time of diagnosis of the index case. Matching criteria were: study centre, age at enrolment (± 6 months), time of day of blood collection (± 1 hour) and time between blood draw and last consumption of food or drink (<3, 3-6, >6 hours). In total, 869 cases and 1174 controls including subjects from the NSHDC (see above) were included in Paper IV of this thesis. Data on stage and grade of disease were collected from each centre where possible.

After the blood draw, blood samples were aliquoted into several 0.5 mL heat sealed plastic straws, divided into plasma with sodium citrate, serum, erythrocytes and buffy coat for DNA. Half of each participant’s straws were then transported to IARC, Lyon, for storage in liquid nitrogen (-196 °C), whereas the remaining straws were stored locally. Exceptions to this procedure were in the Oxford centre were blood samples were collected throughout the UK and transported to a laboratory in Norfolk at ambient temperature, and in Sweden and Denmark were the blood samples were stored locally in tubes in -70°C or -80°C freezers.

All participants gave written informed consent to participate in the study and the study was approved by the local ethics committees in the participating countries. The study of Paper IV in this thesis was approved by the Internal Review Board (IRB) of the International Agency for Research on Cancer.
**Biochemical analyses (Papers II, III, IV and V)**

All assays of plasma IGF1, total- and intact IGFBP3 included in papers II and III were performed at the International Agency for Research on Cancer, Lyon, France, using enzyme-linked immunosorbent assays (ELISA) from Diagnostic Systems Laboratories (DSL, Webster, Texas). The IGF1 assay included an acid-ethanol precipitation step. The inter-batch CVs for IGF1 ranged from 8.4 – 10.5 %, for total IGFBP3 from 7.4 – 8.4 %, and for intact IGFBP3 from 5.5 – 19.3 %. The intra-batch CVs for IGF1 ranged from 3.2 – 3.6 %, for total IGFBP3 from 6.3 – 6.5 %, and for intact IGFBP3 from 5.2 – 5.6 %.

All biochemical analyses included in Papers IV and V were performed at BEVITAL AS (www.bevital.no), Bergen, Norway. Methionine was determined as the sum of methionine and methionine sulfoxide. Levels of methionine, homocysteine, cysteine and MMA were determined using an isotope-dilution gas chromatography–mass spectrometry (GC-MS) method. The within- and between-run CVs ranged from 2.1 – 3.6 % and from 2.1 – 8.1 %, respectively. Levels of methionine sulfoxide, choline and betaine were determined by liquid chromatography-tandem mass spectroscopy (LC-MS/MS). The within- and between-run CVs ranged from 2.1 – 7.2 % and from 3.5 – 8.8 %, respectively. Vitamin B6 and B2 were determined by LC-MS/MS. The within- and between-run CVs ranged from 3 – 20 % and 6 – 22 %, respectively. Levels of folate were determined by a *Lactobacillus casei* microbiological method. The within- and between-run variations were 6.0 and 6.3 %, respectively. Levels of vitamin B12 were determined by a *Lactobacillus leichmannii* microbiological assay. The within- and between-run variations were 5.4 and 6.7 %, respectively. Samples from cases and their matched controls were analysed together and positioned randomly within stratum triplets.

All biochemical assays were performed by laboratory personnel who were blinded to the case-control status of the blood samples.

**Genotyping methods (Papers I – III)**

Throughout papers I – III in this thesis, genotyping was performed at the International Agency for Research on Cancer (IARC), Lyon, France,
using the 5’ nuclease assay (TaqMan). The order of DNA samples from cases and controls was randomised on PCR plates in order to assure that an equal proportion of cases and controls could be analysed simultaneously. TaqMan probes were synthesised by either Applied Biosystems, Foster City, CA, USA (with MGB chemistry), or Proligo, Paris, France (with or without LNA chemistry). The reaction mix included 10 ng genomic DNA, 5 pmol of each primer, 1 pmol of each probe and 2.5 μl of 2× master mix (Applied Biosystems) in a final volume of 5 μl. The thermocycling included 50 cycles with 30 sec at 95°C followed by 60 sec at 60°C. PCR plates were read on an ABI PRISM 7900HT instrument (Applied Biosystems). Laboratory personnel were kept blinded to case–control status and quality control was achieved by analysing the concordance rates between duplicate samples. SNPs were analysed for deviations from Hardy-Weinberg equilibrium using χ² statistics. All SNPs included in Papers I – III conformed to Hardy-Weinberg equilibrium and had high (>99.7 %) concordance rates.

Statistical methods

Throughout the studies included in this thesis we used standard statistical methods, such as linear and logistic regression, in order to investigate various relationships. P-values were assessed as indicators of associations, and p-values of less than 0.05 were considered statistically significant.

Single nucleotide polymorphism (SNP) analysis (Papers I-III)

For each study subject and SNP, genotypes were re-coded using two dummy variables, the first indicating heterozygosity and the second indicating homozygosity of the minor allele. These two dummy variables were then included in the appropriate regression model (see below), using the two-degrees of freedom likelihood ratio test as indicator of association with the endpoint. Thus, genotype specific estimates relative the homozygote subjects of the major allele were achieved. We also assessed associations between SNPs and endpoint using a variable indicating the number of rare alleles carried by an individual, thus achieving a trend model.
Haplotype analysis and selection of haplotype tagging SNPs (htSNPs) (Papers I – III)

Haplotype analyses were only done within regions of limited haplotype diversity (haplotype blocks), i.e. within regions of high linkage disequilibrium (LD). These blocks were defined using genotype data for 30 Caucasian Centre d’Etude Polymorphisme Humain (CEPH) trios, downloaded from the HapMap database. For each gene, we aimed to include 10 kb upstream and 5 kb downstream to the gene. Blocks stretching further than 10 kb outside to the gene were kept intact if they overlapped parts of the gene or the promoter region.

In each haplotype block, haplotype tagging SNPs (htSNPs) were selected aiming to explain the haplotypes with a frequency of more than 5%. We used the criteria $R_h^2 > 0.8$ as described by Stram.

Individual haplotype frequencies were estimated for cases and controls combined, using a maximum likelihood method. For each study individual, the probabilities for being heterozygote and homozygote for each compatible haplotype were then estimated. These estimates, labelled haplotype dosages, were then included as covariates in the appropriate regression model (see below), using the homozygotes of the most common haplotype as reference category. These calculations were performed using the tagSNPs software. We used p-values from likelihood ratio tests as indicators of associations with endpoints for specific haplotypes or overall haplotype blocks.

Analysis of prostate cancer risk associations (Papers I, III, IV and V)

Throughout the papers included in this thesis, we used conditional logistic regression in order to investigate genetic variants or circulating biomarker in relation to prostate cancer risk. These calculations were performed using SAS.

Analysis of genetic variation in relation to circulating biomarkers (Papers II and III)

We used standard linear regression models in order to assess the relation between genetic variants and levels of biomarkers, such as hormone levels of IGF1. These calculations were performed using SAS.
Permutation testing (Paper I)
In order to account for the large amount of statistical tests without prior hypotheses regarding the individual haplotypes carried out in Paper I, we performed permutation testing in order to attain adjusted p-values. New data sets under the null hypothesis were created by permuting the case-control status at least 1000 times. For each new data set (permutation), the association between each haplotype and risk was calculated using unconditional logistic regression adjusting for age-category and region. The smallest p-value for each permutation was thereafter retained, creating a vector of empirical p-values under the null hypothesis. Adjusted p-values were then achieved by comparing the unadjusted p-values with the corresponding position of the sorted empirical p-value vector. These calculations were performed using R.

Meta analysis (Paper II)
Combined estimates of hormone levels in relation to genetic variants were calculated as weighted means, with study specific weights calculated as the inverse of the variance. In SNP analysis, we calculated the within study mean differences in hormone levels between wild type homozygotes and heterozygotes, and between wild type homozygotes and rare type homozygotes, respectively. The estimated differences were then used to calculate the combined genotype specific effect. To assess global significance, study specific beta coefficients (trend estimates) with corresponding confidence intervals were estimated based on the genotype specific level differences. The beta coefficients were then included in a separate meta-analysis as described above. To investigate heterogeneity between studies Cochran’s Q tests were performed. We used the random effects model when heterogeneity was statistically significant, otherwise the fixed effects model. These calculations were performed using the StatsDirect.

Survival analysis (Paper III)
In order to investigate if genetic variants or biomarkers were related to prostate cancer-specific survival probability, we used Cox proportional hazard regression models. The proportional hazard assumptions were tested using Schoenfeld residuals. These calculations were performed using SAS.
RESULTS

Paper I

*Title: Comprehensive evaluation of genetic variation in the IGF1 gene and risk of prostate cancer*¹²⁹

In this study, we aimed to investigate if genetic variation in the *IGF1* gene affects prostate cancer risk. In total, the study included 2,863 cases and 1,737 controls from the CAPS study.

The *IGF1* gene is located on chromosome 12 and stretches approximately 85 kb. Using HapMap phase I data, the gene displayed three distinct haplotype blocks, see Figure 14. For each block, three SNPs were required in order to explain all common haplotypes with an $R_h^2 > 0.8$¹²³.
Figure 14. LD structure of the IGF1 gene. Also showing are the htSNPs and common haplotypes of block 3. Figure adapted from Johansson et al. (Paper II in thesis).  

Genetic variations in blocks 1 and 2 were not associated with overall prostate cancer risk in SNP- or haplotype analyses. In block 3 the rare allele of two SNPs were associated with a borderline increased risk of prostate cancer, see Table 1. When stratifying for cases with high risk disease, defined as local tumour stage T3 or T4, lymph node metastasis, bone metastasis, or serum PSA levels above 50 ng/mL, the rs6220 SNP was also significantly associated with increased risk, whereas the rs2033178 SNP was no longer associated with risk, see Table 1.

<table>
<thead>
<tr>
<th>SNP</th>
<th>All cases</th>
<th>High risk cases</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cases</td>
<td>controls</td>
</tr>
<tr>
<td>rs2033178</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homozygous major</td>
<td>2293</td>
<td>1509</td>
</tr>
<tr>
<td>Heterozygous</td>
<td>278</td>
<td>131</td>
</tr>
<tr>
<td>Homozygous minor</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>rs6220</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homozygous major</td>
<td>1315</td>
<td>881</td>
</tr>
<tr>
<td>Heterozygous</td>
<td>1062</td>
<td>615</td>
</tr>
<tr>
<td>Homozygous minor</td>
<td>209</td>
<td>136</td>
</tr>
<tr>
<td>rs7136446</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homozygous major</td>
<td>1034</td>
<td>727</td>
</tr>
<tr>
<td>Heterozygous</td>
<td>1150</td>
<td>670</td>
</tr>
<tr>
<td>Homozygous minor</td>
<td>332</td>
<td>186</td>
</tr>
</tbody>
</table>

1) Odds ratios for prostate cancer risk calculated by conditional logistic regression, conditioning on age category and recidancy

In haplotype analysis, the heterozygote carriers of the rare TCC haplotype in block 3 had an increased prostate cancer risk, see Table 2. The global p-value for block 3 was 0.004 and the haplotype specific p-value for the TCC haplotype was 0.002. When adjusting for multiple testing by permutation testing, both the global p-value for block 3, and the specific p-value for the TCC haplotype were still significant (both p=0.02).

In conclusion, we noted that the rare allele of several SNPs within the region of block 3 had previously been associated with elevated circulating levels of IGF1. Because elevated levels of IGF1 in the circulation have previously been associated with increased prostate cancer risk, we hypothesized that the observed risk increase associated with the TCC haplotype in block 3 was mediated by circulating IGF1.
levels.

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Control Frequency</th>
<th>Case Frequency</th>
<th>OR (95% CI)</th>
<th>p</th>
<th>Global p</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTT (2 copies)</td>
<td>0.38</td>
<td>0.35</td>
<td>1.00 (reference)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTT (1 copy)</td>
<td>0.46</td>
<td>0.48</td>
<td>0.90 (0.77-1.06)</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>CCC (1 copy)</td>
<td>0.27</td>
<td>0.3</td>
<td>1.19 (1.02-1.4)</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>CCC (2 copies)</td>
<td>0.04</td>
<td>0.03</td>
<td>0.78 (0.55-1.11)</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>CCT (1 copy)</td>
<td>0.2</td>
<td>0.22</td>
<td>1.19 (0.99-1.43)</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>CCT (2 copies)</td>
<td>0.01</td>
<td>0.02</td>
<td>1.18 (0.68-2.05)</td>
<td>0.56</td>
<td>0.004</td>
</tr>
<tr>
<td>CTC (1 copy)</td>
<td>0.09</td>
<td>0.1</td>
<td>1.07 (0.83-1.38)</td>
<td>0.59</td>
<td></td>
</tr>
<tr>
<td>CTC (2 copies)</td>
<td>0.01</td>
<td>0</td>
<td>0.30 (0.10-0.94)</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>TCC (1 copy)</td>
<td>0.08</td>
<td>0.1</td>
<td>1.46 (1.15-1.84)</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>TCC (2 copies)</td>
<td>0</td>
<td>0</td>
<td>0.82 (0.25-2.71)</td>
<td>0.74</td>
<td></td>
</tr>
</tbody>
</table>

1) Dosage estimates from the "tagSNPs" software 2) Risk estimates were assessed by performing conditional logistic regression, conditioning on age and region 3) Global p-value for entire block assessed with likelihood ratio test
Paper II

Title: Implications for Prostate Cancer of IGF1 Genetic Variation and Circulating IGF1 Levels

In paper II, we further investigated the hypothesis postulated in paper I; that genetic variation in the 3’ region of the IGF1 gene affects circulating levels of IGF1 and therefore prostate cancer risk. In total this study included 698 controls from the CAPS study and 575 cases and controls from the prospective NSHDC study.

Plasma levels of IGF1 were analysed in 698 control subjects from the CAPS study. To increase statistical power, we also genotyped the three htSNPs from block 3 of the IGF1 gene (Figure 14), rs2033178, rs6220 and rs713646 in 575 cases and controls from the NSHDC study in which circulating IGF1 had already been measured. An additional SNP, rs2946834 located downstream of block 3, was also included because it had previously been associated with IGF1 levels in another study. In addition, we performed a meta-analysis of these two, and three other association studies in which genetic variation in the 3’ region of the IGF1 gene as well as circulating levels of IGF1 had been measured. Al Zahrani et al. analysed 9 tagging SNPs in both men and women (n=600) within the MRC Ely study and found several SNPs associated with IGF1 levels in the 3’ region of the IGF1 gene. Canzian et al. analysed 5 SNPs selected because of potentially functional roles within a breast cancer case-control study nested within the European Prospective Investigation into Cancer and Nutrition (EPIC) and found one SNP associated with elevated levels of IGF1 within the 3’ region of the IGF1 gene. Verheus et al. analysed 18 haplotype tagging SNPs in relation to breast cancer density, as well as circulating levels of IGF1 in 656 women participating in Prospect-EPIC, and found five SNPs in the 3’ region of the gene, as well as the haplotype corresponding to the TCC haplotype of Paper I in this thesis, associated with elevated IGF1 levels. We also acquired genotype data from Verheus et al. in order to recreate the TCC haplotype and to analyze it using the same regression model as in paper II of this thesis (see Material and methods, Statistical methods).

Heterozygote carriers of the TCC haplotype, that had increased prostate cancer risk in paper I, also had a moderate increase in plasma IGF1
levels in the CAPS controls ($p=0.02$), but not in NSHDC ($p_{\text{trend}}=0.12$). In contrast, the rare alleles of two of the block 3 htSNPs rs6220 and rs7136446 were associated with elevated IGF1 levels in NSHDC ($p_{\text{trend}}=0.03$ and 0.04, respectively), but not in CAPS. The rs2946834 SNP was also associated with IGF1 in both CAPS and NSHDC (both with $p_{\text{trend}}=0.02$).

In the haplotype meta-analysis, including CAPS, NSHDC and Prospect-EPIC, the TCC haplotype was significantly associated with elevated levels of circulating IGF1 ($p=0.001$). In the SNP meta-analysis, the rare allele of all included SNPs that had been analysed in more than one study in relation to circulating IGF1 levels ($n=5$) were significantly associated with elevated levels of IGF1 ($p_{\text{trend}}$ range: 0.04 - < 0.0001), see Figure 15.

In conclusion, paper II provided further evidence that rare alleles in the 3’ region of the IGF1 gene are positively associated with circulating IGF1 levels. Because circulating IGF1 is positively associated with prostate cancer risk, paper II also supported the findings from Paper I, in which the 3’ haplotype carrying the rare alleles on all loci was associated with increased prostate cancer risk.

---

**Figure 15. Meta-analysis graph of the rs6220 SNP in relation to circulating IGF1.**

In conclusion, paper II provided further evidence that rare alleles in the 3’ region of the IGF1 gene are positively associated with circulating IGF1 levels. Because circulating IGF1 is positively associated with prostate cancer risk, paper II also supported the findings from Paper I, in which the 3’ haplotype carrying the rare alleles on all loci was associated with increased prostate cancer risk.
Paper III

Title: Genetic and plasma variation of insulin-like growth factor binding proteins in relation to prostate cancer incidence and survival

In paper III we set out to investigate genetic variation in the IGFBP1, IGFBP3 and IGFALS genes in relation to prostate cancer incidence and survival. In addition, plasma levels of total- and intact IGFBP3 were measured and analysed in relation to genetic variation and prostate cancer-specific survival. In total, this study included 2,774 cases and 1,736 controls available for genotyping, and 1,521 cases and 909 controls available for plasma analysis.

In accordance with paper I, haplotype blocks and htSNPs were defined based on LD information acquired using genotype data from HapMap, see Figures 1 and 2 of paper III. We selected in total 5 htSNPs in the IGFBP1 gene, 5 SNPs (3 htSNPs) in the IGFBP3 gene, and 3 htSNPs in the IGFALS gene for genotyping. Circulating levels of total- and intact IGFBP3 were analysed in all available plasma samples.

Overall, considering the number of statistical tests carried out, no clear associations were observed between genetic variants and prostate cancer incidence or survival. The IGFBP3 SNP rs2854744 was associated with plasma IGFBP3 levels ($p_{\text{trend}}=9*10^{-8}$). Other associations between SNPs and hormone levels were of weak statistical strength ($p>0.01$). In haplotype analysis in relation to hormone levels, several IGFBP3 haplotypes were associated with decreased IGFBP3 levels. However, these associations were attributed to the rare allele of the rs2854744 SNP which was located on the most common haplotype (GTG), i.e. the reference haplotype.

In survival analysis of hormone levels, both total- and intact IGFBP3 plasma levels were significantly associated with increased risk of prostate cancer-specific death, with p-trend values of 0.03 and $6*10^{-14}$, respectively. When adjusting for clinical characteristics and primary treatment, the association for IGFBP3 levels disappeared ($p_{\text{trend}}=0.88$ and 0.72, respectively). The association for intact IGFBP3 was attenuated, albeit still statistically significant with a p-trend value of 0.0001 when adjusting for clinical characteristics, and a p-trend value of 0.0004 when further adjusting for primary treatment. An unadjusted
Kaplan-Meier plot for quintiles of intact IGFBP3 is shown in Figure 16. In order to investigate the effect of primary treatment on hormone levels, we performed linear regression including dummy variables for all treatment categories as well as clinical characteristics as covariates. In this analysis, the palliative treatments chemical (GnRH analogues) and surgical castration were related to elevated levels of intact IGFBP3 ($p_{trend}=0.01$ and $1 \times 10^{-11}$, respectively). Stratified analysis for cases that were castrated (chemically or surgically) was therefore performed and the positive association between levels intact IGFBP3 levels and prostate cancer-specific death was still significant ($p_{trend}=0.0003$). Furthermore, in cases that had not been given any treatment, intact IGFBP3 levels was also associated with increased risk of prostate cancer death ($p_{trend}=0.02$).

Figure 16. Unadjusted Kaplan-Meier plot showing survival probabilities for quintiles of intact IGFBP3.

In conclusion, genetic variation in the \textit{IGFBP3} gene affects circulating levels of IGFBP3 as reported by previous studies.\textsuperscript{84-87} Circulating levels of intact IGFBP3 are positively associated with prostate cancer-specific death, but further studies are warranted to investigate if this observation reflects a causal relationship, or of the association is confounded by some other factor.
Paper IV

Title: Circulating concentrations of folate and vitamin B12 in relation to prostate cancer risk – results from the EPIC study

In paper IV we investigated if circulating levels of folate and vitamin B12 were related to prostate cancer risk as suggested by our previous study within the NSHDC study. Circulating levels of folate and vitamin B12 were analysed in 869 cases and 1174 individually matched controls, nested within the European Prospective Investigation into nutrition and Cancer (EPIC) cohort. The study included prostate cancer cases from seven of the ten participating EPIC countries: Germany, Greece, Italy, the Netherlands, Spain, Sweden (newly diagnosed cases from the NSHDC) and the United Kingdom (UK).

Circulating folate and vitamin B12 were not associated with overall prostate cancer risk ($p_{\text{trend}}=0.62$ and 0.21, respectively). Notable is that the subgroup of subjects from the NSHDC study, i.e. the same cohort in which the association between vitamin B12 and risk was originally found, displayed a null association between vitamin B12 and prostate cancer risk. The only subgroup analysis displaying a significant association was for vitamin B12 in relation to advanced stage prostate cancer cases with an overall relative risk for a doubling in levels of 1.69 (95 % CI: 1.05 – 2.72, $p_{\text{trend}}=0.03$). A potentially important observation was that folate levels varied substantially across the participating countries; Swedish participants displaying low levels, in particular compared to German and UK participants that had almost three-fold the median folate levels of the Swedish participants.

We concluded that Paper IV did not provide strong support for the role of folate and vitamin B12 in prostate cancer aetiology. Further investigations are warranted to elucidate if elevated levels of circulating vitamin B12 increases the risk of advanced stage prostate cancer.
Prostate Cancer Aetiology – Results

Paper V

Title: Prospective investigation of the one-carbon metabolism pathway in relation to prostate cancer risk – results from the NSHDC study

In this study we investigated if additional factors of the one-carbon metabolism pathway are related to prostate cancer susceptibility. Plasma levels of folate, vitamin B12, homocysteine, choline, betaine, methionine, vitamin B2, vitamin B6, cysteine, and methylmalonic acid (MMA) were analysed in 561 cases and 1034 individually matched controls, nested within the Northern Sweden Health and Disease Cohort (NSHDC). In order to further scrutinize the discrepant associations between vitamin B12 and prostate cancer risk noted in previous studies, we also divided the study population into its two study periods\textsuperscript{96,132}. The first period included cases diagnosed between 1994 and 2002 (5\textsuperscript{th} – 95\textsuperscript{th} percentile) that were included in the first study from NSHDC by Hultdin et al\textsuperscript{96}. These cases, together with their matched controls, were labelled Wave 1. The second study period included cases diagnosed between 2001 and 2005 that were included in the EPIC study (Paper IV in thesis)\textsuperscript{132}. These cases, together with their matched controls, were labelled Wave 2.

Baseline Pearson correlation coefficients for all investigated B-vitamins and metabolites are shown in Table 3.

<table>
<thead>
<tr>
<th>Homocysteine</th>
<th>Methionine</th>
<th>MMA</th>
<th>Choline</th>
<th>Betaine</th>
<th>Folate</th>
<th>Vitamin B12</th>
<th>Vitamin B6</th>
<th>Vitamin B2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cysteine</td>
<td>0.12***</td>
<td>-0.05*</td>
<td>0.22***</td>
<td>0.07*</td>
<td>0.11***</td>
<td>-0.06*</td>
<td>0.06*</td>
<td>0.05*</td>
</tr>
<tr>
<td>Methionine</td>
<td>-0.05*</td>
<td>0.15***</td>
<td>0.1***</td>
<td>-0.06*</td>
<td>0</td>
<td>0.06*</td>
<td>-0.01</td>
<td></td>
</tr>
<tr>
<td>MMA</td>
<td>0.02</td>
<td>0</td>
<td>0.01</td>
<td>-0.14***</td>
<td>-0.02</td>
<td>-0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Choline</td>
<td>0.41***</td>
<td>0.11***</td>
<td>-0.02</td>
<td>0.01</td>
<td>0.07*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Betaine</td>
<td>0.24***</td>
<td>-0.01</td>
<td>0.07*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Folate</td>
<td>0.08*</td>
<td>0.2***</td>
<td>0.08*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vitamin B12</td>
<td>0.15***</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vitamin B6</td>
<td>0.06*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* indicates \(P<0.05\), *** indicates \(P<0.0001\)

We observed positive associations for plasma levels of choline (\(p_{\text{trend}}=0.03\)), vitamin B2 (\(p_{\text{trend}}=0.05\)) and vitamin B12 (\(p_{\text{trend}}=0.01\)) in relation to prostate cancer risk, whereas negative associations were observed for plasma levels of homocysteine (\(p_{\text{trend}}=0.02\)) and MMA (\(p_{\text{trend}}=0.03\)). Stratifying the study population into its two study waves revealed a number of differences in case characteristics, see Table 3 in paper V. Comparing cases of wave 2 with cases of wave 1, wave 2 cases had lower stage tumours (\(p=0.01\)), lower PSA levels at blood draw
they were older at diagnosis ($p=0.0006$), they had a lower fraction of malign tissue in biopsies ($p=0.01$), and they had longer time between blood draw and diagnosis ($p<0.0001$). These observations would be consistent with a higher proportion of cases in wave 2 being diagnosed after a health check-up (opportunistic PSA testing) than cases in wave 1 ($p<0.0001$).

The overall associations for vitamin B12, homocysteine and MMA were clearly attributed to subjects of wave 1, see Figure 17. The only analytes displaying overall significant associations with risk and similar relative risks between the study waves were choline and vitamin B2. Stratifying the vitamin B12 analysis for any discrepant baseline characteristics did not explain the inconsistent results between the study waves. Furthermore, smoking status modified the relative risk associated with elevated levels of vitamin B12 ($p_{interaction}=0.03$), see Figure 18. In stratified vitamin B12 analysis by smoking status, non-smokers of both study waves had similar and positive association between vitamin B12 and risk. However, in ever smokers, vitamin B12 was negatively associated with risk in wave 2, but not in wave 1, and in current smoker, vitamin B12 was not associated with risk in any of the study waves.

In conclusion, we observed novel and positive associations for plasma levels of choline and vitamin B2 in relation to prostate cancer risk. Smoking status modified the association between vitamin B12 and risk where positive associations were observed for both study waves in non smokers, whilst ever/current smokers displayed ambiguous associations. Differences in baseline or case characteristics did not explain the conflicting results between vitamin B12 and overall risk as observed in previous studies from the NSHDC. Overall, paper V provided some support for the role of one-carbon metabolism in prostate cancer aetiology.
Figure 17. Forest plot of relative risks associated with a doubling in plasma levels of all analysed B-vitamins and metabolites, stratified for study wave.
Figure 18. Forest plots of prostate cancer RRs associated with a doubling in vitamin B12 levels, stratified by wave and smoking status.
DISCUSSION

The IGF pathway and prostate cancer

The IGF-pathway, and in particular IGF1, has been one of the most frequently implicated factor in prostate cancer aetiology. The evidence underlying this is based on results from cell lines and mouse models, as well as prospective and retrospective case-control studies\(^7^4\). However, when reviewing the prospective case-control studies it becomes clear that the evidence for this relationship is overall weak\(^7^3,7^4,1^1^2\). Recently, the Endogenous Hormones and Prostate Cancer Collaborative Group have again investigated this relation in a pooled analysis of twelve cohorts, including in total 3300 prospective cases and 4450 controls (Roddam et al. submitted). In this study, high circulating IGF1 levels were associated with a modestly increased risk of prostate cancer, a result of high statistical significance (\(p_{\text{trend}}=4\times10^{-7}\)). One observation from Roddam et al. which raises doubt on the causal role of IGF1 is that this association seemed to be located in cases with short lag-time, i.e. short time between blood draw and diagnosis. Given that prostate cancer is a slowly developing disease with neoplastic transformations likely to occur several years before diagnosis, early prostate malignancies may affect the circulating pool of IGF1, thereby reversely causing the association between circulating IGF1 and prostate cancer risk.

Investigating levels of biomarkers in a prospective study in relation to disease risk is statistically a very powerful approach, because continuous variables (e.g. plasma levels) bring more statistical power than discrete
variables (genetic markers). However, reversed causation and confounding may cause spurious associations in case-control studies\textsuperscript{100}. Using data on polymorphic genes known to be related to plasma levels and possibly disease risk may serve as an important complement to biomarker studies. According to Mendel’s second law, carriage of a genetic trait, such as a polymorphism, is subject to random assortment of maternal and paternal alleles at the time of gamete formation. Associations between a germline genetic variant and risk should therefore not be subject to reversed causation or confounding\textsuperscript{100}. Therefore, genetic- and biomarker association studies can be used in conjunction to strengthen the evidence of causality for a risk factor.

With this background, it is interesting that we found associations between specific genetic variants in the \textit{IGF1} gene and prostate cancer risk, as well as to circulating levels of IGF1, as reported in papers I and II\textsuperscript{129,130}. It should be pointed out that the CAPS control population might be prone to bias when investigating alternative hypotheses apart from disease risk, because the controls were frequency matched to cases by age, thus not being truly population based\textsuperscript{133}. However, given the cumulative evidence gained from the meta-analysis including studies of various designs, the association between \textit{IGF1} 3’ variation and IGF1 levels seems convincing. Along the lines of Mendelian randomization, these observations support the causal role for circulating IGF1 in prostate cancer aetiology, see Figure 19.

As illustrated in figure 19 with arrows between the genetic variant and disease pointing one-way, an inherited genetic variant cannot be affected by the disease itself. However, one reservation of the causal inference of IGF1 in prostate cancer from the results of papers I and II, would be the possibility that the genetic variant in the \textit{IGF1} gene affects risk through some other unknown mechanism in addition to its affect on circulating IGF1, a phenomenon referred to as pleiotropy\textsuperscript{134}. It should also be noted that the statistical significance of the association between the \textit{IGF1} haplotype and risk, although significant after adjustments for multiple testing, was not strong enough to convey conclusive evidence for the association. Thus, additional studies are needed to confirm the finding from Paper I.
In paper III we investigated genetic variations in the major IGF-binding proteins, *IGFBP1* and *IGFBP3* as well as *IGFALS*. In parallel with paper I, genetic variation was assessed by haplotype tagging and plasma levels of total- and intact IGFBP3 were measured in plasma samples from both cases and controls. A few nominal associations were noted between genetic variants, prostate cancer risk, survival and hormone levels, but given the large number of tests performed in this study, these findings were most likely due to chance. The one exception was the strong association between the *IGFBP3* SNP rs2854744 and circulating levels of IGFBP3, as has been reported by several other studies \(^{83,85,87,135}\). In paper III of this thesis, the minor homozygote carriers of the rs2854744 SNP had a 16 % increase in total IGFBP3 levels compared to the major homozygotes. Therefore, if circulating total IGFBP3 had a major impact on prostate cancer susceptibility, the rs2854744 SNP would also be expected to be associated with risk. It is also interesting to note that the rs2854744 SNP was not associated with levels of intact IGFBP3.
IGFBP3 exists in two major forms in the circulation; the intact part 43-45 kDa which binds to IGF1, and the inactive proteolyzed 30 kDa part\textsuperscript{136}. If only intact IGFBP3 is important in the regulation of IGF1 in the circulation, this would explain the lack of association between the rs2854744 SNP and risk. Nevertheless, it is intriguing that a germline polymorphism can exert such a strong effect on total IGFBP3, but not on intact IGFBP3.

IGFBP3 has consistently been shown to inhibit proliferation and induce apoptosis in prostate cancer cell lines and animal models independently of IGF1\textsuperscript{77}. Therefore, it was surprising that elevated levels of intact IGFBP3 were strongly associated with increased risk of prostate cancer-specific death in paper III of this thesis. This association was substantially attenuated, but still statistically significant, when adjusting for clinical characteristics and primary treatment. Hormone treatment and castration have previously been found to increase expression of IGFBP3, and we observed that cases who were either chemically (treated with GnRH analogues) or surgically castrated, had substantially elevated levels of intact IGFBP3\textsuperscript{137,138}. Because prostate cancer patients that are castrated usually have an advanced disease with poor prognosis, this observation suggests that the association between intact IGFBP3 and survival might be reversely caused by the treatment. Conflicting with this assumption, in stratified analysis of cases that were castrated, the association between intact IGFBP3 and survival was still significant. Furthermore, the association was also nominally significant in cases that had not been treated. However, we only had data on categorized primary treatment, thus leaving the possibility of residual confounding in adjusted analysis. Overall, given the pro-apoptotic effect of IGFBP3 on malign tissue locally in the prostate, it seems unlikely that circulating intact IGFBP3 would causally increase the risk of prostate cancer-specific death\textsuperscript{77}. Instead, speculatively, in cases with advanced disease, expression of IGFBP3 might be up-regulated, thus reversely causing the observed increase in circulating intact IGFBP3.

Overall, the strong association noted between intact IGFBP3 and prostate cancer survival underlines the importance of the IGF pathway in prostate cancer epidemiology. The result motivates further studies of the biological implication of intact IGFBP3 in other study designs, including prospective- and experimental studies.
The one-carbon metabolism pathway and prostate cancer

Factors of one-carbon metabolism, in particular folate, have been regarded as protective against some cancers by results from prospective studies\(^9^4\). Recently, however, it has been recognized that this relation may be more complex than initially anticipated. In the Aspirin-Folate Polyp Prevention Trial, a significantly increased number of recurrent adenomas, as well as an increased risk of prostate cancer, were reported in the folate supplementation arm of the study\(^8^9\). A dual relationship between folate and cancer development has been suggested, in which high folate levels may protect against cancer in healthy cells but promote cancer development once neoplastic transformations have occurred.

In a study from the NSHDC cohort, Hultdin \textit{et al.} reported that elevated levels of vitamin B12 were strongly associated with increased prostate cancer risk\(^9^6\). In an attempt to replicate this finding, we pooled newly diagnosed cases and matched controls from the NSHDC study with the EPIC study but found no association between folate, vitamin B12 and overall prostate cancer risk (Paper IV). The only significant association found in Paper IV was a positive association between vitamin B12 and advanced stage prostate cancer. Surprisingly, the subjects from the NSHDC study displayed a NULL association between vitamin B12 and overall prostate cancer risk, whereas the other sub-cohorts displayed a positive association when excluding NSHDC (data not shown in Paper IV). One particular feature of Paper IV was that folate levels varied substantially across the participating sub-cohorts; German and UK participants displaying almost three-fold the median folate levels of the Swedish participants (NSHDC). Because folate is a crucial factor of the one-carbon metabolism pathway, we also made a stratified analysis of vitamin B12 in the EPIC study, excluding subjects with folate levels above 10 ng/mL and all subjects from the NSHDC cohort. This analysis – which was not reported in Paper IV – yielded a strong positive association between vitamin B12 and prostate cancer risk, suggesting that the risk increase associated with vitamin B12 in the study from Huldin \textit{et al.} only applies to populations of low folate status\(^9^6,1^3^2\).

In order to further investigate factors of one-carbon metabolism in relation to prostate cancer we analysed eight additional B-vitamins and
related metabolites within the NSHDC study in Paper V of this thesis. We also aimed to elucidate the discrepant results on the association between vitamin B12 and prostate cancer risk by performing a comprehensive comparison between the participants included in the two previous studies from NSHDC. The first study, referred to as wave 1 in Paper V, included prostate cancer cases diagnosed before 2003. The second study, referred to as wave 2 in Paper V, included cases diagnosed after 2002. Most of the analytes studied in Paper V had not been investigated in relation to prostate cancer risk in a prospective setting before.

We observed positive associations between choline, vitamin B2 and B12 and overall prostate cancer risk, whereas homocysteine and MMA were inversely associated with overall prostate cancer risk. Levels of cysteine, methionine, total methionine, betaine and vitamin B6 were not associated with overall prostate cancer risk. When comparing the two study waves it was apparent that the positive association between vitamin B12 and risk, as well as the inverse associations for homocysteine and MMA, were attributed to subjects from wave 1 (Figure 17). Given the NULL association for vitamin B12 in wave 2, the similar behaviour of homocysteine and MMA is not surprising since they are both negatively correlated with vitamin B12 (Table 3). The only analytes displaying overall significant associations with risk, as well as similar relative risk estimates in both study waves, were choline and vitamin B2.

One way in which one-carbon metabolism might affect prostate cancer development is by inducing hypermethylation of important genes, such as the Glutathione S-transferase π (GSTP1) gene (see Background, Candidate pathways). Factors within the one-carbon metabolism pathway that increases the methyl-group availability may therefore increase prostate cancer risk. Choline is a methyl donor particularly important when the folate status is low. Choline might therefore be a strong methylation determinant in populations such as the NSHDC population in which the folate status is low. Choline donates methyl groups via betaine, which also displayed a similar, although not statistically significant, association with risk. This observation would support the hypermethylation hypothesis in prostate cancer aetiology. However, betaine donates its methylation groups in the remethylation of homocysteine to methionine, and methionine displayed a null association
with risk. The latter observation does not support the hypermethylation hypothesis. In parallel with choline, however, vitamin B2 may also increase methyl group availability by acting as a cofactor in the MTHFR reaction, see Figure 10.

One prior hypothesis in Paper V was that differences in tumour characteristics might explain the inconsistent results of the risk associations for vitamin B12 between the study waves. Indeed, several discrepancies between the study waves in tumour characteristics were observed, wave 2 cases generally showing characteristics related to less aggressive disease (Table 3 in paper V). However, performing analyses between vitamin B12 and risk stratified on tumour characteristics did not explain the discrepant results. Furthermore, in Table 4 in paper V showing baseline levels of all investigated B-vitamins and metabolites stratified on case-control status and study wave, the case groups of both waves had similar levels of vitamin B12, whilst controls in wave 1 had significantly lower levels of vitamin B12 than wave 2 controls. Thus, it seems as if the explanation for the discrepant vitamin B12 association might be due to differences in the control group, rather than in the case group, but differences in baseline characteristics between the study waves in controls were minor and stratified analyses did not shed any further light on the relation between vitamin B12 and prostate cancer risk.

Interestingly, smoking appeared to modify the risk association for vitamin B12, and notable was that vitamin B12 was positively associated with risk in non-smokers of both study waves. The association for vitamin B12 was less clear in ever and current smokers, and wave 2 ever smokers displayed a negative association. Testing the overall relative risk trends for interaction by smoking status yielded a significant result. The biological explanation for this effect modification is not clear, but one finding that would support the observed interaction is that smoking has been reported to transform vitamin B12 co-enzymes into biologically inactive compounds\textsuperscript{142}. If smoking reduces the activity of vitamin B12, this would also explain the null association between vitamin B12 and prostate cancer risk in the study by Weinstein \textit{et al.} which was conducted within the Finnish ATBC study, consisting solely of heavy smokers\textsuperscript{97}. However, the inverse association between vitamin B12 and risk in ever smokers, as reported in Paper V, is puzzling and complicates the
interpretation.

In conclusion, we observed associations between choline, vitamin B2 and overall prostate cancer risk, but these associations were statistically weak and should be further evaluated in other large prospective studies. Folate status seems crucial when investigating these relationships, and stratified analysis on subjects with low folate status is therefore warranted. The risk increase associated with elevated levels of vitamin B12 seems to be attributed to non-smokers. Overall, papers IV and V indicate that factors of one-carbon metabolism affect the risk of developing prostate cancer but the result is not conclusive.
Study design – strengths and limitations

Genetic association studies (Papers I – III)
Prostate cancer is regarded as a typical complex disease, in which many environmental and genetic variants contribute and interact in the development of the disease. To date, most case-control studies investigating genetic variation in relation to prostate cancer risk have been candidate gene studies. These studies often analyze genes that have been implicated in prostate cancer aetiology by results from experimental studies or studies investigating biomarkers in relation to risk. Until a few years ago, candidate gene studies mostly analysed single functional SNPs, a strategy not overly successful. Over the last couple of years, however, more studies are utilizing different tagging approaches trying to find associations with the use of genetic markers serving as proxies for potentially unknown causal variants (see Background, Genetics).

Tagging methods can broadly be divided in pair-wise- and haplotype tagging. Several attempts have been made to compare the power and efficiency of different tagging methods using both HapMap and simulated data. These studies have not provided a universal answer for which tagging approach is the most efficient, and the answer is most likely dependant on the LD structure of the particular genomic region of interest. In order to choose the best tagging method for a specific study, it would be useful to have an algorithm in which the power for different tagging methods can be simulated based on the LD structure of the genomic region of interest. The haplotype tagging approach requires definition of haplotype blocks which is not needed using the pair-wise tagging approach. Thus, in regions of high LD with few common haplotypes, one can argue for the haplotype tagging approach as the most efficient, whereas the pair-wise method might be more convenient in regions of low LD. Furthermore, the interpretation of associations found by haplotype tagging might be more difficult since haplotype blocks can stretch over large regions, thus making pinpointing of the causal variant hard. Associations found by pair-wise tagging, on the other hand, are limited to smaller genomic regions and should be more easily interpreted. Overall, haplotype tagging require more hands on work in order to gain in efficiency compared to pair-wise tagging which is easier to automate, a nice feature in large scale projects where efficient hands on definition of haplotype blocks might not be feasible.
Throughout papers I – III of this thesis, we used the haplotype tagging approach in order to investigate common genetic variation within the genes of interest. In papers I and II, the *IGF1* gene displayed three distinct regions of high LD in which hSNPs were readily selected, see Figure 14. In paper III, the *IGFBP1* and *IGFALS* genes displayed high LD across the genes, whereas the *IGFBP3* gene displayed a more heterogeneous LD pattern with a region of low LD where we had to select additional SNPs to get adequate coverage, see Figures 1 and 2 of paper III. Thus, it might have been more convenient to use the pair-wise tagging approach in paper III.

By results from the genome-wide association studies (GWAS) on prostate cancer, it is becoming clear that epidemiologists have been consistently poor in predicting susceptibility genes and polymorphisms. In fact, most findings from GWAS have not been located in previously implicated candidate genes, and many findings have even been in gene deserts\textsuperscript{25,68-72}. Over the last two years, GWAS have by far surpassed all previous candidate gene studies in identifying susceptibility loci for prostate cancer.

**CAPS (Papers I-III)**

The CAPS study, with nearly 3,000 cases and 2,000 controls, is one of the largest case-control studies of prostate cancer today. The statistical power in CAPS to detect a true association at a p-value of 0.05 with odds ratios of 1.2 for heterozygotes and 1.44 for homozygotes using the trend test is shown in Figure 20. At a minor allele frequency of 12 %, CAPS have a power of 80 % to detect such a true association. It should be noted that this power estimation assumes no genotyping error or case-control misclassification.

In the CAPS study, diagnosed prostate cancer cases were invited and controls were thereafter frequency matched to cases by age and geographical region. Considering potentially confounding factors, age would not be expected to confound the results in papers I – III where we investigated genetic variants because the genotype distribution is not likely to differ much between different age groups. In addition, conditioning the analyses on the 5-year age groups should effectively account for any confounding by age. A more serious problem in CAPS, and in genetic association studies overall, is population stratification.
Given the overall participation rates in cases (81 %) and controls (58 %), population stratification and/or selection bias might be a problem in CAPS. Sub-populations with different genetic backgrounds and different willingness to participate may exist within the study population, and conditioning on the two matching regions would not effectively adjust for this. However, the homogeneous nature of the Swedish population and the register based recruitment should minimize this problem\textsuperscript{145}.

**EPIC and NSHDC (Papers IV and V)**

Both the EPIC- and NSHDC studies are of prospective design, i.e. cases were recruited prior to their prostate cancer diagnosis. In retrospective case-control studies, reversed causality and confounding by treatment or diagnosis might pose serious problems. In prospective case-control studies, these issues should be less of a problem. In addition, controls were individually matched to each case by age and date of recruitment, and in all biochemical analyses cases and controls of the same stratum were analysed in the same batch. This design together with the statistical analysis by conditional logistic regression effectively controls for any systematic differences in recruitment or biochemical analysis.

EPIC is the largest single prospective cohort worldwide with in total 500,000 participants of which almost 400,000 donated a blood sample. Comparing the EPIC study with the NSHDC study, the most apparent
advantage is its larger study population. However, the EPIC study includes ten countries across Europe. In some countries follow-up and identification of diagnosed prostate cancer cases were based on cancer registers. In other countries, however, follow-up was based on other methods, such as health insurance records, cancer and pathology registries, and active follow-up through study subjects and their next-of-kin. Speculatively, these follow-up discrepancies, and other differences in clinical diagnostic practice, might cause a larger degree of phenotypic heterogeneity than in other, more homogeneous studies. Some of the sub-cohorts provided sparse information on case characteristics. For instance, tumour grade was only assessed as low/intermediate/high in some cohorts, whereas others included more detailed data by Gleason score.

Furthermore, as indicated in paper IV, large differences in folate levels were observed across the study population. This observation underlines the large lifestyle differences inherent in this type of multi-centre study. This may give the opportunity to investigate a broader range of exposure values, but it may also pose problems if some unknown life-style related factor interacts with the exposure of interest. In other words, an association found in one sub-population might not be identified in another because of different exposure to unknown interacting factors.

The major advantages of the NSHDC study include the register-based follow-up by linkage to the Northern part of the National cancer register which includes 100 % of all diagnosed prostate cancer cases, the detailed tumour characteristics, and the homogeneous study base of the Västerbotten county. In particular the register-based follow-up provides a swift and inexpensive way to achieve complete identification of all incident prostate cancer cases within the cohort, a great advantage compared to prospective studies in countries without cancer registers. Furthermore, Weinehall et al. studied differences in social characteristics between NSHDC participants and non-participants of the Västerbotten and concluded that selection bias was marginal overall.

While the homogeneous nature of the NSHDC cohort provides unique opportunities to study some exposures in a specific setting, e.g. the association between vitamin B12 and risk in a population of low folate status, it also means that we have to question the external validity of results found within the cohort. The associations found within NSHDC
may not be applicable to other cohorts. In addition, because many exposures of interest, such as vitamin levels, vary substantially over time, it would also be very useful to have multiple plasma samples for each subject collected over several years.

The impact of PSA-testing in epidemiological studies of prostate cancer

Prostate specific antigen (PSA) is a protein produced by the prostate gland\textsuperscript{147}. The biological function of PSA is to liquefy the semen in order to allow the sperms to penetrate the cervical plug and increase the probability of a sperm to reach an ovum. PSA can be found at low levels in the circulation of healthy men, but in 1980 Papsidero \textit{et al.} recognized that men with prostate cancer often had elevated serum levels of PSA\textsuperscript{148}. The first commercial PSA kit, the Hybritech Tandem-R PSA test, was released in 1986 and the U.S. Food and Drug Administration (FDA) later approved the PSA test to be used as a screening tool for early detection of prostate cancer. Based on the results from a large multicentre trial it was suggested that a threshold of a minimum of 4 ng/mL of serum PSA should be used to select patients for performing biopsies\textsuperscript{149}. Consequently, 4 ng/mL became the universal PSA threshold for performing biopsy even though many cancers will be missed using this limit\textsuperscript{150}. More recently however, a lower PSA threshold has been suggested, and in Sweden a threshold of 3 ng/mL is now recommended whereas the National Comprehensive Cancer Network in the US were even more aggressive and suggested a threshold of 2.5 ng/mL for performing biopsy\textsuperscript{151}. One problem with a fixed PSA threshold is that PSA levels generally increase during life, presumably because of increased occurrence of benign prostatic enlargement (PBE) in older men, see \textbf{Figure 21}. A fixed PSA threshold does not take this variation into account. Thus, one could argue for using an age-dependant threshold; lower for younger men and higher for older men.

There are several other common conditions that may cause PSA levels to rise, such as infection and inflammation (prostatitis)\textsuperscript{152}. These conditions may therefore cause a positive outcome on the PSA test, with unnecessary biopsy as a consequence.
A number of studies have investigated the performance of the PSA test in a cross-sectional setting where biopsy is performed in men with elevated PSA levels\textsuperscript{154,155}. Cross-sectional studies provide the possibility to assess the positive predictive value (PPV), i.e. the fraction of subjects with PSA levels above the threshold with confirmed malignancy on biopsy. However, these studies can not estimate the negative predictive value (NPV), i.e. the fraction of subjects with low PSA levels without cancer, because these subjects do not undergo biopsy. In order to assess the overall validity of the PSA test, we measured PSA levels in a case-control study nested within the NSHDC study (see Material and methods, Study populations). Analysing PSA levels with conditional logistic regression revealed an association signal rarely seen in epidemiological studies, with an odds ratio of prostate cancer diagnosis of 9.0 (95% CI: 6.5 – 12.6, \( p=9 \times 10^{-39} \)), for PSA levels above vs. below 4 ng/mL, respectively. However, because PSA is not a risk factor of prostate cancer, but rather a tumour marker, attempting to assess the discriminative validity using logistic regression clearly produces misleading results. Because PSA levels were measured at recruitment to the cohort in men who subsequently developed prostate cancer and in controls that remained free of disease over a number of years (mean follow-up time: 6.7 years), it was possible to estimate both the sensitivity, i.e. the fraction of the total number of cases found with a specific PSA
threshold, and the specificity, i.e. the fraction of the total number of controls that remained free of cancer during the follow-up below a certain PSA threshold. Compared to logistic regression, the sensitivity and specificity estimates are better indicators of the discriminative performance, and using the 4 ng/mL threshold, we observed a sensitivity of 0.44 and specificity of 0.92. In other words, using this threshold we classified 44% of the cases correctly and 92% of the controls correctly, see Table 4.

Indeed, these numbers indicate that the performance of the PSA test is very good, but for a biomarker to be considered as a screening tool in a healthy population, very high performance must be achieved. In particular the specificity must be extremely high in order to avoid a large number of healthy individuals being followed up. Therefore, a specificity of at least 0.98 and a sensitivity of at least 0.5 have been suggested as criteria when evaluating biomarkers to be used in screening, and our data indicate that the PSA test does not meet those requirements\textsuperscript{156}. The distribution of the logarithm of total PSA for cases and controls in the NSHDC study shown in Figure 22 illustrates the spread of PSA levels in the Västerbotten county.

In paper V of this thesis, we divided the study into two study waves with cases of wave 1 diagnosed between 1994 and 2003, and cases of wave 2 diagnosed between 2001 and 2005. Analysing differences in tumour characteristics between the study waves clearly shows a shift towards cancers with lower T-stage, lower PSA levels at blood draw and a lower fraction of malign tissue in biopsies in wave 2 compared to wave 1, see Table 3 in paper V. To a large extent, this shift seems to be attributable to a higher proportion of cases in wave 2 being diagnosed after health

### Table 4. Discriminative performance of total PSA in prediction of subsequent prostate cancer diagnosis

<table>
<thead>
<tr>
<th>Threshold (total PSA) [ng/ml]</th>
<th>Cases n=540</th>
<th>Controls n=1047</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;5</td>
<td>180</td>
<td>55</td>
<td>0.33</td>
<td>0.95</td>
</tr>
<tr>
<td>&gt;4</td>
<td>236</td>
<td>84</td>
<td>0.44</td>
<td>0.92</td>
</tr>
<tr>
<td>&gt;3</td>
<td>321</td>
<td>138</td>
<td>0.59</td>
<td>0.87</td>
</tr>
<tr>
<td>&gt;2.5</td>
<td>370</td>
<td>191</td>
<td>0.69</td>
<td>0.82</td>
</tr>
</tbody>
</table>
check-up than cases in wave 1, i.e. by opportunistic PSA testing.

Figure 22. Distribution of log-PSA levels for cases and controls of the NSHDC study. The threshold of 5 ng/mL is marked giving a specificity of 95 % and a sensitivity of 33 %.

On the basis of results from autopsy studies showing that 15 to 60 % of men 60 to 90 years of age have undiagnosed prostate malignancies, and the high incidence – but intermediate mortality – of prostate cancer in the USA (Figure 1), it is evident that prostate cancer is a highly heterogeneous disease with many clinically insignificant diagnoses. Here, clinically insignificant refers to men diagnosed with prostate cancer cases that would have gone through life with little or no symptoms of the disease. As indicated in Paper V, with the advent of opportunistic PSA testing, the case-mix in epidemiological case-controls studies shifts towards less clinically significant cases, thus making the case-population phenotypically more similar to the control population, see Figure 23.

In epidemiological studies, the definition of the endpoint, i.e. the case-control status, is immensely important when trying to find risk factors. In studies with many controls misclassified as cases, i.e. clinically insignificant cases, the power of finding a true association will decrease substantially. In order to make a clearer distinction between the cases and the controls, one may argue for paying more attention to sub-analysis of high risk cases, i.e. cases with poor prognosis with locally
advanced disease (T3/T4), metastasis or very high PSA levels. This sub-analysis may also be motivated by the wish to identify risk factors of high-risk disease rather than low-risk disease, since little is know about why some cases progress, while others do not.

Perhaps it is time to realize that prostate cancer is not a discrete disease, but rather a continuous trait, and that the case-control way of thinking is not appropriate in epidemiological studies of prostate cancer in the PSA era? One potentially powerful approach to attack this problem would be to construct a continuous phenotype based on detailed clinical tumour characteristics and PSA levels, preferably weighted against fatal outcome. This would also enable the possibility to use analytical methods more powerful than logistic regression, such as linear regression.

Figure 23. Schematic figure of misclassification of controls and cases in prostate cancer case-control studies.
Future prospects of prostate cancer epidemiology

One of the major obstacles in prostate cancer epidemiology, as well as in epidemiology as a whole, has been the large body associations reported from inadequately powered studies. Most of these associations, in particular those between genetic variants and risk, have never been replicated. The probability that an association displaying a p-value of less than 0.05, is actually a false positive association, greatly increases when reported by a study with low power of detecting a true association. In fact, Colhoun et al. estimated that 95% of all findings from genetic association studies are false positives\(^{158,159}\). Furthermore, this problem multiplies when investigating variants with no explicit prior hypothesis of its relation to the disease, i.e. as in studies utilizing various tagging approaches and even more so in genome-wide studies\(^{159}\). This issue is now widely recognized and as a response, several large collaborative studies have emerged such as the Breast and Prostate Cancer Cohort Consortium including 9,000 cases (BPC3), and The Endogenous Hormones and Prostate Cancer Collaborative Group including 4,000 cases\(^{35,160}\). Still, in order to reach the levels of statistical significance required for genome-wide significance \((p < 5 \times 10^{-7})\), these studies are only powered to detect the most common disease variants\(^{161}\). Thus, we can expect even larger collaborative studies appearing over the next few years which will lead to the discovery of more, less common, susceptibility loci. These studies will hopefully also lead to conclusive evidence on the importance of previously implicated environmental factors in prostate cancer aetiology. However, in the backwaters of the results from the genome-wide association studies, we will face a massive challenge in deciphering the biological implications of these findings.

It has been widely accepted that the aetiology of complex diseases, e.g. prostate cancer, is not only explained by explicit environmental and genetic exposures, but also their interactions\(^{162}\). An obvious example of a gene-environmental interaction is the much stronger effect of sunlight for fair skinned people than for dark skinned people in the predisposition of skin cancer. Investigations of other less obvious interactions, utilizing biomarker analyses in genetic association studies will also be a major topic in the near future. As pointed out above,
however, sample size is crucial in epidemiological studies overall, and it might be an even stronger limitation in interaction studies\textsuperscript{162,163}. Furthermore, GWAS databases will enable exhaustive explorations of gene-gene interactions, leading to the development of new statistical methods. Given the exploding multiple comparisons issues inherent in these studies, a more feasible approach might be to focus on interactions between genetic variants already showing significant main effects.

It seems clear that the next decade will bring many important clues to the aetiology of prostate cancer, in particular in the identification of genetic risk loci. An important question is what these findings will mean in clinical practice? One can argue that the identification of novel susceptibility loci give clues to mechanisms regarding the biological background to the disease, thus leading to the development of new therapeutic and preventive strategies. Furthermore, early attempts in combining known genetic variants found by GWAS in risk classification have already been made, but the discriminative performance of this strategy is still limited, mostly because of the high number of risk variants required to make such an approach successful\textsuperscript{164,165}. It has been estimated that in order to achieve adequate discriminative performance in prediction of a future diagnosis, more than 50 independent genetic variants with relative risks of at least 1.5 are needed\textsuperscript{164}. It is not unreasonable to expect this number of risk variants to be identified for prostate cancer in the near future by combining results from GWAS. Given the huge investments made in GWAS, it seems inevitable that we will see direct-to-customer genotyping services providing information on customers risk profiles for developing various diseases. In fact, as of November 2007 two companies provide such services, and even though they still seem premature scientifically, the concept is not unfeasible in the future, in particular in the identification of subjects who would benefit from preventive measures\textsuperscript{41,166}. However, this development warrants an extensive debate regarding the ethical and practical implications in future health care.
SUMMARY AND CONCLUSIONS

In the papers included in this thesis, we investigated factors of two pathways and their implications in prostate cancer aetiology. In the insulin-like growth factor pathway, information on both genetic and circulating protein variations were used, and we found a variant of the \textit{IGF1} gene associated with increased prostate cancer risk in paper I. This variant was later related to elevated circulating levels of IGF1 in paper II, suggesting that the increased prostate cancer risk noted in paper I was mediated by circulating IGF1 levels. The results from papers I and II also gave further support for the causal relationship between circulating IGF1 and prostate cancer risk along the lines of Mendelian randomization. In paper III we investigated genetic and plasma variations of the most important IGF-binding proteins and found that elevated levels of circulating intact IGFBP3 were associated with increased risk of prostate cancer-specific death. It is important to note that this association was found in a study in which the blood samples had been collected after the cancer diagnosis. Therefore, given the design of our study, we can not exclude that this association was caused by treatment or the tumour itself. Nevertheless, the association between intact IGFBP3 and survival is interesting and warrants further investigations in other study designs, such as experimental- and prospective studies.
In papers IV and V we investigated circulating factors of the one-carbon metabolism pathway and found some evidence that high levels of choline, vitamin B2 and B12 increase the risk of prostate cancer. Even though these studies were performed in well characterized, relatively large prospective studies, the associations were statistically week overall, and further studies are warranted in order to confirm these findings.

Based on the papers underlying this thesis, the following conclusions can be drawn:

- Genetic variation in the *IGF1* gene affects prostate cancer risk by causing the circulating levels of IGF1 to rise. This result gives further support for the causal relationship between circulating IGF1 and prostate cancer risk.

- Elevated circulating levels of intact IGFBP3 are associated with increased risk of prostate cancer-specific death as measured in retrospectively collected plasma. Further studies are warranted to conclude the biological meaning of this finding.

- Factors of one-carbon metabolism seem to be involved in prostate cancer aetiology. Implicated factors include choline, vitamin B2 and B12.
Prostatan är en könskörtel och en viktig del av mannens fortplantningssystem, se Figur 24. Dess främsta funktion är produktion och lagring av sekret som ingår i sådesvätskan. Trots att prostatacancer är den vanligaste cancersjukdomen i Sverige med drygt 9 000 diagnostiserade fall och 2 500 dödsfall varje år, vet man förvånansvärt lite om faktorer som påverkar risken för att utveckla sjukdomen.

Målet i delarbete I, II och III i denna avhandling var att undersöka om genetisk variation i gener i den så kallade IGF-signalvägen påverkar prosatacancerrisken. IGF1 står för insulin-like growth factor-1 och är ett tillväxthormon som vid höga nivåer gör att cellerna delar sig (prolifererar) snabbare, samt att den programmerade celldöden (apoptos) mattas av. Dessa egenskaper gör att personer med höga nivåer av IGF1-proteinet i blodet har en ökad risk för att få prosatacancer, ett samband som har noterats i flertalet studier. Man vet att nivåerna av IGF1 i blodet påverkas av både kost och genetiska faktorer.

**Figur 24.** Prostatan sitter i bäckenet, nedanför urinblåsan och framför ändtarmen. Bilden är hämtad från wikimedia commons (http://commons.wikimedia.org/wiki/Main_Page)
I delarbete I undersökte vår forskningsgrupp om genetiska varianter av \textit{IGF1}-genen påverkar risken för prostatacancer. Detta gjordes genom att analysera ett antal punktmutationer i genen, så kallade SNPs, i ett antal prostatacancerfall och friska kontroller från \textit{Cancer Prostate in Sweden} – studien (CAPS). CAPS är en så kallad fall-kontroll-studie vilket innebär att både diagnostiserade fall och friska kontroller ingår. I CAPS har nästan 3 000 prostatacancerfall och 2 000 kontroller samlats in. Frekvenserna för SNPs jämfördes sedan mellan cancerfallen och de friska kontrollerna och det visade sig att personer med en specifik genetisk variant i slutet av \textit{IGF1}-genen hade en ökad risk med nästan 50 \% för prostatacancer.

I delarbete II undersökte vi om den genetiska riskvarianten från delarbete I berodde på en genetisk påverkan av IGF1-nivåerna i blodet. Denna studie utfördes i både CAPS-studien (se ovan) och i \textit{Northern Sweden Health and Disease Cohort}–studien (NSHDC). NSHDC-studien är en så kallad prospektiv studie där ett stort antal personer har bjudits in att delta innan de fått sin diagnos. Totalt har knappt 100 000 personer samlats in i NSHDC-studien och av dessa har nästan 600 personer utvecklat prostatacancer. Dessutom sammanställdes data från andra studier i vilka både genetisk variation i \textit{IGF1}-genen samt nivåer av IGF1 i blodet analyserats. Sammanfattningsvis fann vi att de genetiska varianterna som gav ökad risk för prostatacancer i delarbete I, även gav förhöjda nivåer av IGF1 i blodet. Från delarbete I och II drogs därför slutsatsen att genetiska varianter i \textit{IGF1}-genen ökar risken för prostatacancer genom att orsaka förhöjda nivåer av IGF1 i blodet. Dessa fynd ger dessutom ett ökat stöd för att höga nivåer av IGF1 i blodet orsakar en ökad risk för prostatacancer.

En stor del av IGF1 i blodet är bundet till så kallade bindningsproteiner. I delarbete III undersökte vi om genetisk variation i specifika gener som kodar för bindningsproteiner till IGF1, påverkar risken för att få prostatacancer, samt risken för att dö på grund av sjukdomen. Dessa gener inkluderade \textit{IGFBP1}, \textit{IGFBP3} samt \textit{IGFALS}. Dessutom undersöktes om nivåer av totalt IGFBP3 och intakt IGFBP3 i blodet påverkar risken för att dö i prostatacancer. Denna studie utfördes också inom CAPS-studien. Sammanfattningsvis kunde inget samband mellan
genetisk variation och risken för att få prostatacancer, eller att dö på grund av prostatacancer påvisas. Dock var förhöjda nivåer av intakt IGFBP3 i blodet starkt relaterat till en ökad risk för att dö på grund sjukdomen, men vi kunde inte utesluta att detta fynd berodde på andra orsaker som till exempel tumören eller behandlingen.

Flera studier har visat att risken för missbildningar vid födseln minskar när modern har högt dietintag av folsyra. Dessa fynd har lett till att flera länder, inklusive USA och Kanada, har infört obligatorisk folatberikning av vissa livsmedel. Man har också funnit att folat kan minska risken för vissa cancersjukdomar. I en nyligt publicerad studie hade dock personer som fick folatillskott en ökad risk för prostatacancer samt förstadiet till kolorektalcancer. I Sverige pågår för närvarande en intensiv debatt om huruvida folatberikning ska införas, men bland annat på grund av att sambandet mellan folat och cancer är oklart, har folsyratillskott hittills bara rekommenderats till kvinnor i fertile ålder.

Folat är tillsammans med vitamin B12 viktiga komponenter i enkolsmetabolism-signalvägen som påverkar DNA-syntes och metylering (reglering av gener). Vår forskargrupp har tidigare sett att höga nivåer av folat, och framförallt vitamin B12, i blodet ger ökad risk för prostatacancer i en studie utförd inom NSHDC. I delarbete IV ville vi konfirmera dessa fynd i samarbete med European Prospective Investigation into Nutrition and Cancer (EPIC)-studien, ett stort europeiskt samarbete som inkluderar totalt 500 000 personer. Sammanfattningsvis fanns dock inget starkt samband mellan nivåer av folat och vitamin B12 i blodet och prostatacancer risken. I delarbete IV ingick det, förutom personer från övriga EPIC, även personer från NSHDC, den studie där man tidigare hade funnit ett starkt samband mellan vitamin B12 och prostatacancer. Dessa personer hade dock inte varit med i den ursprungliga studien som visade det starka sambandet mellan vitamin B12 och prostatacancer. En märklig observation i delarbete IV var avsaknaden av samband mellan vitamin B12 och prostatacancer risken hos personerna från NSHDC.

För att undersöka de motsägelsetfulla resultaten för vitamin B12 i delarbete V, delade vi därför hela NSHDC-studiegruppen i två delar, varav den första inkluderade personer som var med i den första studien från NSDHC, och den andra gruppen personer från NSHDC som var med i delarbete IV. Dessutom analyserades nivåer av åtta ytterligare B-

Sammanfattningsvis har denna avhandling gett ett ökat stöd för att IGF1 spelar en viktig roll i utvecklingen av prostatacancer. Att intakt IGFBP3 var kraftigt relaterat till ökad risk för död i prostatacancer är intressant, men ytterligare forskning krävs för att utreda den biologiska bakgrunden till detta fynd. Vi har dessutom sett indikationer på att flera faktorer inom enkolsmetabolism-signalvägen såsom kolin, vitamin B2 och B12 har en betydelse för utvecklingen av prostatacancer men dessa fynd var statistiskt svaga och bör därför undersökas i fler prospektiva studier innan några slutsatser kan dras.
ACKNOWLEDGEMENTS

Many have contributed in making this work possible. In particular I want to acknowledge the following:

All study participants in the CAPS, NSHDC and EPIC studies. Without your generous contribution to medical research the studies underlying this thesis would never have been done.

Pär Stattin – my main supervisor and mentor – for introducing me to medical research and scientific writing. For your generosity and enthusiasm. You have always believed in me, giving me support when I’ve been stressed out. I genuinely hope that we will work together for a long time. I wish we had more time for skiing though…

Henrik Grönberg and Fredrik Wiklund – my co-supervisors – for inviting me with open arms into the research group making me feel welcome. For teaching me about genetic epidemiology and giving me the opportunity to work with the CAPS study.

Sara Lindström – my best friend and roommate for the last three years – for giving me the opportunity to work in this field and for teaching me about genetics. You’ve always stood by my side making every day just a little bit shorter, even at times when I’ve been obnoxious. I hope you realize how important you are to me.
Tanja Stocks – my dear friend and colleague – for your genuine smile, straightforwardness and for teaching me about cooking. I’m really going to miss you and our dinners together.

James McKay – IARC collaborating powder hound – for endless Skype conversations, always pushing me to think one step further. For making me feel as if I was actually contributing. BTW, I’m getting fat-skis for 2009…

Everyone at the Urology Department – for showing your interest in my work even though I’m rarely around, Solveig for keeping track of me when I’ve forgot meetings and deadlines, for helping me with all administrative issues always being ice cold. Imagine the chaos that would emerge on the department if you left. Börje for employing me despite my strange background.

All my colleagues at Onkologiskt Centrum – for treating me as one of you. For numerous laughs on coffee breaks and festivities. Lena D for letting me stay at OC and making me feel at home, Åsa for sharing my sick sense of humour and heaps of music, Benjamin, my room mate for the last year, I guess you’re sick and tired of me by now?

All the people in the research group – Fredrik, Björn-Anders, Bea, Kristina, Katarina, Camilla, Monica and Bettan for your warm welcome and for inspiring me to complete this thesis.

Rudolf Kaaks – DKFZ collaborator – for giving me the opportunity to enjoy some of your infinite well of wisdom and for always taking the position as the sceptic reviewer, teaching me to criticize my results. It’s been a true pleasure and I hope we will work together again.

The crew at IARC in Lyon – Sabina, Martijn, Carine, Laure, Becki and Fatiah for taking care of me in a country where everyone spoke a strange language.

Tim Key and all the people at the cancer epi-unit at Oxford University – Paul, Naomi, Ruth and Andrew for giving me the opportunity to work with the EPIC study and for inviting me to Oxford.
The staff at the **Medical Biobank** in Umeå – for being the only reliable players in a field where customer support and delivery deadlines not yet have been introduced. Åsa Ågrén for always managing to schedule our projects in times of uncertainties. Göran Hallmans for your encouragement and for initiating the biobank, creating an invaluable source for medical research.

Everyone I met on meetings and courses around the world, in particular Maike, Fredrik and all the other I met in Florence for drinking mediocre wine into the night on the backside lawn. You made it all worth while. Sara for having fun skiing in Tärna/Hemavan. I also want to thank Amanda for letting me stay in her flat in Boston over the most stressful period. I promise I’ll be more fun the next time we meet.

**All co-authors** – for valuable contributions to this work, Hans-Olov Adami for showing me that simple is always better, Bethany for fruitful discussions and inspirational collaborations, and Anders Bergh always having a minute to spare to share your wisdom.

All my **physics friends** from the university – Tom, David, Ulf and everyone else for giving me the best time of my life, I truly miss you all. Jesper for sharing the horrors of sitting on the edge of a cliff in the middle of a blizzard and for being a really good friend, Kent and Desiré for endless discussions over the more important things in life and for always making me feel welcome in your home.

**My family** – Karl-Axel, Anne-Christine and Johanna for your unconditional love, always making me feel as if there is a place for me to return. I could never express my gratitude in words nor return all your sacrifices over the years. You brought me up to be the person I am today, for better and for worse.

**Sofia** – for supporting me over the last couple of months giving me energy when I needed you the most. For always giving me something to long for, and most of all, for showing me what love is. Jag älskar dig.

This work was supported by the **Medical Faculty** at Umeå University, the **County Council of Västerbotten**, the **Swedish Cancer Society**, **Lion's Cancer Research Foundation**, **World Cancer Research Found** and United States Army Medical Research and Material Command.
REFERENCES


