Gene Expression Patterns in a Rat Model of Human Endometrial Adenocarcinoma

Sandra Karlsson

Department of Cell and Molecular Biology – Genetics
Lundberg Institute, Faculty of Science
2008

School of Life Sciences, Systems Biology – Biomedicine
University of Skövde

UNIVERSITY OF GOTHENBURG

Thesis book:
Gene Expression Patterns in a Rat Model of Human Endometrial Adenocarcinoma

ISBN: 978-91-628-7648-7

© Sandra Karlsson
sandra.karlsson@his.se

Department of Cell and Molecular Biology – Genetics
Lundberg Institute, Faculty of Science
University of Gothenburg

Printed in Sweden 2008
Vasastadens bokbinderi AB, Göteborg

"Do not worry about your difficulties in Mathematics. I can assure you mine are still greater."

Albert Einstein

ABSTRACT

Gene Expression Patterns in a Rat Model of Human Endometrial Adenocarcinoma

Sandra Karlsson

Endometrial cancer develops from the endometrium of the uterus and is the most common pelvic malignancy diagnosed in women in the western society. Similar to all cancer diseases, endometrial cancer is a disorder that results from complex patterns of genetic and epigenetic alterations involved in the malignant transformation. The BDII/Han rat model is unique for spontaneous hormonal carcinogenesis since more than 90% of the female virgins spontaneously develop endometrial cancer. The possibility to perform global gene expression profiling of tumor cells would likely provide important information of the genes and pathways that are aberrant in endometrial adenocarcinoma (EAC). The works in the present thesis have been focused on investigating the expression patterns in endometrial tumors.

The findings in this thesis involve the identification of a novel candidate tumor suppressor region of rat chromosome 10. This genomic segment contains 18 potential tumor suppressor genes. Preliminary microarray data analysis confirmed that this region might contain relevant candidate genes as the EACs on average had 3.8 times lower expression of *Crk* in comparison to the normal/pre-malignant endometrial tissue cultures. Furthermore, an expression analysis using qPCR, revealed a significant down-regulation of *Myo1c* and *Hic*.

We were also able to identify a group of genes associated with the TGF-β pathway that were differentially expressed between endometrial tumors and normal/pre-malignant endometrium. These results suggest that the TGF-β signaling pathway is disrupted in EAC. This has previously been demonstrated in human EAC, although this is the first report on aberrant expression of TGF-β down-stream target genes.

Evaluation of *Gpx3* down-regulation in the rat EAC cell lines revealed an almost complete loss of expression in a majority of the endometrial tumors. From methylation studies, we could conclude that the loss of expression of *Gpx3* is correlated with biallelic hypermethylation in the *Gpx3* promoter region. This result was confirmed with a demethylation study of EAC cell lines, where the *Gpx3* mRNA expression was restored after treatment with a demethylation agent and a deacetylation inhibitor. We also showed that mRNA expression of the well-known oncogene, *Met*, was slightly higher in endometrial tumors with loss of *Gpx3* expression. A likely consequence of loss of *Gpx3* function is a higher amount of reactive oxygen species (ROS) in the cancer cell environment. Since it has been proposed that overproduction of ROS is required for the hypoxic activation of HIF-1, we suggest that loss of *Gpx3* expression activates transcription of *Met* through induction of the transcription factor HIF-1. The loss of the protective properties of GPX3 most likely makes the endometrial cells more vulnerable to ROS damage and genome instability.

We extended the results obtained from the rat endometrial tumors to human material, and conducted expression analysis of *GPX3* in 30 endometrial human tumors using qPCR. The results showed a uniformly down-regulation of *GPX3* in 29 of the tumors, independent of tumor grade. We thus concluded that the down-regulation of GPX3 probably occurs at an early stage of EAC and therefore contributes to the EAC carcinogenesis. These results suggest that there are important clinical implications of *GPX3* expression in EAC, both as a biomarker for EAC and as a potential target for therapeutics.

PAPERS INCLUDED IN THE STUDY

The present study is based on four papers, which will be referred to in the text by their Roman numerals:

I. Nordlander C, **Karlsson S**, Karlsson A, Sjöling A, Winnes M, Klinga-Levan K, Behboudi A. *Analysis of chromosome 10 aberrations in rat endometrial cancer-evidence for a tumor suppressor locus distal to Tp53.* Int J Cancer. 2007 Apr 1;120(7):1472-81.

II. **Karlsson S**, Holmberg E, Askerlund A and Klinga-Levan K. *Altered TGF-β pathway expression pattern in rat endometrial cancer*. Cancer Genet Cytogenet. 2007 Aug;177(1):43-50.

III. **Karlsson S**, Olsson B & Klinga-Levan K. *Gene expression profiling predicts a three-gene expression signature of endometrial adenocarcinoma in a rat model*. Submitted 2008.

IV. **Karlsson S**, Falck E, Carlsson J, Helenius G, Karlsson M & Klinga-Levan K. *Loss of expression of Glutathione peroxidase 3 in endometrial cancer is correlated with epigenetic mechanisms*. Manuscript.

TABLE OF CONTENTS

TABLE OF CONTENTS - 6 -
LIST OF ABBREVIATIONS - 7 -
INTRODUCTION - 8 -
CANCER - 8 -
Cancer – a complex genetic disease - 8 -
Cancer genes - 9 -
Endometrial cancer - 9 -
Animal models - 10 -
GLOBAL GENE EXPRESSION PROFILING - 11 -
Statistical analysis of gene expression data - 12 -
AIMS OF THE STUDY - 14 -
MATERIALS AND METHODS - 15 -
EXPERIMENTAL MATERIALS - 15 -
Animal crosses and tumor material - 15 -
Human endometrial tumor material - 17 -
METHODS - 17 -
Global gene expression profiling - cDNA microarray experiments - 17 -
Design of the microarray experiments - 17 -
Microarray hybridizations - 18 -
Statistical analysis of the microarray data - 18 -
Exploratory data analysis - Hierarchical Clustering - 22 -
Statistical inference analysis of significantly differentially expressed genes - 22 -
Transmission Disequilibrium Test (TDT) - 23 -
Reverse Transcription PCR (RT-PCR) and real time quantitative PCR (qPCR) - 23 -
Chromosome paint and dual-color fluorescent in situ hybridization (FISH) - 24 -
Mutation screening by DNA sequencing - 25 -
Methylation-specific PCR (MSP) - 25 -
IDENTIFICATION AND ANALYSIS OF GENES INVOLVED IN EAC - 26 -
GLOBAL GENE EXPRESSION ANALYSIS - 26 -
Identifying potential tumor suppressor gene candidates on RNO10 (paper I) - 26 -
Aberrant expression of genes associated with the TGF- β signaling pathway (paper II) - 28 -
A three-gene signature of EAC (paper III) - 30 -
EVALUATION OF GENES IDENTIFIED FROM THE MICROARRAY EXPRESSION STUDIES (PAPER IV) - 33 -
Epigenetic mechanisms responsible for loss of Gpx3 mRNA expression - 33 -
EXPRESSION OF *GPX3* IN HUMAN ENDOMETRIAL TUMORS (PAPER IV) - 35 -
CONCLUDING REMARKS - 37 -
ACKNOWLEDGEMENTS - 39 -
REFERENCES - 41 -

LIST OF ABBREVIATIONS

BASE	Bioarray Software Environment
BDII	BDII/Han (inbred rat strain)
BGN	Betaglycan
BN	Brown Norway (inbred rat strain)
CAT	Catalase
cDNA	complementary DeoxyriboNucleic Acid
CGH	Comparative Genomic Hybridization
cRNA	complementary Ribonucleic Acid
Cy3, Cy5	Cyanine3, Cyanine5
DAVID	Database for Annotation, Visualization and Integrated Discovery
DNA	DeoxyriboNucleic Acid
EAC	Endometrial AdenoCarcinoma
EC	Endometrial Cancer
EST	Expressed Sequence Tags
F1	First generation of a cross, first filial
F2	Second generation intercross (F1xF1)
FFPE	Formalin Fixed Paraffin Embedded
FISH	Fluorescent in Situ Hybridization
FWER	Family Wise Error Rate
FDR	False Discovery Rate
GPX3	Glutathione PeroXidase 3
LOESS	LOcal Scatterplot Smoothing
mRNA	messengerRiboNucleicAcid
MET	Mesenchymal-Epithelial Transition factor
N1	Backcross generation
NUT	rat uterine tumor developed in the backcross (N1) progeny
NME	Non-malignant Endometrium (or Normal/pre-malignant Endometrium)
PCR	Polymerase Chain Reaction
qPCR	Real time Quantitative PCR
RNA	RiboNucleic Acid
RNO	Rattus Norvegicus (rat chromosome)
ROS	Reactive Oxygen Species
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
SPRD-Cu3	Sprague-Dawley-curly3 inbred rat strain
SOD	SuperOxide Dismutase
TGFB3	Transforming Growth Factor beta 3
TDT	Transmission Disequilibrium Test
Weka	Waikato Environment for Knowledge Analysis

Notes on nomenclature: Gene symbols contain letters and Arabic numerals. Human gene symbols are written with all capitals, whereas those for rat are in lower case letter, initialized by capital. Gene symbols are italicized in the text. Protein designations are the same as the gene symbol, but are not italicized; all letters are in uppercase.

INTRODUCTION

Cancer

Cancer – a complex genetic disease

Cancer is the general name for a class of more than 200 neoplastic diseases affecting more or less all organs and tissues in the body. Although there are many different cancers, they all start as abnormal cells growing beyond their usual boundaries. Metastases, the major cause of death from cancer, are cancerous cells that have gained the capacity of invading adjoining parts of the bodies and are then spread to other organs. According to the World Health Organization (WHO), 7.9 million people in the world died from cancer during 2007 and it is estimated to cause the deaths of 12 million people in 2030 [1].

Cancer is per definition a genetic disease and during the last decades it has become clear that only a minor proportion of cancers are caused by one single highly penetrant gene. The majority of cancer diseases are rather caused by intricate interactions among genetically or epigenetically altered genes. Single mutations, which might be inherited or spontaneous, are generally not sufficient to give rise to cancer, but they may initiate cells to turn to a malignant growth. Additional changes in other genes, caused by damages from the environment, progress the cells' malignant transformation [2-6]. Hence, cancer is a multi-step process that involves initiation, promotion, transformation and progression. For common epithelial cancers development, it has been estimated that 4-7 rate-limiting genetic events are required (Figure 1) [7]. The last step of the accumulated genetic changes, is the promotion of the pre-malignant cells to true neoplasias, which are characterized by uncontrolled proliferation, loss of normal cell function and morphology, sustained angiogenesis and the ability to metastasize and invade tissues beyond the immediate primary tumor location [8].

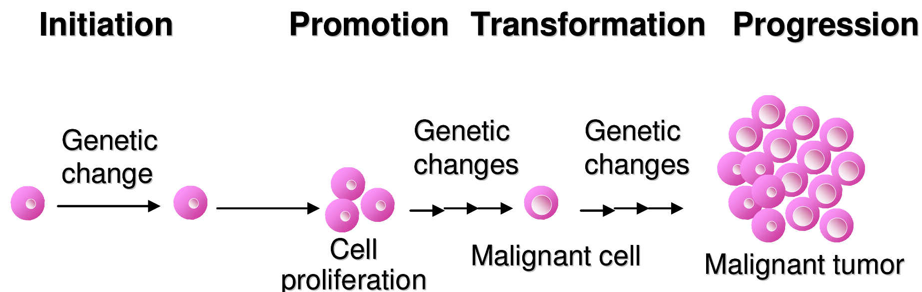

Figure 1. A series of genetic changes that lead to cancer. Cancer develops through a multi-step process of multiple genetic changes during an extended period of time. Each change enables pre-cancerous cells to acquire some of the characteristics that together produce the malignant growth of cancer cells. This figure illustrates only a few genetic changes, but carcinogenesis probably involves about 4-7 changes [7].

The inherited tendency to develop cancer varies among individuals. Highly penetrant mutations cause strong genetic predisposition to cancer disorders and confer Mendelian patterns of inheritance. More than 200 of such cancer susceptibility syndromes have been described, but they are rare in the human population accounting for only 5-10% of all cancers [9]. Several population based epidemiological studies have shown that the genetic predisposition to the majority of all cancers involves a collective impact of several low

penetrance alleles with minor/modest effects. Hence most inherited cancers are considered to be complex, polygenic disorders which rarely follow simple Mendelian rules.

The penetrance of mutations associated with carcinogenesis may be dependent on genetic background, life style and environmental factors. The genetic heterogeneity present in the human population poses the greatest hurdle when analyzing the contribution of low-penetrance genes to cancer etiology. Similar tumor phenotypes may, for instance, be the result of alterations in different genes. It has been shown in mouse models that predisposition caused by combinations of weak genetic variants can exert a profound influence on cancer susceptibility, thus it is likely that the inheritance of most common cancers is polygenic [9-12].

Cancer genes

Generally, there are three classes of genes with great importance in tumor etiology; i.e. oncogenes (tumor promoting genes), tumor suppressor genes (tumor inhibiting genes) and stability genes (care-taker genes). Oncogenes are mutated normal cellular genes, so called proto-oncogenes, whose products participate in cellular growth and controlling pathways. Proto-oncogenes are generally activated via i) intragenic gain-of-function mutations that might result in a changed protein activity, ii) chromosomal translocations and iii) gene amplifications [13, 14].

In contrast, the activity of tumor suppressor genes is reduced by genetic alterations such as miss-sense mutations, mutations that result in truncated protein, deletions and insertions or from epigenetic silencing [2]. In addition, tumor suppressor genes usually follow the two-hit hypothesis, initially proposed by AG Knudson [15, 16], which means that both alleles must be affected in order to manifest loss of function of the specific tumor suppressor gene. This means that mutant tumor suppressor alleles are recessive, whereas mutant oncogenes alleles are dominant. The functions of the proteins encoded by tumor suppressor genes involve repression of genes essential for cell division, coupling of the cell cycle to DNA damage and cell adhesion [17, 18]. One important and well-known tumor suppressor gene is *Tp53* (Tumor protein 53). Homozygous loss of *Tp53* has been found in 70 % of colon cancers, 30-50 % of breast cancers and 50% of lung cancers. The anti-cancer mechanisms of TP53 activates DNA repair proteins, facilitates the repair system by holding the cell cycle at the G_1/S regulation point and induces apoptosis [19-21].

The third group, the stability genes, or care takers, include mismatch repair genes, nucleotide-excision repair genes and genes that control processes involving large portions of chromosomes, such as those responsible for mitotic recombination and chromosomal segregation. The stability genes thus minimize genetic alterations and decrease mutation rate when active [22].

Endometrial cancer

In the western society, endometrial carcinomas represent the most prevalent neoplasms of the female pelvis and are the third most common cause of gynaecological cancer deaths, only exceeded by ovarian and cervical cancer. The incidence rate varies worldwide but is highest among white women in western populations, post-menopausal women being predominantly affected [23]. Endometrial carcinoma refers to different cancer diseases that arise from the endometrium, the inner lining of the uterus. Most endometrial cancers are usually

adenocarcinomas (also known as endometrioid), meaning that they originate from epithelial cells that line the endometrium where they form the glandular cells in the uterus. Roughly, endometrial carcinomas can be categorized into two subgroups based on histopathology. The most common subtype, the low-grade endometroid type I, typically debuts prior and during menopause, and displays a relatively low aggressiveness. They arise in an environment of excessive estrogen exposure and are frequently preceded by endometrial hyperplasias. Type II endometrial carcinomas on the other hand, usually debut in older post-menopausal women and are not associated with increased exposure to estrogen. These tumors are typically of high-grade endometroid adenocarcinomas, papillary serous or clear cell types, and generally carry a poor prognosis. The most common therapeutic approach is a total abdomal hysterectomy with a bilateral salpingo-oophorectomy (i.e. surgical removal of both ovaries). Hormonal treatment with progestins and antiestrogens has also been used in the therapy of endometrial stromal sarcomas [24-27].

Animal models

To reduce and control the factors contributing to cancer diseases, it is appropriate to turn to an animal model system. Inbred rodent models have contributed enormous to our understanding of biology and etiology of a variety of traits and have been widely used for studies of many complex diseases. More than 500 different inbred rat strains, carefully characterized with respect to genetical and physiological characteristics, are available and many of them constitute excellent models of human complex diseases. Compared to the mouse, which also is one of the most widely used model system for genetic diseases, the rat model has several important advantages. Rat pregnancies are more size consistent, rat cycling is relatively non-pheromonal (similar to human) and rats can be bred quickly after parturition. The adequate size of the rat also allows for many important measurements to be quantified, for example invasive procedures [28-31]. After identification of potential disease genes and their function in rats, the pathophysiological mechanisms can be elucidated and human genetic counterparts can thus be more easily identified.

During the last decades, large amount of genome data from human and animal models have been generated, including the complete DNA sequence of rat and human. By employing methods such as comparative mapping, it is possible to take advantage of the results from experiments performed in rats when analyzing human diseases. It has been estimated that approximately 90% of the coding sequences in the rat posses strict orthology to genes in both human and mouse genomes [32-34].

The BDII/Han inbred rat model

Currently, four experimental rat models for spontaneous endometrial tumorigenesis are available (Wistar/Han, DA/Han, Donroy and BDII/Han) [35]. However, the BDII/Han strain is unique, since the incidence of spontaneously developed EAC is high with more than 90% of the female virgins affected. The present thesis is based on studies performed on endometrial cell lines from tumors developed in F1, F2 and N1 (backcross) progeny from crosses between the BDII females and two non-susceptible inbred strains (SPRD*Cu3* and BN). Since the endometrial tumors in the BDII rat strain are estrogen dependent and their histopathology and pre-malignant stages of development resembles human endometrial carcinoma, it represent an excellent model for human EAC.

Global gene expression profiling

The advance of the Human Genome project and the availability of genome sequence information, have paved the way for gene expression studies on a genome-wide scale. DNA microarrays are powerful and versatile tools that allow comparison between different conditions across tens of thousands of specific mRNAs in one single experiment. The microarray technology is relatively new but has already rendered a great impact on cancer research. The applications range from identification of new drug targets, new molecular tools for diagnosis and prognosis, as well as for a tailored treatment that will take the molecular determinants of a given tumor into account [36].

The DNA microarray technologies are performed on oligonucleotide chips, glass slide cDNA arrays or nylon-based cDNA arrays. Microarraying allows the comparison of gene expression profiles from two or more tissues or from the same tissue in different biological conditions. The technologies have had some drawbacks but continue to develop. Each platform has its own specific advantages and disadvantages; however the most important consideration is the ability of the technology to address the chosen hypothesis. The two most commonly used DNA microarray platforms, customized cDNA microarrays (two-channel format) and commercially produced high-density oligonucleotide microarrays (one-channel format), differ mainly in the type of solid support on which arrayed elements are immobilized and the method of arraying (Figure 2).

In more detail, the two-channel format (cDNA microarrays) employs PCR amplified expressed sequence tag (EST) clones, full-length cDNAs or oligonucleotides (50-70mers) that are spotted onto glass slides (generally microscope slides). With present technology, up to 30 000 elements can be printed on one microscope slide. Two differently labeled samples (typically Cyanine3, Cy3, and Cyanine5, Cy5) are simultaneously hybridized to one array for a period of time, and subsequently the excess labels are washed off and the glass is scanned under laser light producing a relative level for each RNA molecule [36, 37]. High density arrays (i.e Affymetrix, Santa Clara, CA) contain between 11 and 20 pairs (perfect match (PM) vs single mismatch (SM)) of 20-25mer oligonucleotide probes for a target RNA that are synthesized *in situ* by photolithography on silicon wafers. The oligonucleotides used as probes on the array are usually designed from nucleotide sequences or expressed sequence tags, ESTs, available from public databases (such as GenBank, UniGene and RefSeq) and often represent the most unique part of the sequence [38]. The SM probes are identical to the PM probes apart from a single nucleotide mismatch at the center position. RNA extracted from the biological sample is biotin labeled during the complementary RNA (cRNA) synthesis step, hybridized to the array and fluorescently detected through the streptavidin-phycoerythrin method. The average hybridization signal at each set of PM sequences provides a quantitative measure of the specific gene’s transcript. The reduced signals at each of the SM locations validate the specificity of the hybridization [36-38].

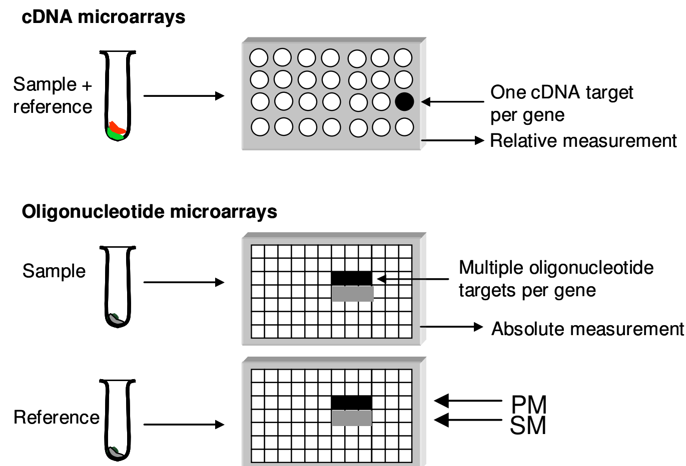

Figure 2. cDNA microarrays vs high density oligonucleotide arrays. Custom made cDNA microarrays and high density oligonucleotide arrays (i.e. Affymetrix) are the two most commonly used DNA microarray platforms. They differ mainly in the type of solid support the probes are printed on (glass and silicon, respectively) and the method of arraying. cDNA microarrays are in two channel format, i.e. two different cDNA samples labeled with different fluorophores (generally Cy3 and Cy5) are hybridized together to one array, competing for their complementary target and thus providing relative measurements. The probes are usually 50-70mer oligonucleotides resulting in high specificity of the hybridizations. High density oligonucleotide arrays are on the other hand in single channel format, where only one biotin labeled cRNA sample is hybridized to the array, resulting in an absolute measurement. The oligonucleotide arrays contain 11-20 pairs of 20-25mer oligonucleotides (perfect match (PM) and single mismatch (SM)), where the SM is identical to the PM except from a single nucleotide mismatch at the center position. The PM/SM design allows for validation of hybridization specificity [36-38].

Statistical analysis of gene expression data

A carefully chosen design at the beginning of a microarray experiment is a prerequisite for generating high quality data and to maximize the efficiency of the data analysis. One of the greatest challenges in microarray data analysis is to distinguish changes in gene expression specific for the cell type, from the noise and variability inherent within the microarray technique.

There is no standardized way to analyze the vast amount of data generated by a microarray experiment and thus the analytical method selected should be directed against the specific biological hypothesis tested. However, the fundamental steps in the data analysis can be divided into two categories; low level analysis and high level analysis. Hence, the data analysis starts with the low level analysis which includes image acquisition, image analysis (i.e. exclusion of poor quality spots, background correction etc), data-preprocessing (log-transformation of the data) and normalization of the data. The low level analysis is succeeded by the high level analysis involving statistical inference of differentially expressed genes, various exploratory data analysis, classification of samples and pathway analysis (Figure 3). [39].

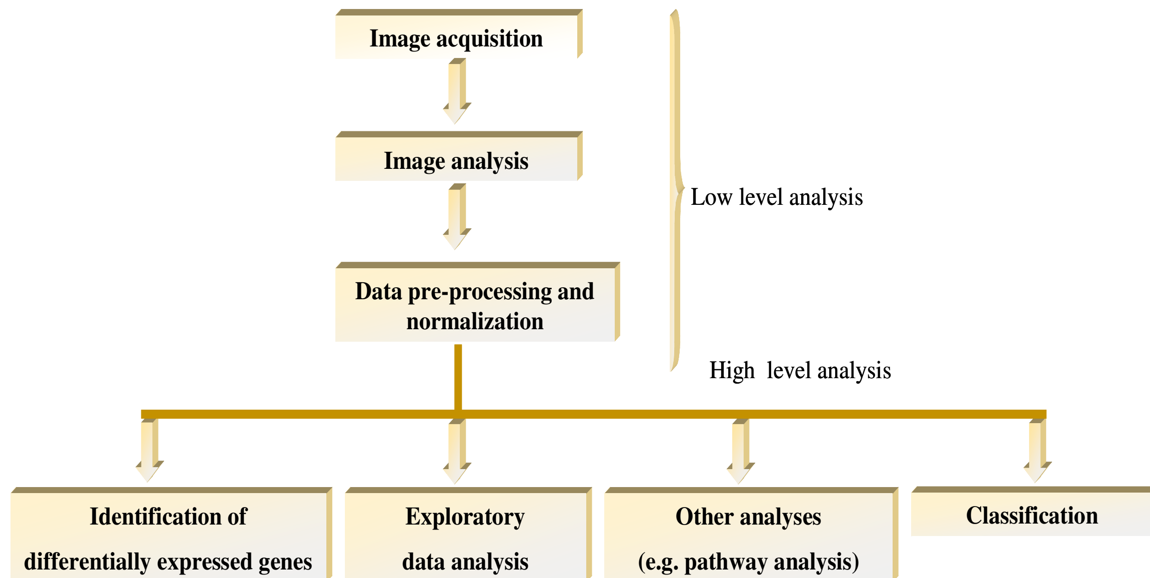

Figure 3. General work-flow of cDNA microarray data analysis. Low level analysis is the first step in the data analysis and involves image acquisition, image processing and data pre-processing and normalization. The low level analysis is followed by the high level analysis which includes identification of differentially expressed genes, exploratory data analysis, other analysis such as pathway analysis and classification. The selection of which high level analysis to perform should be directed to the biological hypothesis in the specific experiment. The figure is adapted from Leung and Cavalieri, (2003).

AIMS OF THE STUDY

The overall aim of this PhD project is to investigate the expression patterns in a rat model of human endometrial adenocarcinoma (EAC) by means of global gene expression profiling.

The specific objectives of this thesis are:

- to compare gene expression patterns between endometrial tumors and other cell types from the endometrium in order to identify genes and cellular pathways involved in EAC.
- to identify marker genes that might be used for diagnosis of human EAC.
- to confirm and evaluate potential candidate genes for EAC carcinogenesis.

MATERIALS AND METHODS

Experimental materials

Animal crosses and tumor material

Among virgin females of the BDII/Han inbred rat strain, more than 90% spontaneously develop EAC during their life span. The two other inbred rat strains (SPRD*Cu3/*Han and BN/Han) used in the crossing experiments rarely develop EAC. Crosses between BDII females and the non-susceptible BN or SPRD males were made to produce F1 progenies. The F1 progeny was subsequently backcrossed to BDII females to produce N1 progenies and by brother sister mating, F2 progenies were produced (Figure 4). Females in the N1, F1 and F2 progenies with suspected tumors were euthanized and tumors were surgically removed and subsequently subjected to pathological characterization. Normal tissue from liver was collected from the entire progeny for DNA extraction. Tumor tissues were collected for DNA extraction and cell culture establishment.

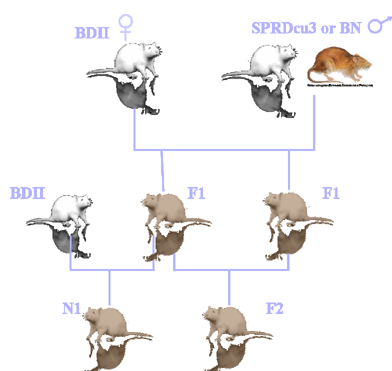

Figure 4. Animal crosses. Females of the EAC susceptible BDII rat strain were crossed with one non-susceptible strain (BN or SPRD, respectively) to produce an F1 offspring. The F1 progeny was subsequently backcrossed to BDII females to produce an N1 generation or intercrossed to produce an F2 intercross generation.

The tumors that developed in the N1, F1 and F2 progeny were pathologically classified as EAC, or other uterine tumors. In some cases, no cancer cells were detected in the removed cell mass when pathologically analyzed, and these are referred to normal/pre-malignant endometrium (NME). In the backcross progeny, the majority of the removed tissues were classified as NME, i.e. the tumors did not exhibit the morphological characteristics specific for EAC or other uterine tumors.

In this study, global gene expression analysis was performed on cDNA from 12 cell lines classified as NMEs, 26 endometrial tumor cell lines and from 7 cell lines classified as other tumors of the uterus/endometrium (Table 1).

Table 1. EAC tissue cultures used in the gene expression profiling experiments.

Tissue culture	Pathology	Genetic background	Tissue
RUT30	Anaplastic carcinoma	(BDIIxBN)F2	Uterus
RUT7	Endometrial adenocarcinoma	(BDIIxBN)F2	Endometrium
RUT12	Endometrial adenocarcinoma	(BDIIxBN)F2	Endometrium
RUT5	Endometrial squamous cell cancer	(BDIIxBN)F2	Endometrium
NUT114	Cervical cell squamous cell polyps	(BDIIxBN)xBDII	Cervix
NUT6	Endometrial adenocarcinoma	(BDIIxBN)xBDII	Endometrium
NUT31	Endometrial adenocarcinoma	(BDIIxBN)xBDII	Endometrium
NUT43	Endometrial adenocarcinoma	(BDIIxBN)xBDII	Endometrium
NUT46	Endometrial adenocarcinoma	(BDIIxBN)xBDII	Endometrium
NUT50	Endometrial adenocarcinoma	(BDIIxBN)xBDII	Endometrium
NUT51	Endometrial adenocarcinoma	(BDIIxBN)xBDII	Endometrium
NUT52	Endometrial adenocarcinoma	(BDIIxBN)xBDII	Endometrium
NUT81	Endometrial adenocarcinoma	(BDIIxBN)xBDII	Endometrium
NUT82	Endometrial adenocarcinoma	(BDIIxBN)xBDII	Endometrium
NUT97	Endometrial adenocarcinoma	(BDIIxBN)xBDII	Endometrium
NUT98	Endometrial adenocarcinoma	(BDIIxBN)xBDII	Endometrium
NUT99	Endometrial adenocarcinoma	(BDIIxBN)xBDII	Endometrium
NUT100	Endometrial adenocarcinoma	(BDIIxBN)xBDII	Endometrium
NUT127	Endometrial adenocarcinoma	(BDIIxBN)xBDII	Endometrium
NUT128	Endometrial adenocarcinoma	(BDIIxBN)xBDII	Endometrium
NUT37	Malignant uterus tumor	(BDIIxBN)xBDII	Uterus
NUT61	Malignant uterus tumor	(BDIIxBN)xBDII	Uterus
NUT48	Non-malignant endometrium	(BDIIxBN)xBDII	Endometrium
NUT75	Non-malignant endometrium	(BDIIxBN)xBDII	Endometrium
NUT110	Non-malignant endometrium	(BDIIxBN)xBDII	Endometrium
NUT118	Non-malignant endometrium	(BDIIxBN)xBDII	Endometrium
NUT122	Non-malignant endometrium	(BDIIxBN)xBDII	Endometrium
NUT123	Non-malignant endometrium	(BDIIxBN)xBDII	Endometrium
NUT129	Non-malignant endometrium	(BDIIxBN)xBDII	Endometrium
RUT2	Endometrial adenocarcinoma	(BDIIxSPRD-Cu3)F2	Endometrium
RUT13	Endometrial adenocarcinoma	(BDIIxSPRD-Cu3)F2	Endometrium
NUT7	Endometrial adenocarcinoma	(BDIIxSPRD-Cu3)xBDII	Endometrium
NUT12	Endometrial adenocarcinoma	(BDIIxSPRD-Cu3)xBDII	Endometrium
NUT39	Endometrial adenocarcinoma	(BDIIxSPRD-Cu3)xBDII	Endometrium
NUT41	Endometrial adenocarcinoma	(BDIIxSPRD-Cu3)xBDII	Endometrium
NUT42	Endometrial adenocarcinoma	(BDIIxSPRD-Cu3)xBDII	Endometrium
NUT47	Endometrial adenocarcinoma	(BDIIxSPRD-Cu3)xBDII	Endometrium
NUT84	Endometrial adenocarcinoma	(BDIIxSPRD-Cu3)xBDII	Endometrium
NUT15	Endometrial papillary adenoma	(BDIIxSPRD-Cu3)xBDII	Uterus
NUT1	Malignant uterus tumor	(BDIIxSPRD-Cu3)xBDII	Cervix
NUT18	Non-malignant endometrium	(BDIIxSPRD-Cu3)xBDII	Endometrium
NUT56	Non-malignant endometrium	(BDIIxSPRD-Cu3)xBDII	Endometrium
NUT58	Non-malignant endometrium	(BDIIxSPRD-Cu3)xBDII	Endometrium
NUT89	Non-malignant endometrium	(BDIIxSPRD-Cu3)xBDII	Endometrium
NUT91	Non-malignant endometrium	(BDIIxSPRD-Cu3)xBDII	Endometrium

* NUT designates tumors derived backcross progeny whereas RUT designates tumors derived from first generation and intercross progeny.

Human endometrial tumor material

In paper IV, we investigated the mRNA expression of *GPX3*, an EAC candidate gene identified in paper III, in human endometrial adenocarcinomas using real time quantitative PCR (qPCR). A total of 30 endometrial tumors (EACs) embedded in archival formalin fixed paraffin (FFPE) were used in the study. All samples were anonymous endometrial adenocarcinomas and as reference material, benign endometrial tissue and lung tissue was used. A pathologist marked the tumor area at the hematoxylin and eosin slide. Using a Tissue Micro Array-equipment (Pathology Devices), 3-4 cores (∅0.6mm) of tumor tissue was punched out from the paraffin block. Total RNA was then extracted from the paraffin block and used for the real time qPCR.

We compared the mRNA expression between grade I, II and III endometroid tumors. The grade of an endometrioid cancer is based on how much the cancer forms glands that look similar to the glands found in normal, healthy endometrium. Grade I tumors have more than 95% of the cancerous tissue forming glands, grade II tumors have between 50-94% of the cancerous tissue forming glands whereas grade III tumors have less than half of the cancerous tissue forming glands. The latter tumors tend to be more aggressive and carry a poorer prognosis than do low grade cancers.

Methods

Global gene expression profiling - cDNA microarray experiments

In this work, the two-channel cDNA microarray format was employed. The 18K (6000 clones in triplicates) rat 70mer oligonucleotide arrays used were printed at the Swegene DNA microarray resource center in Lund. Each probe in the probe set (Rat 70mer oligonucleotide set, ver 1.0, OPERON) were printed in triplicates at random positions on the arrays and thus serve as technical replicates within the array. The tumor samples used in these experiments served as biological replicates since they come from the same tumor phenotype, but from different individuals in cross progenies between inbred rat strains.

Design of the microarray experiments

As a reference for the microarray experiments, we have consistently used a specific Universal Rat Reference RNA (Stratagene), comprising a calibrated mix of defined cell lines from 14 different tissues for expression studies in rat cell lines, as a common reference for all hybridizations (Figure 5). In this manner, a high portion of expressed genes will be present on the chip and consequently, positive hybridization signals at each probe element are obtained, thus avoiding having small, near zero denominators in calculating ratios [40].

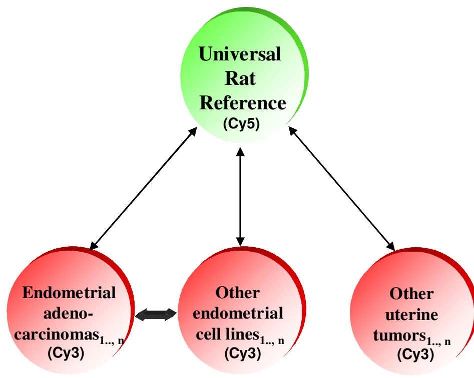

Figure 5. The reference design of the microarray experiments performed in this study. The use of a common reference allows comparisons of a large number of experiments performed under different time periods. In this work, the expression pattern in endometrial tumor cell lines was compared with the expression pattern in NME cell lines as well as with the expression pattern in cell lines of other uterine tumors. Thus, the universal reference serves as an internal control and allows for intra- and inter-comparisons among the different cell types.

Microarray hybridizations

The general procedure for the cDNA microarray hybridizations, involves extraction of total RNA from the biological samples under study. The RNA extracted from the biological samples and the universal rat reference RNA were subjected to reverse transcription to single stranded cDNA and simultaneously labeled with fluorophores (Cy3 and Cy5). The two differently labeled cDNA samples were pooled and subsequently hybridized to one array. The array was then scanned by a laser scanner, producing one image for each fluorophor (Figure 6).

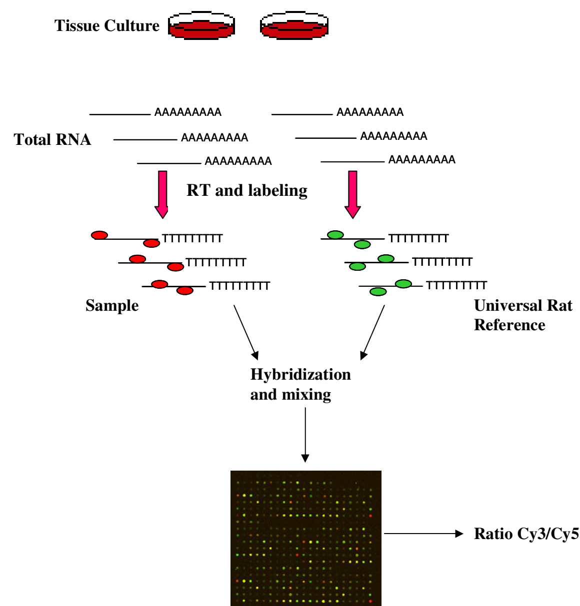

Figure 6. Workflow of the cDNA microarray experiments. The cDNA experiment begins with extraction of total RNA from the biological samples under study. The amount of total RNA required for a single hybridization ranges from 5-20 µg. The RNA extracted from the biological samples and the reference are subjected to reverse transcription to single stranded cDNA and simultaneously labeled with fluorophores (Cy3 and Cy5). The two differently labeled cDNA samples are pooled and subsequently hybridized to one array. The array is then scanned by a laser scanner, producing one image for each fluorophor. By merging the images, a ratio between the two channels can be calculated and the data is exported for various analyses.

The result files (.gpr files) generated from GenePix Pro 6.0 were imported into the R software environment (http://www.R-project.org) (where the R/limma package and limmaGUI were used for constructing diagnostic plots for evaluation of the data) and subsequently into BioArray Software Environment (BASE) for background correction, normalization, filtering, clustering analysis and significance analysis of expression changes. Bad quality spots (flagged as "bad" in GenePix) and low intensity spots were filtered out. The arrays were annotated to three classes: Endometrial adenocarcinoma, normal/pre-malignant lesions (NMEs) and other uterine tumors. The software used for constructing diagnostic plots is freely available from the Bioconductor project site http://www.bioconductor.org. The bioconductor packages limma, limmaGUI and arrayQuality applied in the present work, employ the free statistical programming environment R [41]. *MA* scatterplots (Figure 9) for all arrays were constructed in order to identify spot artifacts and to detect intensity-dependent patterns in the $\log_2$ ratios M, where M_{kj} ($\log_2$ ratio of background-adjusted intensities for gene k on array j) is plotted against A_{kj} (the average of the red and green channels with respect to background-adjusted intensities for gene k on array j). For each array k, and each gene j, M and A are calculated as follows [42, 43]:

$$M_{jk} = \log_2\left(\frac{R_{jk}}{G_{jk}}\right),$$

where R and G are the background-corrected red and green intensities for each spot and

$$A_{jk} = \frac{1}{2}(\log_2(R_{jk}) + \log_2(G_{jk})).$$

The *MA* scatterplots were also used for the purpose of deciding the normalization algorithm (within-slide) to use. Additionally, *MA* plots for all print-tip groups were constructed since there may exist systematic differences between the print tips, such as slight differences in the length or in the opening of the tips, and deformation after many hours of printing (Figure 10).

Image plots for all arrays were constructed by using the arrayQuality package in limma (Figure 7). M boxplots for all arrays (of raw data, within-slide normalized data and between-slides normalized data) were constructed in order to compare the distribution of M values within (print-tip groups) and between all arrays and to investigate whether scale-normalization (between-array normalization) was required (Figure 8).

Since the data showed dye bias and a skewed distribution of the M values, the intensity-dependent normalization method print-tip group loess (implemented in BASE) was performed. After normalization, density-, *MA*-, M box- and spatial plots were constructed from the normalized data to reinvestigate the distribution of M values (Figure 7-10). The idea behind the print-tip loess algorithm is that each M value is normalized by subtracting the corresponding value of the tip-group loess curve from the M value. The normalized log-ratios N are the residuals from the tip group loess regressions, i.e,

$$N_i = M_i - loess(A_i)$$

where $loess(A)$ is the loess curve as a function of A for the ith tip.

By performing a series of local regressions for each point in the scatterplot, a loess curve for each print tip group is constructed. Print-tip loess is considered to be the most suitable

normalization algorithm for cDNA microarrays and it is recommended to be used as a default method by many groups. It corrects the *M* values for spatial variation within the array and for intensity trends [42, 43].

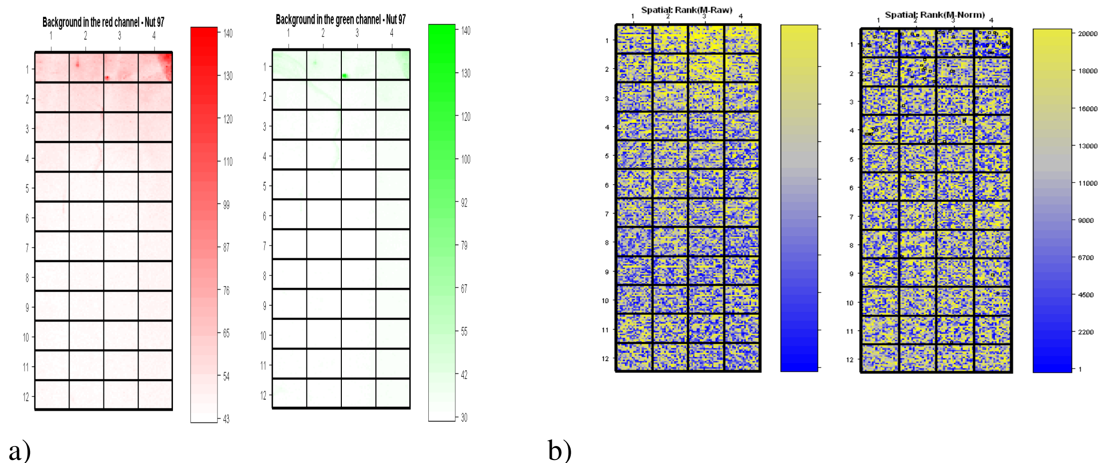

Figure 7. Image plots of an example data set/array (NUT97). The spatial image plots display the background in the green (Cy3) and red (Cy5) channel using a white-red and green-white color palette (a). A more intense background in the top right corner was observed. b) displays the distribution of spots ranked according the their *M* values, pre- and post print-tip loess normalization using a blue-yellow color palette. Higher ranks are denoted by blue and low ranks by yellow. The plots displaying log ratios *M* in the individual print-tip groups pre-normalization demonstrate an uneven distribution of low ranked spots in the top right corner. The print-tip loess normalization procedure resulted in a balanced distribution of high- and low ranked spots across both arrays.

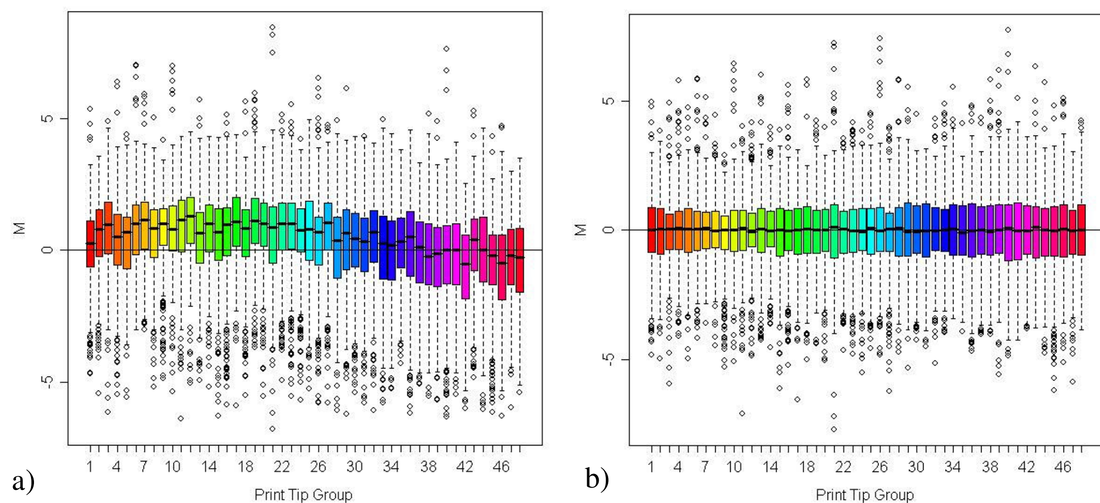

Figure 8. *M* boxplots of the individual print-tip groups pre- (a) and post-normalization (b). The x-axis denotes the individual print-tip groups and the y-axis denotes the log ratios, *M*. Variation in the spread in the log ratios of the individual print-tip groups on the array can be seen in a). The *M* values are skewed to $M<0$ in the last print-tip groups, suggesting the need for intensity and spatial based normalization. In b) the median of the *M* values are scaled to $M=0$ post print-tip loess normalization and the widths of the boxes are fairly consistent.

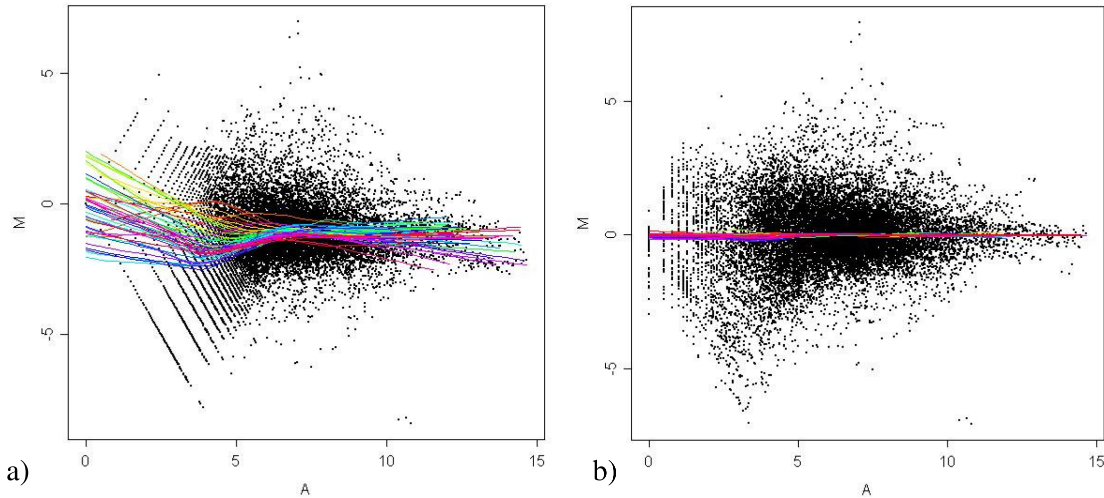

Figure 9. *MA* scatterplots. The loess curves for each print-tip group are represented with different colors. The *MA* scatterplot in a) demonstrates a non-linearity of the loess fits which indicates the need for within-array intensity- and location based normalization. Conducting print-tip loess normalization b) resulted in convergence of the loess curves and *M* values centered about *M*=0.

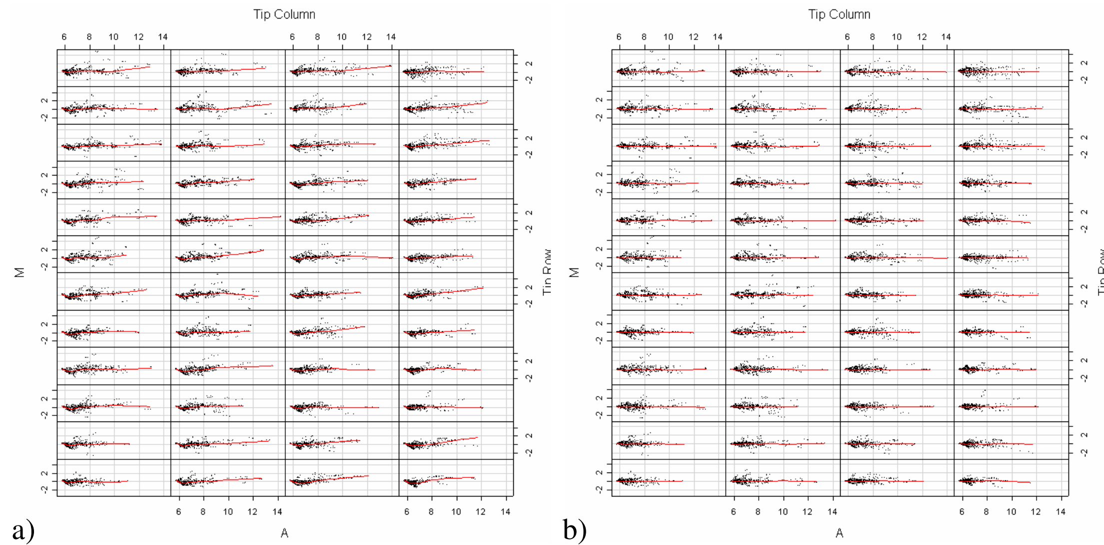

Figure 10. *MA* scatterplots for the individual print-tip groups pre- (a) and post print-tip loess normalization (b). The *MA* scatterplot of pre-normalized data for NUT97 clearly displays variation between the individual print-tip groups, which indicates that print-tip loess normalization could be worthwhile. The post print-tip loess normalization procedure resulted in scatterplots displaying *M* values centered around 0 across the range of *A* intensities.

The plug-in median/mean centering implemented in BASE was applied in order to scale the *M* values of the data. Centering median was used since it is more robust to outliers than centering based on the mean and it was performed for both genes and arrays. The number of centering cycles was set to 5.

Spots that were present on less than 30 arrays (of the total 45) were rejected. The data was also subjected to variation filtering, i.e. all position-reporter pairs with standard deviation (SD) smaller than 0.8 were rejected. After cleaning, filtering and normalization procedures, 4336 probes/reporters out of 6000 remained.

Exploratory data analysis - Hierarchical Clustering

For an initial exploratory analysis of the microarray data, the hierarchical clustering algorithm was applied to all sets of the data. Cluster analysis is a powerful method tool for reducing the complexity of the large amount of data generated in microarray experiments, which is the chief purpose with this algorithm. Hierarchical clustering is an agglomerative and unsupervised technique that in an iterative manner builds clusters of genes that share high similarity in the expression pattern and where the number of clusters is unspecified. This is accomplished by using a distance metric (also known as dissimilarity measure) that characterizes "the distance" between the expression patterns of, for example, different tumors [44, 45]. Genes with no difference in expression were filtered out prior to the cluster analysis. Euclidean distance was used for the array distance metric and the Pearson correlation coefficient was used for the gene distance metric. In this way, both genes and assays were clustered. Average linkage was used as the linking distance (the average of all pair-wise distances between members of the two clusters). Duplicate/triplicate reporters/spots were merged and averaged.

The clustering process does however not test for statistical validation and hence statistical inference analysis was applied for this purpose.

Statistical inference analysis of significantly differentially expressed genes

Significant differences in expression for reporters between the EAC cell lines and normal/pre-malignant endometrial cell lines were assessed by applying Wilcoxon Mann-Whitney test and a traditional student's *t* test, with a significance threshold of 0.05. Wilcoxon Mann-Whitney statistics is a non-parametric *t*-statistic computed, making no assumption that the data is normally distributed. It is computed by ranking the expression values of each gene across experiments from low to high, disregarding to which class each experiment (array) belongs [46, 47]. As multiple testing of thousands of genes usually generate a high proportion of false positives and false negatives, it is necessary to perform *P* value adjustments. Correction of the *P* value was therefore performed using the False Discovery Rate (FDR) procedure [48]. Applying FDR, the expected proportion of false positives among the rejected hypotheses is controlled. Other conventional methods for *P* value adjustements controlling the family wise error rates (FWER), such as Bonferroni, Holm's method and the Hochberg's method, are generally to stringent and resulting in an increase of false negatives and hence limit the power to identifying differentially expressed genes. The correlation among expression levels between different genes is not taken into considerations with the FWER approach [39, 47, 49, 50].

Classification analysis using Weka

In order to identify genes that might be used for discriminating between endometrial tumors and normal/pre-malignant endometrium, classification analysis using Waikato environment for knowledge analysis (Weka, version 3.4.12) was employed [51]. For each of the 29 samples, a "flag" (1 or 0) was set to signify group membership (cell lines from EAC tumors and non/pre-malignant lesions, respectively). The Weka software includes 70 different machine learning algorithms, each of which can be used to generate a classifier by learning from examples to distinguish between groups.

Gene functional classification

The web-accessible program, the Database for Annotation, Visualization and Integrated Discovery, DAVID, was used to obtain an overview of the gene functions of the 50 genes with the highest differential expression between endometrial tumors and normal/pre-malignant endometrium. DAVID provides tools for functional annotation of genes and gene functional classification, in which large lists of genes can be rapidly reduced into functionally related groups of genes to help unravel the biological content [52]. We wanted to investigate whether these genes were involved in pathways/processes contributing to the cancer phenotype (increased proliferation, increased apoptosis etc) and thus only cellular processes recognized as typical cancer hallmarks were selected.

Transmission Disequilibrium Test (TDT)

TDT statistics was performed on genotype data (Falck *et al.* manuscript in preparation) from microsatellite markers located adjacent to chromosomal regions harboring the identified classifiers and the top 50 genes with the most significant differential expression between endometrial tumors and non/pre-malignant endometrium. The TDT statistic is defined as $(H-A)^2/(H+A)$ [53], where H is the number of heterozygous animals and A is the number of animals homozygous for the BDII allele. Thus, in the TDT, the number of times that heterozygous parents pass one marker allele to the affected offspring is compared to the number of times affected offspring have received the other marker allele. The test has a χ^2 distribution with one degree of freedom. TDT statistics were calculated for markers adjacent to each gene in the EAC tumors versus non/pre-malignant lesions, and for differences between the two backgrounds, BDII/BN and BDII/SPRD, respectively.

Reverse Transcription PCR (RT-PCR) and real time quantitative PCR (qPCR)

The traditional, semi-quantitative reverse transcription PCR for investigating mRNA expression was used for verification of the genes identified as differentially expressed between the groups in the microarray experiments. Beta-actin (*Actb*) was used as an endogenous control for the PCR experiments and thus co-amplified in the reactions.

Real time RT-PCR, or real time quantitative PCR (qPCR), was performed to analyze the mRNA expression of *GPX3* in 30 human endometrial tumors and Glyceraldehyde 3-phosphate dehydrogenase (*GADPH*) was used as endogenous control. The advantage of using the real time RT-PCR, compared to traditional RT-PCR, is that it enables both detection and quantification (as absolute number of copies or relative amount when normalized to DNA input or additional normalizing genes) of a specific sequence in a DNA sample and is more sensitive when comparing expression levels between samples. In this study, we have employed the Taqman assay, which measures the accumulation of a product via a fluorophore during the exponential stages of the PCR, rather than at the end point as in the traditional PCR. The threshold cycle, C_T, i.e. the number of PCR cycles at which a significant exponential increase in fluorescence is detected, is determined by the exponential increase of the PCR product (Figure 11). The C_T value is directly correlated with the number of copies of DNA template in the reaction.

The comparative C_T method (ΔC_T) was used for assessing the relative changes in mRNA expression between the different groups investigated and is calculated as follows:

$$\frac{X_{test}}{X_{control}} = 2^{\Delta\Delta C_T} = 2^{(C_{T,X} - C_{T,R})_{control} - (C_{T,X} - C_{T,R})_{test}},$$

where CT,X is the threshold cycle of the gene of interest and CT,R is the threshold cycle of the endogenous reference gene (i.e. GAPDH). Test refers to the tumor cDNA sample and control refers to the calibrator cDNA sample.

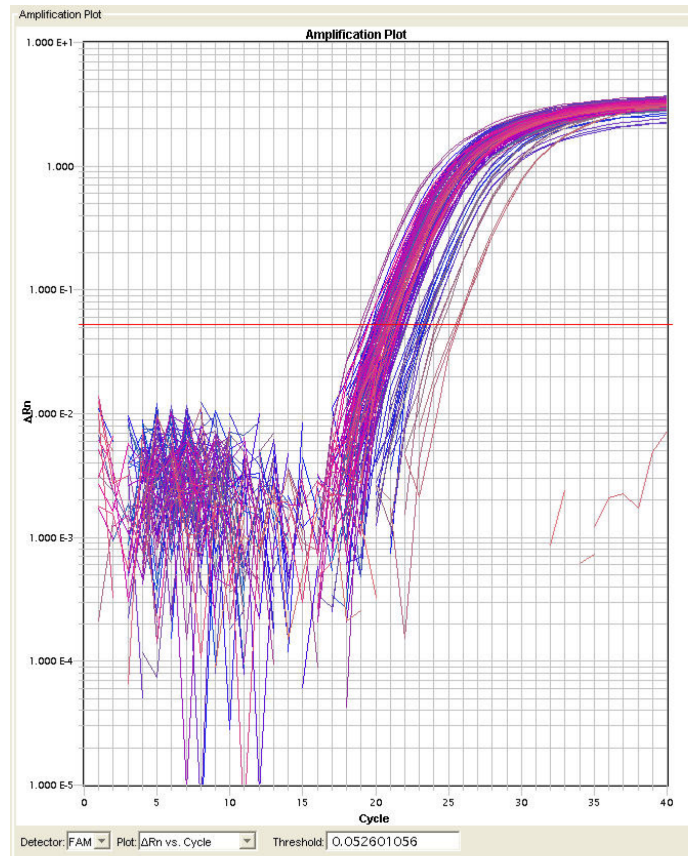

Figure 11. Real time qPCR amplification plot. The cycle number is plotted against the fluroscence emission. The red vertical line is the threshold line which is set in the exponential phase of the fluorescence emission curves.

Chromosome paint and dual-color Fluorescent In Situ Hybridization (FISH)

Chromosome paint and dual-color FISH were used to map breaks and deletions on rat chromosome 10. FISH is used to detect deletions or amplifications of specific genomic targets using probes that are labeled with fluorochromes (usually biotin or with targets for antibodies). The single-stranded probe is then applied to interphase or metaphase chromosome preparations and incubated for approximately 12 hours, while hybridizing. The results are visualized and quantified using a fluorescence microscope. There are different

types of probes, each of which has different applications. For detecting locus specific deletions, locus specific probes are used, whereas for examining chromosomal abnormalities, whole chromosome probes are used (also called whole chromosome paint). By using several overlapping probes, it is possible to detect breakpoints of translocations. The application of these techniques facilitates analysis of chromosomal aberrations and genetic abnormalities in various human diseases including cancer

Mutation screening by DNA sequencing

DNA sequencing for mutations and analysis of allelic imbalance (AI) was performed to investigate the status of *Tp53* and for mutation screening of *Myo1c* in endometrial tumors. Briefly, the DNA regions of interest were first amplified by PCR using genomic DNA and/or cDNA as template. The PCR products were subsequently purified and subjected to cycle sequencing using a fluorescent dye-labeled dideoxy procedure (BigDye^{TM}Terminator Cycle Sequencing Ready Reaction).

Methylation-Specific PCR (MSP)

Methylation-specific PCR (MSP) was used to investigate whether the loss of expression/down-regulation of the *Gpx3* gene was due to hypermethylation. Using the MSP method, the methylation status of virtually any group of CpG sites within a CpG island, independent of the use of cloning or methylation-sensitive restriction enzymes, can be assessed. The assay involves initial modification of DNA by sodium bisulfite, converting all unmethylated cytosines to uracil, and subsequent amplification with primers specific for methylated versus unmethylated DNA. The primers were designed using the freely available web-based program Methprimer, publicly available at http://www.urogene.org/methprimer/ (Figure 12).

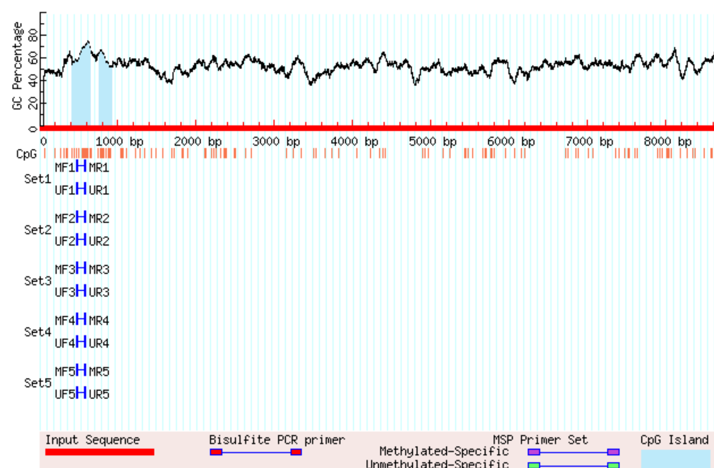

Figure 12. Results from Methprimer. Two CpG islands were found in the promoter region of *Gpx3*. The first primer sets generated from the Methprimer software were used for investigating the methylation status of the *Gpx3* promoter.

IDENTIFICATION AND ANALYSIS OF GENES INVOLVED IN EAC

Global gene expression analysis

Previous work with the BDII model system has been focused on finding chromosomal regions associated with susceptibility and development of EAC. By means of genome wide screens with microsatellite markers of tumors developed in BDII crosses, several chromosomal regions associated with susceptibility to EAC were identified. This infers that several susceptibility genes with minor, but co-operating, effects are responsible for the EAC susceptibility. In the SPRDCu3 background, three chromosomal regions (RNO1q35-36, RNO11q23 and RNO17p11-q11) were found to be associated with susceptibility for EAC, whereas in the BN background, there was only one region (RNO20p12). Thus, the onset of tumors depends not only on the presence of susceptibility alleles from the EAC-prone strain, but also on the contribution of genetic components derived from the non-susceptible strains [54, 55]. Previous investigations also include studies of chromosomal aberrations that occurred in tumor samples of the BDII crosses. By conducting Comparative Genome Hybridization (CGH) in tumor samples developed in the crosses, it could be concluded that certain chromosomal regions were recurrently engaged in increases or decreases in copy number (e.g. hyperploidy/ amplifications or hypoploidy/deletions) [56-58].

Thus, some genetic factors and chromosomal aberrations that might contribute to initiation and malignant progression into EAC are known. However, identifying causative cancer-related genes within recurrent genomic aberrations is not always uncomplicated since the affected regions often harbor several hundreds of genes and many of these might contribute to the malignant transformation. In order to find genes with aberrant expression and to identify expression profiles typical to EAC, thus elucidating crucial molecular events occurring during EAC development, we performed cDNA microarray experiments on a set of cell lines established from EAC tumors, normal/pre-malignant endometrium and other uterus tumors.

Identifying potential tumor suppressor gene candidates on RNO10 (paper I)

Rat chromosome 10 (RNO10) has been shown to be frequently involved in chromosomal aberrations in EAC. By means of cytogenetic studies and CGH analyses of the solid EAC tumors and cell lines, common deletions in the proximal part of RNO10 in EAC could be determined [59]. In additional allelic imbalance (AI) studies, three deleted sub-regions in the proximal region on RNO10 were identified in several separate EACs [60, 61]. One of the commonly deleted regions was located in the central part of the chromosome and since the *Tp53* gene is located within that region at the border between bands 10q24-q25, it was selected as a candidate gene for EAC tumorigenesis. The main aim of paper I was to investigate whether the *Tp53* gene is the molecular target of the frequent allelic losses in the region RNO10q24-q25. To determine the frequency and map position of the chromosome breaks along RNO10 presumably involved in the allelic losses, dual-color gene-specific fluorescent *in situ* hybridization (FISH) and chromosome paint analysis were performed. We also investigated the mutation status of *Tp53* in the tumor materials and combined the results of sequencing for gene mutations with the analysis of allelic imbalance results. For a more detailed deletion analysis, the FISH study was extended with eight additional probes evenly distributed between the *Tp53* gene and *Thra1*.

From the measurements described above, we could show recurrent breaks and losses distal to *Tp53* along with high frequency of allelic imbalance in RNO10q24-q25. 35% of the tumors with AI and/or deletions at RNO10q24-q25 did not show any mutations in the *Tp53* gene, and thus we concluded that RNO10q24-q25 may harbor another tumor suppressor gene with important implications in EAC development. We also concluded that this potential tumor suppressor gene is located close to and distal of the *Tp53* gene. Our conclusions is in accordance with earlier studies, since low frequencies of *Tp53* mutations accompanied by high frequency of LOH at HSA17p13.3 is common in a variety of human malignancies including breast, lung and hepatocellular carcinomas and neuronal tumors [62-69].

By comparing the rat and human DNA sequence, it became clear that several intra-chromosomal rearrangements have taken place in this region during the divergence between human and rat. Thus, the conservation with respect to gene order is not very extensive. By a reciprocal comparison of the positions of the genes located in the proximity of the break in the rat genome, with human gene positions and information from other human cancer studies, we could reduce the size of the suggested candidate tumor suppressor region to approximately 0.64 Mb. This genomic segment contains 18 genes (Table 3). In order to identify potential tumor suppressor candidate genes in this region, we investigated the mRNA expressions from microarray data. The microarray chips contained 6000 clones but most of these 18 genes were not represented. A preliminary analysis however, revealed that the endometrial tumor tissue cultures on average had 3.8 times lower expression of *Crk* in comparison to the normal/pre-malignant endometrial tissue cultures. This implied that this region might contain relevant candidate genes.

Table 3. Suggested tumor suppressor candidate region on RNO10q24

Position in rat sequence			**Position in human sequence**		
Start (Mb)	**Symbol**	**Cytoband**	**Start (Mb)**	**Symbol**	**Cytoband**
62.23	*Est1a*	10q24	1.91	*EST1A*	17p13.3
62.47	*Hic1*	10q24	1.91	*HIC1*	17p13.3
62.48	*Ovca2*	10q24	1.89	*OVCA2*	17p13.3
62.49	*Dph1*	10q24	1.88	*DPH1*	17p13.3
			1.86	*D17S831*	
62.50	*Rtn4rl1*	10q24	1.79	*RTN4RL1*	17p13.3
			1.76	*D17S1574*	
62.61	*Rpa1*	10q24	1.68	*RPA1*	17p13.3
62.66	*Smyd4*	10q24	1.63	*SMYD4*	17p13.3
62.71	*Serpinf1*	10q24	1.61	*SERPINF1*	17p13.1
62.75	*Serpinf2*	10q24	1.60	*SERPINF2*	17p13
62.81	*Prpf8*	10q24	1.50	*PRPF8*	17p13.3
62.83	*Rilp*	10q24	1.50	*RILP*	17p13.3
62.84	*Scarf1*	10q24	1.48	*SCARF1*	17p13.3
62.85	*Slc43a2*	10q24	1.42	*SLC43A2*	17p13.3
62.91	*Pitpn*	10q24	1.37	*PITPN*	17p13.3
62.95	*Skip*	10q24	1.34	*SKIP*	17p13.3
62.99	*Myo1c*	10q24	1.31	*MYO1C*	17p13
63.02	*Crk*	10q24	1.27	*CRK*	17p13.3
63.07	*Ywhae*	10q24	1.19	*YWHAE*	17p13.3

To conclude, the association between observed patterns of chromosomal deletions and AI at the proximal end of RNO10 and *Tp53* mutation was analyzed in 27 rat EAC cell lines. Together with similar observations in human, the present study provides further evidence for the presence of a putative tumor suppressor gene close to, but distal of the *Tp53* gene.

Further studies involve expression analysis of the 18 genes presented in Table 3 above, using qPCR. These results show a significant down-regulation of *Myo1c* and *Hic* in the EAC tumors, suggestive of two novel candidate genes for EAC carcinogenesis in RNO10q24 cytoband (Figure 13).

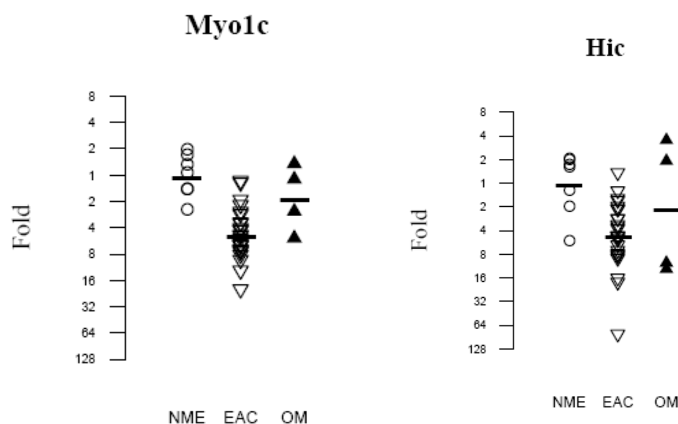

Figure 13. Scatterplots from the qPCR analysis. The relative expression of *Myo1c* and *Hic1* in EAC tumors, normal/pre-malignant endometrium (NME) and other uterine tumors (OM) is shown. Expression is presented as fold change of C_T (log_2).

Aberrant expression of genes associated with the TGF-β signaling pathway (paper II)

Since the cDNA microarray experiments generate a huge amount of expression data, it is desirable to get an initial overview of the expression profiles. Cluster analysis is one of the most commonly used techniques applied for identification of patterns in large scale gene expression data. Hierarchical clustering is an agglomerative and unsupervised technique that in an iterative manner builds clusters of genes, which share high similarity in expression patterns and where the number of clusters is unspecified [44, 45]. An initial hierarchical clustering of the expression data was performed to explore the expression profiles and to identify possible subgroups in the tumor material. This clustering analysis revealed two significant results; 1) a cluster of genes associated with the TGF-β pathway was found to be differentially expressed between the EAC samples and normal/pre-malignant samples. The mRNA expressions of several pro-collagens (*Col1a1*, *Col1a2*, *Col3a1*, *Col5a1*), *Ctgf*, *Ltbp2 and Tgfb1i1* were decreased in a majority of the endometrial tumors, 2) a few of the endometrial tumors clustered together with a majority of the normal/pre-malignant samples and a few of the normal/pre-malignant samples clustered together with the majority of the EACs.

Furthermore, we performed a separate hierarchical clustering on the TGF-β signaling pathway associated genes, and found three distinct clusters of endometrial samples; Group I) contained mainly normal/pre-malignant EACs and up-regulated TGF-β associated genes, Group II) contained mainly EACs and down-regulated TGF-β associated genes, and Group III) contained EACs, normal/pre-malignant endometrial samples and other uterine tumors, all

exhibiting a mixed expression pattern of the TGF-β associated genes. This implies the need of adjustment of the tumor classification of the endometrial tumors and tissues.

From these data, we concluded that the irregularities found in the TGF-β pathway among the majority of the EAC tumor cell lines, have substantial impact on EAC carcinogenesis. Our conclusions are supported by other studies that have demonstrated aberrant TGF-β signaling in human EAC, due to a decreased expression of the TGF-β receptors I and II at both transcriptional and translational levels. In addition, in these studies, a decreased protein expression of phosphorylated Smad2 was reported [70, 71]. Smad2 is an indicator of active TGF-β signaling and together with Co-Smads, it induces transcription of specific target genes, in both early and more progressed stages of EACs [72-74]. The decreased expression of Smad2 in both early and late stages, suggests a loss of TGF-β signaling at an early stage in EAC carcinogenesis. The TGF-β pathway associated genes found to be differentially expressed between EACs and normal/pre-malignant endometrium in this work, are directly or indirectly regulated by TGF-β signaling. In addition, some of the genes might activate TGF-β signaling (such as *CTGF*, *TGFB1I1 and*, *LTBP2*) and thus the down-regulated expression might result in a hampered TGF-β signaling. Our work is the first report presenting aberrations in expression of TGF-β down-stream target genes and we propose that the down-regulation of these genes might be due to abrogated TGF-β signaling.

The TGF-β superfamily of cytokines can be attributed to regulate a wide range of cellular processes, including cell growth, proliferation, differentiation, apoptosis and homeostasis in adult tissues (Figure 14) [72-74]. The growth suppressor effects exerted by the TGF-β family in normal cells makes the members of the TGF-β family potent tumor suppressors that cancer cells must escape for malignant evolution. Resistance to the TGF-β growth inhibitory effects in early tumorigenesis has been demonstrated in numerous studies on epithelial cancer cell lines [70, 75-80]. In tumor cells that are responsive to TGF-β signaling, TGF-β induces cell cycle arrest in late G1, apoptosis and expression of cell adhesion molecules, including fibronectin, laminin, CEA, collagen and integrins [75]. However, TGF-β signaling has dual implications in cancer, since it in some cancer diseases modulates processes such as cell invasion and immune regulation that cancer cells may make use of to their advantage. Consequently, the output of a TGF-β response is highly contextual in different types and stages of tumors [81].

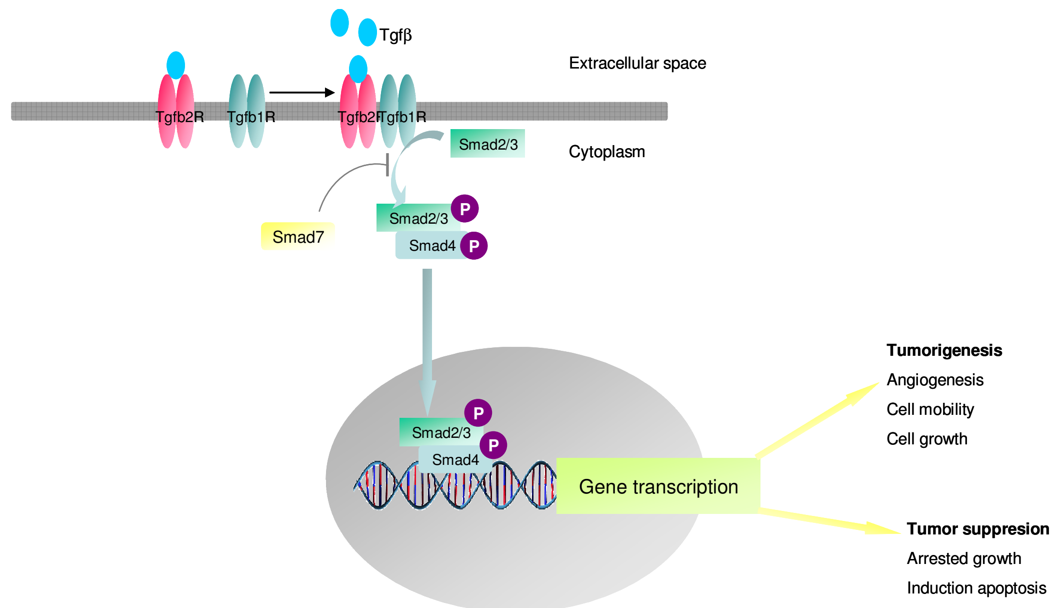

Figure 14. A schematic diagram of the TGF-β signaling pathway. The mechanisms for TGF-β signaling transduction involves the binding of TGFβ to TGFβ2R which induces phosphorylation and activation of the TGFβ1R. The activated TGFβ1R phosphorylates SMAD2 and SMAD3, which in turn form complex with SMAD4 mediator. The complex is translocated into the nucleus and regulates gene transcription. SMADs regulate transcription in several ways, including binding to DNA, interactions with other transcription factors, and transcription corepressors and coactivators, like p300 and CBP. The SMAD signaling cascade can be repressed by SMAD-7. The TGF-β pathway also interacts with other signaling pathways, such as the MAP kinase-ERK pathway, that modulate SMAD activation.

Another important finding in paper II, as previously mentioned, was that a few of the endometrial tumors shared similar expression pattern with the majority of the normal/pre-malignant cell lines and vice versa. This was shown both in the clustering analysis on the global expression data, as well as in the clustering analysis of the expression of the TGF-β signaling associated genes. The expression of the TGF-β associated genes was up-regulated in the normal/pre-malignant cell types, referred to as group I and down-regulated in the endometrial tumors, referred to as group II. We therefore concluded that these samples needed to be re-classified, which was also supported by the observations made from the *in vitro* cell culturing. The EAC cell lines that clustered together with the normal/pre-malignant cell lines were more similar to the typical non-malignant cell lines when cultured, and the non-malignant cell lines that clustered to group II, displayed similar morphological and physiological characteristics (i.e. increased growth rate, irregular shapes etc) as the fully malignant EAC cell lines. We have therefore chosen to focus our continued analysis on the two groups representing fully malignant of EAC and normal/pre-malignant respectively.

A three-gene signature of EAC (paper III)

The main objective of paper III was to identify potential marker genes and molecular signatures that correlate with early and late stages of endometrial cancer by applying the classification tool, Waikato environment for knowledge analysis, (Weka, version 3.4.12) on the microarray expression data and conducting a traditional Student's *t* test and Mann-Whitney test for assessing genes with differential expression between the two groups (EAC

cell lines and normal/pre-malignant endometrial cell lines). Even when applying *P* value adjustments (False Discovery Rate, FDR [48]), we found a high number of genes (354) that were significantly differentially expressed between the two groups. In Figure 15, we present a dendrogram from hierarchical clustering of the 200 genes that had the most significant differential expression between EACs and normal/pre-malignant samples. As shown, the majority of the differentially expressed genes was generally down-regulated in the EAC group whereas the majority of the genes was more or less normally expressed in the normal/pre-malignant samples, implying that these genes are dysfunctional in EAC.

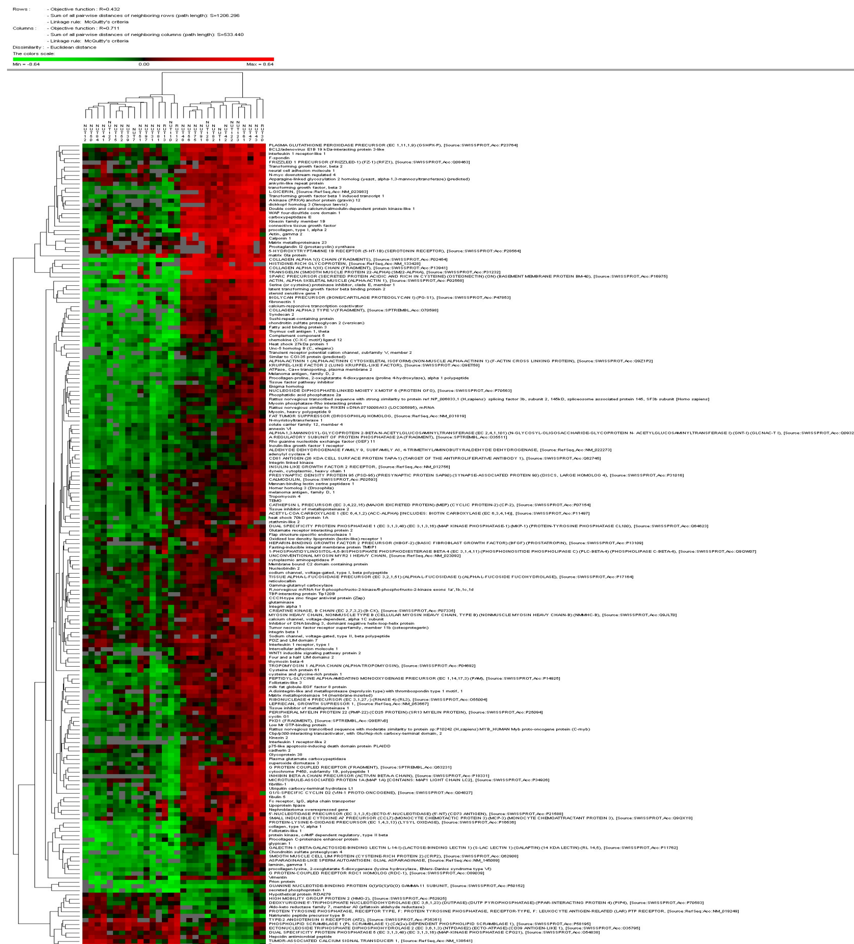

Figure 15. Dendrogram of the top 200 genes with the highest differential expression. The resulting dendrogram from the hierarchical clustering analysis of the top 200 genes with highest differential expression between EAC cell lines and normal/pre-malignant cell lines.

Since the number of differentially expressed genes identified was relatively high, we wanted to unravel whether some of them were involved in cellular processes known to be frequently disturbed in cancerous cells. By employing the freely available gene function clustering tool, the Database for Annotation, Visualization and Integrated Discovery, DAVID [52], of the 50 genes with the highest differential expression, we found that 30 were involved in cellular processes typically aberrant in tumors (i.e. differentiation, proliferation, apoptosis, mobility etc). Moreover, several of these genes were involved in more than one process, which reflects the complexity of cancer etiology.

To explore whether a specific gene expression signature for EAC could be generated, we applied the Weka classification tool. The Weka software includes 70 different machine learning algorithms, each of which, by learning from examples, can be used to generate a classifier which distinguish groups from each other [51]. We excluded all algorithms that required discrete-valued input data set, since these are not applicable to real-valued gene expression data set. The fold change values from the microarray data were imported into Weka. In total, 29 samples from group I and II were analyzed, and the group membership was signified. The two genes that were most frequently identified as top genes in the classifiers were *Gpx3* (nine classifiers) and *Bgn* (seven classifiers). These two genes were also among the highest ranked among the 50 genes with highest differential expression. In total, 16 algorithms derived classifiers could determine the most important gene for classification, whereas 26 algorithms used many or all genes for the classification. The majority of the algorithms reached accuracies of 97% (a single misclassification) or 100%. We also applied Weka using the original histopathological classification, and found a significant low accuracy rate of classification. This confirms our previous re-classification of the two groups.

The gene with the highest differential expression was plasma glutathione peroxidase 3, *Gpx3*, and thus considered as a highly interesting candidate gene for EAC carcinogenesis. GPX3 exhibits a critical role in detoxifying reactive oxidative species and maintaining the genetic integrity of mammalian cells. *Gpx3* has been found to be either deleted or highly methylated in exon 1 in prostate cancer cell lines [82, 83] and in Barret's tumorigenesis [84] and has been suggested to exhibit tumor suppressor activity [82]. The tumor suppressor activity of *GPX3* is thought to be associated with its ability to repress the expression of the *MET* oncogene. Expression data of *Met* in the present work implies that samples exhibiting an up-regulation of *Gpx3* also show a low expression of *Met*.

Biglycan (BGN) belongs to the family of small leucine-rich proteoglycans [85] and is functionally involved in matrix assembly, cellular migration, adhesion, and the regulation of growth factors. Several studies have shown that gene expression of *Bgn* is regulated by TGF-β signaling [86, 87]. This result is in accordance with our previous hypothesis of a disruptive TGF-β signaling, since we have shown down-regulation of several genes controlled by the TGF-β signaling pathway. Furthermore, it has been shown that exogenously administered BGN, induced pancreatic cancer cells to arrest in the G1 phase of the cell cycle, indicating a direct inhibiting effect on proliferation in cancerous cells and thus exhibits tumor suppressive activities [88].

Previous work with the BDII inbred rat model included identification of susceptibility regions by means of genome wide screening with microsatellite markers on the female cross progenies that developed EAC [54, 55]. When we examined the genotyping data in the chromosomal regions harboring the 50 genes identified as the most differentially expressed between EACs and normal/pre-malignant endometrium, we found only one gene (*Tgfb3*) to be located in a susceptibility region (RNO6q31). Regardless of genetic background, we found a significant χ^2 value (5.26) from the TDT analysis ($p<0.05$). This means that *Tgfb3* is a possible susceptibility gene candidate and implies that the BDII strain might harbor a single nucleotide polymorphism, SNP, involved in the initiation of EAC. However, the expression data only correlated with grade of tumors, i.e. the expression of *Tgfb3* was down-regulated in EACs and up-regulated in normal/pre-malignant cell lines regardless of the genetic background. Since the TGFB3 protein is involved in multiple cellular processes, the effects of genetic variation on mRNA expression may vary during tumor progression due to interactions

with other affected genes. Thus, if the hypothesized genetic change in *Tgfb3* is inherited, the effect on the expression level might be different in early and late stages of EAC.

The mRNA expression of *Tgfb3* has been suggested to be stimulated by Collagen 1 through the PI3K/ERK pathway in lung adenocarcinomas, [89]. The decreased expression of *Tgfb3* shown here, might be due to the decreased expression of several pro-collagens as we previously have demonstrated [90]. Interestingly, it has been shown that TGFB3 expression is up-regulated in human EAC and that TGFβ3 exerts promoting effects on invasiveness [91], which is contradictive to our results. However, Van Themsche and co-workers [91] used two commercial cell lines (i.e. KLE and HEC-1A) originating from elderly women (64 and 71, respectively). These cell lines may represent the more invasive and non-estrogen dependent type II endometrial carcinoma, whereas our model represents estrogen-dependent endometrial carcinoma. This indicates that TGF-β signaling in the more aggressive type II endometrial carcinoma has a tumor promoting effects, whereas it in the estrogen-dependent type I endometrial carcinoma seem to have a tumor inhibiting role [70, 71, 92]. Further studies are needed to clarify the implications of down-regulated expression of *Tgfb3* in rat EAC.

In conclusion, we were able to present a three-gene signature that can be utilized as a starting point for establishing a panel of endometrial cancer biomarkers. In addition, the data set also provides potential candidate genes that should be further evaluated for possible implications in the carcinogenesis of EAC.

Evaluation of genes identified from the microarray expression studies

Confirmations of the gene candidates identified from the analyses of the microarray experiments were performed using the traditional semi-quantitative RT-PCR. All gene expressions investigated could be verified and rendered almost identical results when comparing the two methods.

Epigenetic mechanisms responsible for loss of *Gpx3* mRNA expression (paper IV)

In paper III, we identified Glutathione peroxidase 3 (*Gpx3*) as a potential molecular marker of rat EAC by using the Waikato Environment for Knowledge Analysis (Weka) classification tool on the microarray expression data. In paper IV, we performed RT-PCR on a larger selection of endometrial tumors and found that the expression of *Gpx3* was indeed down-regulated in all tumors. GPX3 is one of three key enzymes, along with catalase (CAT) and superoxide dismutase (SOD), involved in the cellular defence against reactive oxidative species (ROS). SOD catalyzes the dismutation of the superoxide anion (O_2-•) into hydrogen peroxide and GPX3 in turn reduces hydrogen peroxide (and other organic hydroperoxides) to water using reduced glutathione (GSH) as the electron donor (Figure 16). SOD and GPX3 are present in the cytosol and in the mitochondria. CAT, on the other hand, reduces hydrogen peroxide in the peroxisomes [93, 94]. It has been shown that cancer cells are constantly under oxidative stress and that they produce ROS at a much higher rate than do non-transformed cell lines [95]. ROS are known to cause severe damages on DNA, by introducing strand breaks of the DNA, altering guanine and thymine bases, and by inducing sister chromatide exchanges. The ROS-induced damage to cells also includes alteration of macromolecules such as polyunsaturated fatty acids in membrane lipids and essential proteins. The consequences might be inactivation of tumor suppressor genes or increased expression of proto-oncogenes.

The persistent oxidative stress on carcinoma cells may eventually lead to genetic instability which in turn increases the malignant potential of the tumor [96].

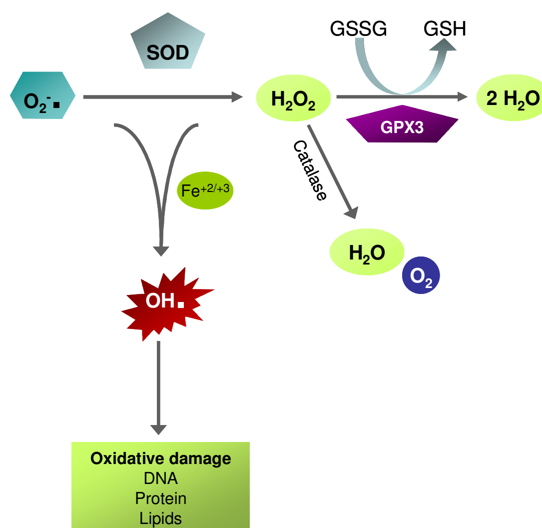

Figure 16. Oxidative stress pathway. Glutathione peroxidase 3 (GPX3), superoxide dismutase (SOD) and catalase (CAT) constitute the backbone in the cellular defence against reactive oxidative species. SOD and GPX3 are present in the cytosol and the mitochondria and reduces superoxide and hydrogen peroxide, respectively. GPX3 catalyses the reduction of hydrogen peroxide using reduced glutathione (GSH) as the electron donor. CAT, on the other hand, reduces hydrogen peroxide in the peroxisomes.

As epigenetic silencing by hypermethylation of the *GPX3* promoter region was previously shown in prostate cancer [82, 83] and in Barret's disease [84], we wanted to examine whether the highly down-regulated expression of *Gpx3* in the endometrial carcinoma samples was due to hypermethylation. We therefore performed methylation-specific PCR on available DNA from tumor cell lines (n=12) and found that 91% of the tumors displayed hypermethylation of the *Gpx3* promoter, and that 9 out of 10 tumors displayed biallelic methylation (i.e. both alleles are methylated). This finding suggests that the GPX3 function is impaired in rat EAC, and a likely consequence is an increased amount of hydrogen peroxide and other reactive oxidative species in the endometrial tumor microenvironment.

Yu *et al* also found that GPX3 might harbor tumor suppressor activities, since they could show that induced over-expression of GPX3 in prostate cancer cell lines decreased invasiveness, anchorage independent growth and colony formation [82]. Moreover, xenografted prostate cancer cells expressing GPX3, showed reduction of tumor size, elimination of metastasis and reduction of animal death. Yu and co-workers findings also suggest that the GPX3 tumor suppressor activity involves transcriptional regulation by the tumor transforming gene Mesenchymal-Epithelial Transition factor (*MET),* since the induction of GPX3 expression decreased the mRNA expression of *MET* [82]. Engagement of MET induces multiple signaling transduction pathways that regulate cellular processes, such as growth promotion, motility and invasiveness (Figure 17) [97-99]. We therefore investigated whether we could find a correlation between the expression of *Gpx3* and *Met* in the BDII rat model by performing a semi-quantitative RT-PCR of *Met* as well. We found that the expression of *Met* was slightly higher in the endometrial tumors that had a loss of *Gpx3* expression. When applying a two-way ANOVA on the RT-PCR expression data, we found a highly significant ($P < 0.0005$) interaction between the *Met* and *Gpx3* expression. Thus,

besides introducing mutagenic events by potential increased levels of H_2O_2, this finding implies that the down-regulation of *Gpx3* presented herein, might result in tumor promoting effects by up-regulating the transcription of the *Met* oncogene.

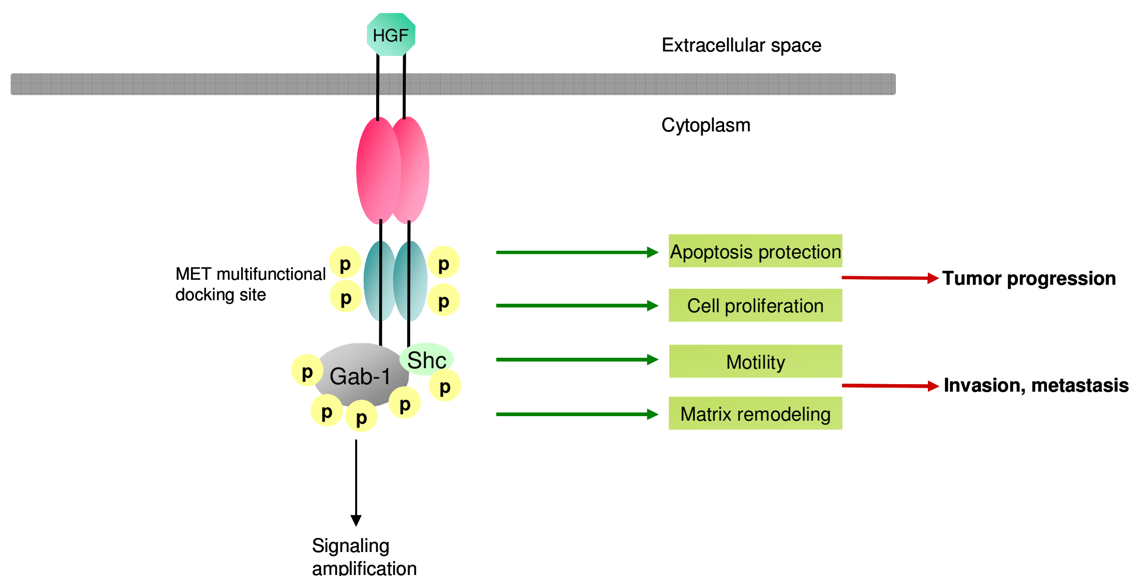

Figure 17. Oncogenic MET signaling transduction. MET engagement by the binding of hepatocyte growth factor (HGF) to the MET receptor, results in the activation of several signaling transduction pathways that regulates tumor promoting cellular processes such as proliferation, motility and invasiveness [97-99].

The transcription of *MET* involves the binding of hypoxia inducible factor 1 (HIF-1) and AP-1, to hypoxia responsible elements present in the *MET* promoter region. In addition, it has been proposed that overproduction of ROS is required for the hypoxic activation of *HIF-1* and thus transcription of *MET* [100]. A possible consequence of the observed down-regulation of *GPX3* presented in this work might be a higher production of ROS. In a preliminary study, we measured the production of hydrogen peroxide in an endometrial tumor cell line with loss of expression of *Gpx3* and in two non-malignant endometrial cell lines with *Gpx3* expression. The results indicate a higher production of hydrogen peroxide in EAC. Hence, the co-occurrence of loss of expression of *GPX3* and induction of *MET* expression suggest the possibility that the GPX3 activates MET transcription through increased levels of ROS and thus the induction of HIF-1. Clearly, further functional analyses to elucidate the role of GPX3 and MET in EAC are required.

Expression of *GPX3* in human endometrial tumors (paper IV)

As mentioned previously, animal models are exceptionally valuable when analyzing genetic components involved in the pathogenesis of human complex disorders. In the end, however, it is important to be able to translate the findings from the animal studies to humans to develop new diagnosis tools and/or for therapeutic interventions. We strongly believe that the aberrant mRNA expression of *Gpx3* identified in rat endometrial adenocarcinoma have important implications in the human endometrial tumor etiology as well. Thus, in order to evaluate the results from the BDII rat model, we measured the mRNA expressions of *GPX3* in 30 human

grade I-III EACs. We found that the expression of *GPX3* was down-regulated in all endometrial tumors, compared to two benign endometrial samples. We could not find any differences in expression between the tumor stages. This finding implies that down-regulated expression of *GPX3* is a universal event occurring at an early stage in EAC. However, since only two benign samples from the human endometrium were used as controls, these results can not be inferred as statistically significant with certainty. The results do however support the conclusions made from the BDII rat model and should be further evaluated.

In conclusion, we propose that the silencing of *GPX3* by DNA hypermethylation of cytosines in CpG islands located in the promoter region may augment the susceptibility of the endometrium to oxidative injury and subsequent carcinogenic transformation. Thus *GPX3* should be considered an important marker gene for human EAC.

CONCLUDING REMARKS

Cancer is a complex disease caused by complex patterns of interactions of hundreds of genes. Also, the influence of the genetic background/heterogeneity contributes to the complexity of tumor etiology. Our still relatively limited knowledge of how tumors develop comes from the studies of single genes or chromosome regions. However, the availability of the human genome information has rapidly increased the development of high-throughput techniques making it possibly to perform studies on thousands of genes at the same time. cDNA microarrays provide a powerful and versatile tool for the identification of differences in the expression profiles of tumor cells in comparison to their normal counterpart, and thus offers an increased understanding of the process of tumor etiology and development. Other important implications of cDNA microarrays in cancer research includes classification of tumors to different grades, identification of tumor subgroups and the discovery of novel diagnostic and prognostic biomarkers.

The main aim of the present thesis work was to investigate the expression patterns in rat EAC by the use of global gene expression profiling. The expression patterns of total 26 EAC cell lines, 12 normal/pre-malignant cell lines and 7 other uterine tumors were investigated by conducting two-channel cDNA microarrays.

The findings from the studies that the thesis is based are summarized below:

I. **Identification of a candidate tumor suppressor region on RNO10.** In this study, we found that the tumor suppressor gene, *Tp53*, is not probably the only target of the chromosomal breaks and deletions demonstrated at RNO10q24-25. We propose a new candidate tumor suppressor gene region of approximately 0.64 Mb. This genomic segment contains 18 potential tumor suppressor genes. Preliminary analysis of the microarray data showed that the endometrial tumors on average had 3.8 times lower expression of *Crk* in comparison to the normal/pre-malignant endometrial tissue cultures. This implied that this region might contain relevant candidate genes.

.

II. **Identification of an expression profile.** An initial hierarchical clustering analysis of the microarray data aiming to investigate the expression profiles between EACs and normal/pre-malignant cell lines, revealed a cluster of genes regulated by the TGF-β signaling pathway that were differentially expressed between the EACs and normal/pre-malignant samples. These results indicated a disruptive TGF-β signaling in the rat EAC cell lines. Previous studies on human endometrial tumors suggest loss of TGF-β signaling at an early stage of EAC, thus supporting our results. However, our findings provide the first report on aberrant expression of TGF-β down-stream target genes.

III. **Classification of the tumors.** We were able to make minor adjustment of the initial pathological classification of the EAC and normal/pre-malignant samples, based on the expression profiles from the microarray data and the observations made on cell morphology and physiology when the cells were cultured *in vitro*.

IV. **Identification of EAC biomarkers.** By applying the Weka classification tool on the microarray data, we identified a two-gene signature of EAC (*Gpx3* and *Bgn*). In addition, a TDT analysis on the genomic locations of 50 genes with the highest

differential expression revealed that only one gene, *Tgfb3*, was located in a susceptibility region. Thus, *Tgfb3* become the third part of a three-gene EAC signature.

V. **Differentially expressed genes between tumors and normal/pre-malignant counterparts.** Statistical inference analysis of the microarray data resulted in 354 genes with differential expression between EACs and normal/pre-malignant samples. A gene functional annotation tool, DAVID, was used to find out the function of 50 genes with the highest differential expression. With this analysis we were able to pin-point 30 possible cancer-related genes among the 50 differentially expressed genes. The observation that several of these genes were involved in many cellular processes, typically aberrant in cancer, reflects the complexity of cancer etiology.

VI. **Verification of the microarray expression data analysis by RT-PCR**. All genes that were identified as being differentially expressed from the microarray data analysis were confirmed by the traditional semi-quantitative RT-PCR, which in most cases rendered identical expressions compared to the microarray data. This confirms the accuracy of the data produced from the microarray experiments and the results from the statistical analyses applied.

VII. **Loss of expression of *Gpx3* in EAC is correlated with epigenetic silencing.** Evaluation of *Gpx3* down-regulation in the rat EAC cell lines revealed an almost complete loss of expression in a majority of the endometrial tumors. From the methylation studies, we could conclude that the loss of expression of *Gpx3* is correlated with biallelic hypermethylation in the *Gpx3* promoter region. This result was confirmed with a demethylation study of EAC cell lines, where the *Gpx3* mRNA expression was restored after treatment with a demethylation agent and a deacetylation inhibitor. We also showed that mRNA expression of the well-known oncogene, *Met*, was slightly higher in endometrial tumors with loss of *Gpx3* expression. A likely consequence of loss of *Gpx3* function is a higher amount of ROS in the cancer cell environment. Since it has been proposed that overproduction of ROS is required for the hypoxic activation of *HIF*-1, we suggest that loss of *Gpx3* expression activates transcription of *Met* through induction of the transcription factor *Hif-1*. The loss of the protective properties of GPX3 most likely makes the endometrial cells more vulnerable to ROS damage and genome instability.

VIII. ***GPX3* is down-regulated in early and late stages of human EAC**. Since we wanted to extend the results obtained from the rat endometrial tumors to human material, expression analysis of *GPX3* was performed in 30 endometrial human tumors using qPCR. The results showed a uniformly down-regulation of *GPX3* in all tumors, except one, independent of tumor grade. We thus concluded that the down-regulation of GPX3 probably occurs at an early stage of EAC. These results suggest that there are important clinical implications of *GPX3* expression in EAC, both as a biomarker for EAC and as a potential target for therapeutic interventions.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to all people who has helped, supported and encouraged me while working with this project. In particular, I would like to thank:

Karin Klinga Levan, my supervisor, for giving me the opportunity to work within the intriguing field of cancer genetics, for always believing in my capability to work with this project and giving me lots of independence, for your guidance and excellent help when revising my thesis and for always being such a warm and wonderful person.

Björn Olsson, my co-supervisor, for scientific advice in the data analysis and proof-reading my thesis.

Professor **Göran Levan** and all other previous and present people working in the BDII rat model, my special thanks go to: **Afrouz, Carola**, **Fredrik and Lisbeth.**

All the people at Swegene DNA microarray Resource Center in Lund. My special thanks go to **Jeanette Valchich** and **Johan Staaf** who helped me with the hybridizations and scanning of the microarrays.

Gisela Helenius and **Mats Karlsson (and all other people involved in the human EAC project)** at Örebro University hospital, for our newly, but fruitful, collaboration in endometrial cancer research.

My colleagues in the biomedicine group: **Kajsa**, **Dennis**, **Fredrik, Viktoria**, **Rose-Lill**, **Janne**, **Eva**, **Ingela**, **Lena, Hong** and **Cecilia,** for creating such a warm and friendly atmosphere. My warmest thanks to: **Eva**, my Phd colleague, my "room mate" for fun chats and letting me share problems and frustrations, **Kajsa**, my former "room mate", for always being a support in the lab and for fun chats. **Jessica**, our new PhD student, my former exam- and present project worker, for helping me with the last experiments and for being patient when I have stressed out during the period of writing this thesis, **Jonathan**, for just being such a nice person and **Dennis**, for valuable, pleasant and inspiring conversations regarding both science and pleasure and for just for being such as great person.

All former and present colleagues at the molecular biology and ecology, especially: **Sazzad**, **Linda**, **Andreas**, **Diana**, **Magnus**, **Peter**, **Elie**, **Henric, Helene, Mikael, Tjowe, Emma, Åsa, Patrik, Simon, Niclas, Noel, Jenny** and **Anna-Karin**. My warmest thanks go to **Maria**, for being a wonderful friend, a former great tennis partner and colleague, **Erik** for our fun and philosophical chats about life in general, and **Håkan**, for introducing me to tennis and for being such an inspiring and excellent tennis coach.

Lars, for the pleasant, but hard, work camps at Od and for sharing the "spikningen av avhandlingen" with me. The process of writing our thesis's at the same time has truly, in one way, been a shared journey.

All my work-unrelated friends for keeping me in touch with "the real world", especially: **Sara and Maurus**.

Nixon, **Birgit**, **Tess**, **Mirri, Misse, Oskar** and **King**: A great thank for making my everyday life rich and full of fur. My special thanks goes to **Gustav** who constantly has kept me company during late nights work and **Vicky** who has had enormous appeasing effects on me during the stressful period of thesis writing.

Mette, for bringing great fun to tuesdays and weekends and for constantly making sure that I work hard 😉

My beloved **family**, my mother, my father, my brothers (and brother's in law), for constantly believing in me, your support and for always being there for me.

Nova and **Wilma**, for being such lovely and beautiful small creatures enriching my life...

Last, but definitely not the least: **Fredrik**: Three words only – I love you!

This study was financially supported by the National Research School in Bioinformatics and Genomics, The Nilsson-Ehle Foundation and the Erik Philip-Sörensen foundation.

REFERENCES

1. **Cancer** [http://www.who.int/mediacentre/factsheets/fs297/en/index.html]
2. Knudson AG: **Cancer genetics**. *Am J Med Genet* 2002, **111**(1):96-102.
3. Lynch HT, Fusaro RM, Lynch JF: **Cancer genetics in the new era of molecular biology**. *Annals of the New York Acadamy of Science* 1997, **833**:1-28.
4. Peters J, Loud J, Dimond E, Jenkins J: **Cancer genetics fundamentals**. *Cancer Nurs* 2001, **24**(6):446-461; quiz 462.
5. Ponder BA: **Cancer genetics**. *Nature* 2001, **411**:336-341.
6. Vogelstein B, Kinzler KW: **The multistep nature of cancer**. *Trends Genet* 1993, **9**(4):138-141.
7. Renan MJ: **How many mutations are required for tumorigenesis? Implications from human cancer data**. *Mol Carcinog* 1993, **7**(3):139-146.
8. Hanahan D, Weinberg RA: **The hallmarks of cancer**. *Cell* 2000, **100**(1):57-70.
9. Pharoah PD, Dunning AM, Ponder BA, Easton DF: **Association studies for finding cancer-susceptibility genetic variants**. *Nat Rev Cancer* 2004, **4**(11):850-860.
10. Frank SA: **Genetic predisposition to cancer - insights from population genetics**. *Nat Rev Genet* 2004, **5**(10):764-772.
11. Demant P: **Cancer susceptibility in the mouse: genetics, biology and implications for human cancer**. *Nature Reviews Genetics* 2003, **4**(9):721-734.
12. Balmain A: **Cancer as a complex genetic trait: tumor susceptibility in humans and mouse models**. *Cell* 2002, **108**(2):145-152.
13. Todd R, Wong DT: **Oncogenes**. *Anticancer Research* 1999, **19**(6A):4729-4746.
14. Weinberg RA: **Oncogenes, tumor suppressor genes, and cell transformation: Trying to put it all together**. In: *Origins of human Cancer – A comprehensive review.* Edited by Brugge J, Curran T, Harlow E, McCormick F. New York: Cold Spring Harbor Laboratory Press; 1991: 1-16.
15. Knudson Jr AG: **Mutation and cancer: statistical study of retinoblastoma**. *Proceedings of the National Academy of Sciences of The United States of America* 1971, **68**(4):820-823.
16. Knudson AG: **Hereditary cancer: two hits revisited**. *Journal of Cancer Research and Clinical Oncology* 1996, **122**(3):135-140.
17. Sherr CJ: **Principles of tumor suppression**. *Cell* 2004, **116**(2):235-246.
18. Schwab M: **Genetic principles of tumor suppression**. *Biochim Biophys Acta* 1989, **989**(1):49-64.
19. Hansen R, Oren M: **p53; from inductive signal to cellular effect**. *Curr Opin Genet Dev* 1997, **7**(1):46-51.
20. Levine AJ: **p53, the cellular gatekeeper for growth and division**. *Cell* 1997, **88**(3):323-331.
21. Guimaraes DP, Hainaut P: **TP53: a key gene in human cancer**. *Biochimie* 2002, **84**(1):83-93.
22. Friedberg EC: **DNA damage and repair**. *Nature* 2003, **421**(6921):436-440.
23. Purdie DM, Green AC: **Epidemiology of endometrial cancer**. *Best Pract Res Clin Obstet Gynaecol* 2001, **15**(3):341-354.
24. Amant F, Moerman P, Neven P, Timmerman D, Van Limbergen E, Vergote I: **Endometrial cancer**. *Lancet* 2005, **366**(9484):491-505.
25. Bokhman. J, V: **Two pathogenetic types of endometrial carcinoma**. *Gynecologic Oncology* 1983, **15**:10-17.

26. Prat J: **Prognostic parameters of endometrial carcinoma**. *Hum Pathol* 2004, **35**(6):649-662.
27. Deligdisch L, Holinka CF: **Endometrial carcinoma: two diseases?** *Cancer Detection and Prevention* 1987, **10**(3-4):237-246.
28. Aitman TJ, Critser JK, Cuppen E, Dominiczak A, Fernandez-Suarez XM, Flint J, Gauguier D, Geurts AM, Gould M, Harris PC *et al*: **Progress and prospects in rat genetics: a community view**. *Nat Genet* 2008, **40**(5):516-522.
29. James MR, Lindpaintner K: **Why map the rat?** *Trends in Genetics* 1997, **13**(5):171-173.
30. Gill TJ, 3rd, Smith GJ, Wissler RW, Kunz HW: **The rat as an experimental animal**. *Science* 1989, **245**(4915):269-276.
31. Greenhouse DD, Festing MFW, Hasan S, Cohen AL: **Catalogue of inbred strains of rats**. In: *Genetic Monitoring of Inbred Strains of Rats.* Edited by Hedrich HJ. Stuttgart-New York: Gustav Fischer Verlag; 1990: 410-480.
32. Huang H, Winter EE, Wang H, Weinstock KG, Xing H, Goodstadt L, Stenson PD, Cooper DN, Smith D, Alba MM *et al*: **Evolutionary conservation and selection of human disease gene orthologs in the rat and mouse genomes**. *Genome Biol* 2004, **5**(7):R47.
33. Abbott A: **Laboratory animals: the Renaissance rat**. *Nature* 2004, **428**(6982):464-466.
34. Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, Steffen D, Worley KC, Burch PE *et al*: **Genome sequence of the Brown Norway rat yields insights into mammalian evolution**. *Nature* 2004, **428**(6982):493-521.
35. Vollmer G: **Endometrial cancer: experimental models useful for studies on molecular aspects of endometrial cancer and carcinogenesis**. *Endocr Relat Cancer* 2003, **10**(1):23-42.
36. Luo J, Isaacs WB, Trent JM, Duggan DJ: **Looking beyond morphology: cancer gene expression profiling using DNA microarrays**. *Cancer Invest* 2003, **21**(6):937-949.
37. Butte A: **The use and analysis of microarray data**. *Nat Rev Drug Discov* 2002, **1**(12):951-960.
38. Schulze A, Downward J: **Navigating gene expression using microarrays--a technology review**. *Nat Cell Biol* 2001, **3**(8):E190-195.
39. Leung YF, Cavalieri D: **Fundamentals of cDNA microarray data analysis**. *Trends Genet* 2003, **19**(11):649-659.
40. Novoradovskaya N, Whitfield ML, Basehore LS, Novoradovsky A, Pesich R, Usary J, Karaca M, Wong WK, Aprelikova O, Fero M *et al*: **Universal Reference RNA as a standard for microarray experiments**. *BMC Genomics* 2004, **5**(1):20.
41. Dudoit S, Yang YH, Bolstad B: **Using R for the analysis of DNA microarray data**. *R news* 2002, **2**(1):24-32.
42. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP: **Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation**. *Nucleic Acids Res* 2002, **30**(4):e15.
43. Smyth GK, Speed T: **Normalization of cDNA microarray data**. *Methods* 2003, **31**(4):265-273.
44. Satagopan JM, Panageas KS: **A statistical perspective on gene expression data analysis**. *Stat Med* 2003, **22**(3):481-499.
45. Quackenbush J: **Computational analysis of microarray data**. *Nat Rev Genet* 2001, **2**(6):418-427.
46. Wilcoxon F: **Individual comparisons by ranking methods** *Biometrics* 1945, **1**:80-83.

47. Speed T: **Statitistical Analysis of gene expression microarray data**: Chapman & Hall/CRC; 2003.
48. Benjamini Y, and Hochberg, Y.: **Controlling the false discovery rate: a practical and powerful approach to multiple testing**. *Journal of the Royal Statistical Society Series B* 1995, **57**:289-300.
49. Nadon R, Shoemaker C: **Statistical issues with microarrays: processing and analysis**. *Trends Genet* 2002, **18**(5):265-271.
50. Slonim DK: **From patterns to pathways: gene expression data analysis comes of age**. *Nature genetics supplement* 2002, **32**:502-508.
51. Frank E, Hall M, Trigg L, Holmes G, Witten IH: **Data mining in bioinformatics using Weka**. *Bioinformatics* 2004, **20**(15):2479-2481.
52. Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA: **DAVID: Database for Annotation, Visualization, and Integrated Discovery**. *Genome Biol* 2003, **4**(5):P3.
53. Spielman RS, McGinnis RE, Ewens WJ: **Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM)**. *American Journal of Human Genetics* 1993, **52**(3):506-516.
54. Roshani L, Wedekind D, Szpirer J, Taib Z, Szpirer C, Beckmann B, Riviere M, Hedrich HJ, Klinga-Levan K: **Genetic identification of multiple susceptibility genes involved in the development of endometrial carcinoma in a rat model**. *Int J Cancer* 2001, **94**(6):795-799.
55. Roshani L, Mallon P, Sjostrand E, Wedekind D, Szpirer J, Szpirer C, Hedrich HJ, Klinga-Levan K: **Genetic analysis of susceptibility to endometrial adenocarcinoma in the BDII rat model**. *Cancer Genet Cytogenet* 2005, **158**(2):137-141.
56. Helou K, Walentinsson A, Beckmann B, Johansson Å, Hedrich HJ, Szpirer C, Klinga-Levan K, Levan G: **Analysis of genetic changes in rat endometrial carcinomas by means of comparative genome hybridization**. *Cancer Genetics and Cytogenetics* 2001, **127**(2):118-127.
57. Hamta A, Adamovic T, Helou K, Levan G: **Cytogenetic aberrations in spontaneous endometrial adenocarcinomas in the BDII rat model as revealed by chromosome banding and comparative genome hybridization**. *Cancer Genetics and Cytogenetics* 2005, **159**:123-128.
58. Helou K, Walentinsson A, Beckman B, Samuelson E, Hedrich HJ, Szpirer C, Klinga-Levan K, Levan G: **Comparative genome hybridzation (CGH) analysis in rat uterine endometrial carcinoma**. *Rat Genome* 2000, **6**:78.
59. Helou K, Walentinsson, A., Beckmann, B., Johansson, Å., Hedrich, HJ., Szpirer C., Klinga-Levan, K., Levan, G: **Analysis of genetic changes in rat endometrial carcinomas by means of comparative genome hybridization**. *Cancer Genet Cytogenet* 2001, **172**(2):118-127.
60. Behboudi A, Levan G, Hedrich HJ, Klinga-Levan K: **High-density marker loss of heterozygosity analysis of rat chromosome 10 in endometrial adenocarcinoma**. *Genes Chromosomes Cancer* 2001, **32**(4):330-341.
61. Nordlander C, Behboudi A, Levan G, Klinga-Levan K: **Allelic imbalance on chromosome 10 in rat endometrial adenocarcinomas**. *Cancer Genetics and Cytogenetics* 2005, **156**:158-166.
62. Coles C, Thompson AM, Elder PA, Cohen BB, Mackenzie IM, Cranston G, Chetty U, Mackay J, Macdonald M, Nakamura Y *et al*: **Evidence implicating at least two genes on chromosome 17p in breast carcinogenesis**. *Lancet* 1990, **336**(8718):761-763.

63. Guan XY, Sham JS, Tai LS, Fang Y, Li H, Liang Q: **Evidence for another tumor suppressor gene at 17p13.3 distal to TP53 in hepatocellular carcinoma**. *Cancer Genet Cytogenet* 2003, **140**(1):45-48.
64. Jung HL, Wang KC, Kim SK, Sung KW, Koo HH, Shin HY, Ahn HS, Shin HJ, Cho BK: **Loss of heterozygosity analysis of chromosome 17p13.1-13.3 and its correlation with clinical outcome in medulloblastomas**. *J Neurooncol* 2004, **67**(1-2):41-46.
65. Konishi H, Sugiyama M, Mizuno K, Saito H, Yatabe Y, Takahashi T, Osada H: **Detailed characterization of a homozygously deleted region corresponding to a candidate tumor suppressor locus at distal 17p13.3 in human lung cancer**. *Oncogene* 2003, **22**(12):1892-1905.
66. Phillips NJ, Ziegler MR, Radford DM, Fair KL, Steinbrueck T, Xynos FP, Donis-Keller H: **Allelic deletion on chromosome 17p13.3 in early ovarian cancer**. *Cancer Res* 1996, **56**(3):606-611.
67. Roncuzzi L, Brognara I, Baiocchi D, Amadori D, Gasperi-Campani A: **Loss of heterozygosity at 17p13.3-ter, distal to TP53, correlates with negative hormonal phenotype in sporadic breast cancer**. *Oncol Rep* 2005, **14**(2):471-474.
68. Sarkar C, Chattopadhyay P, Ralte AM, Mahapatra AK, Sinha S: **Loss of heterozygosity of a locus in the chromosomal region 17p13.3 is associated with increased cell proliferation in astrocytic tumors**. *Cancer Genet Cytogenet* 2003, **144**(2):156-164.
69. Zhao X, He M, Wan D, Ye Y, He Y, Han L, Guo M, Huang Y, Qin W, Wang MW *et al*: **The minimum LOH region defined on chromosome 17p13.3 in human hepatocellular carcinoma with gene content analysis**. *Cancer Lett* 2003, **190**(2):221-232.
70. Gold LI, Parekh TV: **Loss of growth regulation by transforming growth factor-beta (TGF-beta) in human cancers: studies on endometrial carcinoma**. *Semin Reprod Endocrinol* 1999, **17**(1):73-92.
71. Parekh TV, Gama P, Wen X, Demopoulos R, Munger JS, Carcangiu ML, Reiss M, Gold LI: **Transforming growth factor beta signaling is disabled early in human endometrial carcinogenesis concomitant with loss of growth inhibition**. *Cancer Res* 2002, **62**(10):2778-2790.
72. Massague J: **TGF-beta signal transduction**. *Annu Rev Biochem* 1998, **67**:753-791.
73. Massague J: **How cells read TGF-beta signals**. *Nat Rev Mol Cell Biol* 2000, **1**(3):169-178.
74. Shi Y, Massague J: **Mechanisms of TGF-beta signaling from cell membrane to the nucleus**. *Cell* 2003, **113**(6):685-700.
75. Markowitz SD, Roberts AB: **Tumor suppressor activity of the TGF-beta pathway in human cancers**. *Cytokine Growth Factor Rev* 1996, **7**(1):93-102.
76. Massague J, Blain SW, Lo RS: **TGFbeta signaling in growth control, cancer, and heritable disorders**. *Cell* 2000, **103**(2):295-309.
77. Blobe CG, Schiemann, W.P. & Lodish, H.F: **Role of transforming growth factor beta in human disease**. *The New England Journal of Medicine* 2000, **342**(18):1350-1358.
78. Levy L, Hill CS: **Alterations in components of the TGF-beta superfamily signaling pathways in human cancer**. *Cytokine Growth Factor Rev* 2006, **17**(1-2):41-58.
79. Bierie B, Moses HL: **TGF-beta and cancer**. *Cytokine Growth Factor Rev* 2006, **17**(1-2):29-40.
80. Jakowlew SB: **Transforming growth factor-beta in cancer and metastasis**. *Cancer Metastasis Rev* 2006.

81. Derynck R, Akhurst RJ, Balmain A: **TGF-beta signaling in tumor suppression and cancer progression**. *Nat Genet* 2001, **29**(2):117-129.
82. Yu YP, Yu G, Tseng G, Cieply K, Nelson J, Defrances M, Zarnegar R, Michalopoulos G, Luo JH: **Glutathione peroxidase 3, deleted or methylated in prostate cancer, suppresses prostate cancer growth and metastasis**. *Cancer Res* 2007, **67**(17):8043-8050.
83. Lodygin D, Epanchintsev A, Menssen A, Diebold J, Hermeking H: **Functional epigenomics identifies genes frequently silenced in prostate cancer**. *Cancer Res* 2005, **65**(10):4218-4227.
84. Lee OJ, Schneider-Stock R, McChesney PA, Kuester D, Roessner A, Vieth M, Moskaluk CA, El-Rifai W: **Hypermethylation and loss of expression of glutathione peroxidase-3 in Barrett's tumorigenesis**. *Neoplasia* 2005, **7**(9):854-861.
85. Iozzo RV, Murdoch AD: **Proteoglycans of the extracellular environment: clues from the gene and protein side offer novel perspectives in molecular diversity and function**. *FASEB J* 1996, **10**(5):598-614.
86. Wadhwa S, Embree MC, Bi Y, Young MF: **Regulation, regulatory activities, and function of biglycan**. *Crit Rev Eukaryot Gene Expr* 2004, **14**(4):301-315.
87. Kinsella MG, Bressler SL, Wight TN: **The regulated synthesis of versican, decorin, and biglycan: extracellular matrix proteoglycans that influence cellular phenotype**. *Crit Rev Eukaryot Gene Expr* 2004, **14**(3):203-234.
88. Weber CK, Sommer G, Michl P, Fensterer H, Weimer M, Gansauge F, Leder G, Adler G, Gress TM: **Biglycan is overexpressed in pancreatic cancer and induces G1-arrest in pancreatic cancer cell lines**. *Gastroenterology* 2001, **121**(3):657-667.
89. Shintani Y, Maeda M, Chaika N, Johnson KR, Wheelock MJ: **Collagen I promotes epithelial-to-mesenchymal transition in lung cancer cells via transforming growth factor-beta signaling**. *Am J Respir Cell Mol Biol* 2008, **38**(1):95-104.
90. Karlsson S, Holmberg E, Askerlund A, Levan KK: **Altered transforming growth factor-beta pathway expression pattern in rat endometrial cancer**. *Cancer Genet Cytogenet* 2007, **177**(1):43-50.
91. Van Themsche C, Mathieu I, Parent S, Asselin E: **Transforming growth factor-beta3 increases the invasiveness of endometrial carcinoma cells through phosphatidylinositol 3-kinase-dependent up-regulation of X-linked inhibitor of apoptosis and protein kinase c-dependent induction of matrix metalloproteinase-9**. *J Biol Chem* 2007, **282**(7):4794-4802.
92. Sakaguchi J, Kyo S, Kanaya T, Maida Y, Hashimoto M, Nakamura M, Yamada K, Inoue M: **Aberrant expression and mutations of TGF-beta receptor type II gene in endometrial cancer**. *Gynecol Oncol* 2005, **98**(3):427-433.
93. Klaunig JE, Kamendulis LM: **The role of oxidative stress in carcinogenesis**. *Annu Rev Pharmacol Toxicol* 2004, **44**:239-267.
94. Yu BP: **Cellular defenses against damage from reactive oxygen species**. *Physiol Rev* 1994, **74**(1):139-162.
95. Szatrowski TP, Nathan CF: **Production of large amounts of hydrogen peroxide by human tumor cells**. *Cancer Res* 1991, **51**(3):794-798.
96. Wiseman H, Halliwell B: **Damage to DNA by reactive oxygen and nitrogen species: role in inflammatory disease and progression to cancer**. *Biochem J* 1996, **313 (Pt 1)**:17-29.
97. Bottaro DP, Rubin JS, Faletto DL, Chan AM-L, Kmiecik TE, Vande Woude GF, Aaronson SA: **Identification of the hepatocyte growth factor receptor as the c-met proto-oncogene product**. *Science* 1991, **251**:802-804.

98. Gentile A, Trusolino L, Comoglio PM: **The Met tyrosine kinase receptor in development and cancer**. *Cancer Metastasis Rev* 2008, **27**(1):85-94.
99. Peruzzi B, Bottaro DP: **Targeting the c-Met signaling pathway in cancer**. *Clin Cancer Res* 2006, **12**(12):3657-3660.
100. Chandel NS, Maltepe E, Goldwasser E, Mathieu CE, Simon MC, Schumacker PT: **Mitochondrial reactive oxygen species trigger hypoxia-induced transcription**. *Proc Natl Acad Sci U S A* 1998, **95**(20):11715-11720.