The testicle, the breast and systems biology

Belling, Kirstine González-Izarzugaza; Brunak, Søren; Jensen, Thomas Skøt; Nielsen, Henrik Bjørn; Gupta, Ramneek

Publication date:
2011

Document Version
Publisher's PDF, also known as Version of record

The testicle, the breast and systems biology

PhD thesis

Kirstine Christensen Belling

August 31st 2011

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

Preface

This thesis was prepared between September 1st 2008 and August 31st 2011 at the Center for Biological Sequence Analysis (CBS), Department of Systems Biology, Technical University of Denmark under supervision of Professor Søren Brunak, previous employee Thomas Skøt Jensen, Associate Professor H. Bjørn Nielsen and Associate Professor Ramneek Gupta. The PhD was funded by the Villum Kann Rasmussen Foundation and fulfill the requirements for acquiring a PhD degree.

The thesis is based on a combination of data analysis and experimental work carried out primarily at CBS in collaboration with Department for Growth and Reproduction, Rigshospitalet, University Hospital of Copenhagen and in collaboration with the Sino-Danish Breast Cancer Research Centre at Faculty of Life Sciences, University of Copenhagen. The work for the project presented in chapter 7 was performed during my six months external stay with Professor Richard Sharpe's group at MRC Human Reproductive Sciences Unit, Centre for Reproductive Biology, The Queens Medical Research Institute, Edinburgh, UK.

It has been a great pleasure to work as a PhD student at CBS. My special thanks to my supervisors, it has been a great pleasure and very rewarding to collaborate with you. My thanks to the people in the Integrative Systems Biology group and also thanks to Bent, David, Rachita, Simon and Tejal. I would like to thank the girls in the CBS administration: Annette, Dorthe, Lone, Louise, Marlene, Sara, Stine and Ulla. It is always a pleasure to walk into your office and you are always very helpful. My thanks to the guys in Systems Administration, Kristoffer Rapacki, Peter Wad Sackett, Hans-Henrik Stærfeldt and John Damm Sørensen, for being very helpful and patience with a person who does not have a computational background.

My thanks to the people at Department for Growth and Reproduction for a great and close collaboration and especially thanks to Henrik Leffers and Marlene Dalgaard who have been very helpful in all projects. Also, my thanks to the people at the Sino-Danish Breast Cancer Research Centre for a good collaboration. A big thanks to Professor Richard Sharpe and his scientific group for including me in the group and for making making me a part of some interesting projects and letting me "play" in the laboratory. A very special thanks to Afshan for being my very close friend during my six months in Edinburgh in 2010.

I would like to give a big thanks to my friends for whom I am truly grateful. Last but definitely not least, I am very thankful to my parents for their continuous love and support.

Kirstine C. Belling

Contents

Abstract

This PhD thesis presents work carried out primarily at Center for Biological Sequence Analysis, Technical University of Denmark, but also during my six months external stay at MRC Human Reproductive Sciences Unit, Centre for Reproductive Biology, The Queens Medical Research Institute, Edinburgh, UK. The projects presented are a combination of experimental work, data analysis and integrative systems biology. Included in this thesis are five projects that focus on the biology in fetal development, the testis and in breast cancer.

The first project described in chapter 6 illustrates an integrative systems biology project on Octamer-binding transcription factor 4 (OCT4) in embryonic stem cell (ESC) pluripotency and differentiation. Based on integration of ChIP-chip data, silencing data, and gene expression data, we found a highly interconnected protein-protein interaction network of OCT4-regulated targets that appear important for ESC differentiation.

The following three chapters in the thesis present projects that study testis biology. Chapter 7 presents a pure experimental study that investigates the importance of androgen receptor (AR) signaling in peritubular myoid cells for proper Leydig cell development and function, which is essential for testosterone and sperm production. Chapter 8 describes a study of adult spermatogenesis and cellular and transcriptional changes after low dose irradiation. The study is a combination of experimental work and gene expression profiling. By clustering analysis of the gene expression data, we found five clusters each representing separate cells in the testis. We compared the length of the differentiation stages with the same stages in postnatal spermatogenesis and found they had the same length. We also identified Leydig cell changes after irradiation, which might impact sperm production in testis cancer patients after radiotherapy. Chapter 9 presents a study of the microRNA (miRNA) profile in carcinoma *in situ* (CIS) cells, which are the precursor cells for testicular germ cell tumors (TGCTs). We found miRNA expression similarities with fetal gonocytes from which CIS cells arise because of developmental arrest. We also found some of the same miRNAs expressed in TGCTs and the miRNA profile identified in CIS cells is properly important for testis carcinogenesis.

The last project included in this thesis (chapter 10) presents a project on a potential marker of endocrine resistance, Tissue metallopeptidase inhibitor-1 (TIMP-1). The study combines experimental and analytical approaches and identified the progesterone receptor to be important in the increase resistance to antiestrogens in high TIMP-1 expressing hormone receptor-positive breast cancer.

Taken together, this thesis presents five projects that have contributed to the understanding of fetal development, testis biology and breast cancer. The projects show that the combination of systems biology and experimental work is a strong approach that can help elucidate complex biology and help to understand diseases and improve treatment in the future.

Resumé

Denne PhD afhandling præsenterer arbejde primært udført på Center for Biologisk Sekvenanalyse, Danmarks Tekniske Universitet, samt under mit 6 måneders eksterne ophold på MRC Human Reproductive Sciences Unit, Centre for Reproductive Biology, The Queens Medical Research Institute, Edinburgh, Storbritanien. Projekterne er en kombination af eksperimentelt arbejde, dataanalyse og integrativ systembiologi. Der er fem projekter inkluderet i denne afhandling, som fokuserer på biologien i fosterudviklingen, testiklerne og brystcancer.

Det første projekt (kapitel 6) illustrerer et integrativt systembiologisk projekt omhandlende Octamer-bindende transskriptionsfaktor 4 (OCT4) i embryonale stamceller (ESC) differentiering. Analysen er baseret på integration af ChIP-chip data, silencing data og genekspression data. Vi identificerede et meget tæt forbundet protein-protein interaktionsnetværk af OCT4-regulerede proteiner, der synes vigtig for ESC differentiering.

De følgende tre kapitler i afhandlingen præsenterer projekter, som udforsker biologien i testiklen. I kapitel 7 præsenteres et eksperimentelt studie, der undersøger vigtigheden af androgen receptor (AR) signalering i peritubular myoid celler for korrekt Leydig celle udvikling og funktion, hvilke er essentiel for testosteronproduktion og normal sædproduktion. Kapitel 8 beskriver et projekt, hvor vi undersøgte sædcelleproduktionen i voksne, samt de cellulære og transskriptionelle ændringer i testiklen efter lav dosis bestråling. Ved clusteringanalyse af genekspressionsdata identificerede vi fem grupper med unikke genekspressionsmønstre, der hver repræsenterer specifikke testikelceller. Vi identificerede ligeledes Leydig celleforandringer efter bestråling, hvilke måske påvirker testikelcancer patienters sædcelleproduction efter strålebehandling. Kapitel 9 præsenterer en undersøgelse af microRNA (miRNA) profilen i carcinoma *in situ* (CIS) celler, hvilke er forløberen til kønscelletumorer i testikler. Vi fandt ligheder med føtale gonocytter, hvorfra CIS celler udvikles pga. genetisk prædisposition eller hormonal forstyrrelse under fosterudviklingen. Vi fandt ligeledes nogle af de samme miRNAer udtrykt i testikulære kønscelletumorer, hvilket indikerer at miRNA profilen identificeret i CIS celler er vigtig for testikelcancerudvikling.

Det sidste projekt i afhandlingen (kapitel 10) beskriver et projekt af en potentiel markør for antihormonbehandling i brystcancer, Tissue metallopeptidase inhibitor-1 (TIMP-1). Studiet er en kombination af eksperimentelt arbejde og dataanalyse og sammen påvises at progesteronreceptoren virker til at have en vigtig rolle i den øgede resistens mod antiøstrogenbehandling i brystcancer med høj ekspression af TIMP-1.

Overordnet præsenterer denne afhandling fem projekter, der bidrager til forståelsen af biologien i fosterudvikling, i testikel og i brystcancer. Projekterne viser at kombinationen af systembiologi og eksperimentelt arbejde er en stærk analytisk kombination, som kan hjælpe med at belyse kompleks biologi og bidrage til forståelsen af sygdomme og forbedre behandling i fremtiden.

Papers included in the thesis

M Welsh, L Moffat, **K Belling**, LR de Franca, TM Segatelli, PTK Saunders, RM Sharpe and LB Smith. *Androgen receptor signalling in peritubular myoid cells is essential for normal differentiation and function of adult Leydig cells.* Int J Androl 2011 [Epub ahead of print]

KC Belling, M Tanaka, M Dalgaard, JE Nielsen, HB Nielsen, K Almstrup, H Leffers. *Recovery of spermatogenesis after irradiation.* Manuscript in preparation

GW Novotny*, **KC Belling***, JB Bramsen, JE Nielsen, J Bork-Jensen, K Almstrup, SB Sonne, J Kjems, E Rajpert-De Meyts and H Leffers. *Gonocyte-Like microRNA Expression in Carcinoma In Situ Cells of the Testis.* Manuscript submitted

C Bjerre*, L Vinther*, **KC Belling***, AS Rasmussen, R Yadav, J Wang, R Gupta, U Lademann, N Brünner and J Stenvang. *TIMP-1 overexpression confers resistance of MCF-7 breast cancer cells to fulvestrant.* Manuscript in preparation

*These authors contributed equally

Papers not included in the thesis

M Welsh, L Moffat, **KC Belling**, LR de França, TM Segatelli, K De Gendt, G Verhoeven, PTK Saunders, RM Sharpe, LB Smith. *Investigation of Androgen-Dependent Stromal-Epithelial Interactions in the Testis Reveals Novel Pathways Promoting Spermatogenesis.* Manuscript in preparation

L Fogh*, O Hekmat*, S Munk, NF Jensen, **KC Belling**, R Gupta, N Brünner, J Stenvang, U Lademann and JV Olsen. *Phosphoproteomic analysis of TIMP-1 overexpressing reveals increased expression and phosphorylation of Topoisomerase proteins.* Manuscript in preparation.

M Chrusciel*, **KC Belling***, M Doroszko, S Vuorenoja, M Dalgaard, I Huhtaniemi, H Leffers, HB Nielsen, J Toppari, N Rahman. *Regulation of Gene Expression in Adrenocortical Tumorigenesis after Luteinizing Hormone Receptor Ablation or Blocking Gonadotropin Releasing Hormone Action.* Manuscript in preparation

*These authors contributed equally

Abbreviations

4-OH-TAM	4-hydroxytamoxifen
A_{al}	$A_{aligned}$
AFP	Alphafetoprotein
AHSG	Alpha-2-HS-glycoprotein
AI	Aromastase inhibitor
AMH	Anti-müllerian hormone
ANOVA	Analysis of variance
A_{pr}	A_{paired}
APOA1	Apolipoprotein A-I
APOA2	Apolipoprotein A-II
AR	Androgen receptor
A_s	A_{single}
BRCA1	Breast cancer 1
BRCA2	Breast cancer 2
CBP	Calmodulin binding protein
CCD	Charged-coupled device
cDNA	Complementary DNA
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation-sequencing
CIS	Carcinoma *in situ*
cRNA	Complementary RNA
Dazl	Deleted in azoospermia-like
DBCG	Danish breast cancer cooperative group
DHT	Dihydrotestosterone
E2	17β-estradiol
E2F1	E2F transcription factor 1
EB	Embryoid body
EC	Embryonal carcinoma
ELISA	Enzyme-linked immunosorbent assay
ER	Estrogen receptor
ERα	Estrogen receptor alpha
ERBB2	*v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2*
ESC	Embryonic stem cell
FDR	False discovery rate
FGB	Fibrinogen beta chain
FISH	Fluorescence *in situ* hybridization
FPKM	Fragments per kilobase of transcript per million mapped reads
FSH	Follicle-stimulating hormone
GCN	Gene copy number
GnRH	Gonadotropin-releasing hormone
Gy	Gray
HCG	Human chorionic gonadotrophin
HER2	Human epidermal growth factor receptor 2
HR	Hormone receptor
Hsd3b	3beta-hydroxysteroid-dehydrogenase
IGF-1R	Insulin-like growth factor 1 receptor
IgG	Immunoglobulin G
IHC	Immunohistochemistry
ISH	*in situ* hybridization
In	Intermediate
LDH	Lactate dehydrogenase
LH	Luteinizing hormone
LHRH	Luteinizing-hormone-releasing hormone
MAPQ	Mapping quality
miRNA	microRNA

mRNA	Messenger RNA
MS	Mass spectrometry
NGS	Next generation sequencing
NSEM	Non-seminomas
nt	Nucleotide
OCT4	Octamer-binding transcription factor 4
PAM	Partitioning around medoids
PCA	Principal component analysis
PCR	Polymerase chain reaction
PGC	Primordial germ cell
pi	Post-irradiation
pn	Postnatal
poly(A)	Polyadenylated
PPI	Protein-protein interaction
PR	Progesterone receptor
PTM	Peritubular myoid
QRT-PCR	Quantitative reverse transcription polymerase chain reaction
RNA-seq	RNA sequencing
RPKM	Reads per kilobase of transcript per million mapped reads
rRNA	Ribosomal RNA
RT-PCR	Reverse transcription polymerase chain reaction
SCO	Sertoli-cell-only
SSC	Spermatogonial stem cell
SDS/PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEM	Seminoma
SERD	Selective estrogen receptor downregulator
SERM	Selective estrogen receptor modulator
SNP	Single nucleotide polymorphism
SRY	Sex determining region Y
TAP	Tandem affinity purification
TDLU	Terminal duct lobular unit
TDS	Testicular dysgenesis syndrome
TGCT	Testicular germ cell tumor
TGF	Transforming growth factor
TGFB3	Transforming growth factor, beta 3
Tgfbr3	Transforming growth factor, beta receptor III
TIMP-1	Tissue metallopeptidase inhibitor-1
TOP2A	Topoisomerase (DNA) II alpha
TP53	Tumor protein p53
UAS	Upstream activation sequence
UHC	Unsupervised hierarchal clustering
UTR	Untranslated region
Y2H	Yeast two-hybrid

Part I

Introduction

Chapter 1

Introductory remarks

Cells come in all sizes and shapes with all kinds of functions that are important for the tissue and the body they comprise. The molecules interaction in various ways and influence the function of each other. An understanding of the cell and its function as a whole is important in both the study of health and disease, and systems biology enable the study of the complex networks of molecules in a cell as a system.

Classical molecular biology has in the last centuries identified many of the components and interactions in cells. During the last decades, the development of high throughput technologies have revolutionized experimental biology and enabled large-scale detection and quantification of molecules such as genomic DNA, messenger RNA, protein and metabolites. Integrative systems biology enable integration and analysis of these large datasets whereby the complex interactions and regulatory mechanisms in cells can be elucidated.

The incidence of both poor semen quality and testis cancer is increasing with Denmark being one of the countries in the world with the highest incidence rates (1; 2). Breast cancer is likewise a disease that affect many people worldwide where women in developed countries have a life time risk of 12% of developing breast cancer (3). Thus, all three are pivotal diseases and increased knowledge and understanding of the biology of these tissues and diseases is essential to improve treatment and maybe even prevent some of the cases.

The approach we have taken in the projects presented here largely represents both the systems biological and the experimental approach and illustrates the benefit of a close collaboration between the fields. With this work, we produced results that improve our understanding of testis and breast diseases that might help patients in the future.

Chapter 2

The testis

The two functions of the testis are to produce germ cells and synthesize hormones, primarily androgens. The testis is comprised of numerous seminiferous tubules with the interstitium in between, where the production of sperm and hormones take place, respectively. The seminiferous tubules consist of an outer layer of peritubular myoid (PTM) cells that surround the basement membrane. Inside the tubules, Sertoli cells nurse the germ cells during spermatogenesis (4). The interstitium contains microvasculature, lymph vessels, nerve fibers, fibroblasts, connective tissue and Leydig cells (5). The Leydig cells produce hormones, primarily androgens. The testis is by far the major source of testosterone and also one of the main targets for male androgen action.

Testicular dysgenesis syndrome

Common disease phenotypes of the male reproductive system are cryptorchidism (undescended testis), hypospadias (abnormally placed urethral opening), poor semen quality and testis cancer. These four phenotypes are assumed to arise from male fetal maldevelopment and linking these phenotypes together as one development syndrome has been collectively referred to as the testicular dysgenesis syndrome (TDS) (6) (Figure 2.1). Evidence for a common syndrome is that the four phenotypes often co-occur: men with testis cancer often have extreme low sperm counts, also much lower than expected in a man with only one functioning testis (7), vice versa men with poor semen quality have higher risk of testis cancer than the average cancer risk in the population (8) and cryptorchidism is a risk factor for testis cancer (9).

TDS arises from a combination of genetic, environmental and lifestyle factors. Some genetic causes of TDS have been established such as 45,X/46,XY

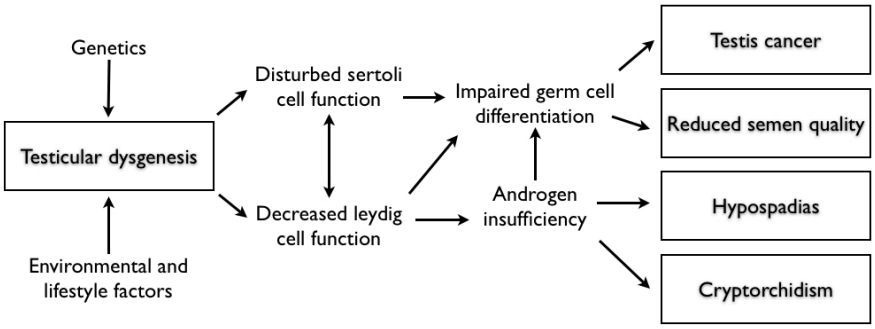


Figure 2.1. The testicular dysgenesis syndrome (TDS). Modified from Skakkebaek *et al.*, 2001 (6).

mosaic karyotype and *sex determining region Y (SRY)* mutations (10). As mentioned, environmental factors strongly influence TDS occurrence, since genetic factors alone cannot explain the 400% increase in testis cancer incidence in two generations (1). Another group of environmental risk factors are endocrine disruptors that either interfere with the production of testosterone or estrogens or disturb proper activation of their receptors, which are in an essential and fine-tuned balance during male fetus development (11; 12). Endocrine disruptors have been a large focus area and increasing evidence show that these chemicals can affect and disrupt proper embryogenesis. An example are boys whose mothers have been exposed to pesticides during pregnancy have impaired fertility (13) and pesticides in breast milk have been associated with cryptorchidism in the sons (14). Also, painkillers that inhibit prostaglandin synthesis have shown endocrine disruptor effects by preventing proper development of the Leydig cells (15). Life style also affects testis development, such as smoking (16) and alcohol consumption during pregnancy (17) also affects testis development. The TDS phenotype(s) present in a patient and the degree of severity depend on the degree of genetic predisposition and environmental effects that the patient was exposed to during fetal development. Thus, TDS is a complex disease and although the knowledge of the syndrome continues to increase there is still a long way before we have the full picture.

2.1 Testis development

Fetal development can be divided into two stages: sex determination and sex differentiation. In the beginning of fetal life there is no morphological or functional differences between females and males. The urogenital ridges develop in early fetal life with common urinary and genital systems for both sexes. In this stage, gonads consist of the common precursor cells, which for males are Sertoli cells and for females are granulosa cells, before they are

colonized with primordial germ cells (PGCs). At this stage the fetus has the potential to develop into both sexes (18). The Wollfian ducts have the potential to differentiate into the male genital organs, whereas the Müllerian ducts can give rise to the female genitals (19).

Sex-specific development is triggered by the expression of *SRY* in the male fetus that activates differentiation of Sertoli cells and directly or indirectly suppresses the female sex-determining pathway. In the absence of *SRY* in females, and also in males that lack functional SRY, fetuses enter the female sex-determining program (19). During male sex differentiation, the gonadal cells segregate into two compartments: the seminiferous tubules and the interstitium. The Sertoli cells assemble into tubules containing the PGCs and both cell types are surrounded by the basement membrane and the PTM cells. The Sertoli cells stimulate the sex-specific development of germ cells and produce anti-müllerian hormone (AMH), that causes regression of the Müllerian ducts. The somatic Leydig cells differentiate and secrete androgens (20) through the coordinated action of steroidogenic enzymes (21) and this testosterone action is essential for proper development of male reproductive organs (22). Testosterone from fetal Leydig cells is responsible for Wollfian duct differentiation in the first trimester, but without sufficient testosterone action the ducts will degenerate (23).

Androgens mediate its actions through the androgen receptor (AR) that is expressed predominantly in Sertoli, PTM, Leydig and perivascular smooth muscle cells (12). Testosterone production is during the first half of the embryogenesis stimulated by the placental human chorionic gonadotrophin (HCG), whereas luteinizing hormone (LH) is the main stimulant afterwards and onwards in adult life (24). Testosterone is converted by 5α-reductase to dihydrotestosterone (DHT) that is the hormone responsible for virilisation of the internal and external genitalia (25). DHT has a higher affinity for binding to the AR than testosterone

2.2 Spermatogenesis

Spermatogenesis is the process in the seminiferous tubules whereby sperm is produced. Spermatogenesis is initiated at puberty by a steep increase in testosterone levels. The increased testosterone production is induced by bursts of gonadotropin-releasing hormone (GnRH) secreted from the hypothalamic neurons that evoke corresponding pulses of LH and follicle-stimulating hormone (FSH) from the pituitary. LH stimulates testosterone production in Leydig cells and the amount is controlled by a negative feedback loop of testosterone acting on the hypothalamic-pituitary axis with suppression of GnRH and LH secretion (26).

FSH is also essential for spermatogenesis and mediates its action via binding to the FSH receptor on Sertoli cells. FSH activates a number of signaling pathways in Sertoli cells, whereof one effect is that FSH has a supportive action for testosterone as AR expression in Sertoli cells is upregulated in response to FSH (26). The presence of AR in germ cells is more controversial

and the theory is that the main androgen-stimulation of spermatogenesis occur indirectly via Sertoli cells (12), but the receptor is also present in PTM cells. FSH thus to a large extent control testicular size, germ cell numbers per testis and sperm production (27).

Spermatogenesis is a process of multiple divisions and differentiations that is sustained by the spermatogonial stem cells (SSCs). The division of SSCs is the initiation of spermatogenesis. Since SSCs are stem cells the daughter cells can take one of two paths: either the cells complete cytokinesis and become two new stem cells or they start to undergo division (28). Differentiating spermatogonia undergo multiple mitotic divisions before they differentiate to primary spermatocytes (29). Primary spermatocytes recombine the DNA before they undergo the first meiotic division that produces secondary spermatocytes, which immediately proceeds with the second meiotic division whereby haploid round spermatids are produced (30), see Figure 2.2.

The germ cells move from the outer part of the tubule, close to the basement membrane, towards the lumen in the middle whilst they develop (5). Once the germ cells have differentiated to mature elongated spermatids they are released into the tubular lumen and then called spermatozoa. The spermatozoa further undergo multiple morphological changes during maturation in the tubular lumen whilst they passively are transported in the fluid through rete testis where they leave the testis and continue into the epididymis (28; 30).

The developing germ cells are surrounded by Sertoli cells in the seminiferous tubules during spermatogenesis. Sertoli cells communicate with germ cells autocrinely and paracrinely and through Sertoli cell-germ cell junctions (32). Sertoli cells have multiple functions such as nutrition, production of seminiferous fluid and production and secretion of spermatogenesis-regulating factors (5). Sertoli cells are interconnected by tight-junctions forming the blood-testis barrier of the testis dividing the seminiferous epithelium into a basal and an adluminal compartment. Germ cells move through the blood-testis barrier during maturation and once they have passed the barrier, the germ cells are protected from the immune system and other extraneous substances (33).

The problem with poor semen quality is increasing, and Danish men have one of the lowest semen qualities in the world (2; 34). Only a small percentage of infertile or subfertile men have the option of an effective treatment. An example is in cases where hormonal disturbances cause the infertility where compensating treatment often is a success. Yet, for most couples, where the male have poor semen quality the only solution is by assisted reproductive techniques, i.e. *in vivo* or *in vitro* fertilization (35). Thus, a further understanding of spermatogenesis and the origin of poor semen quality is needed in order to be able to hinder development of TDS and to be able to offer a better treatment for men with poor semen quality.

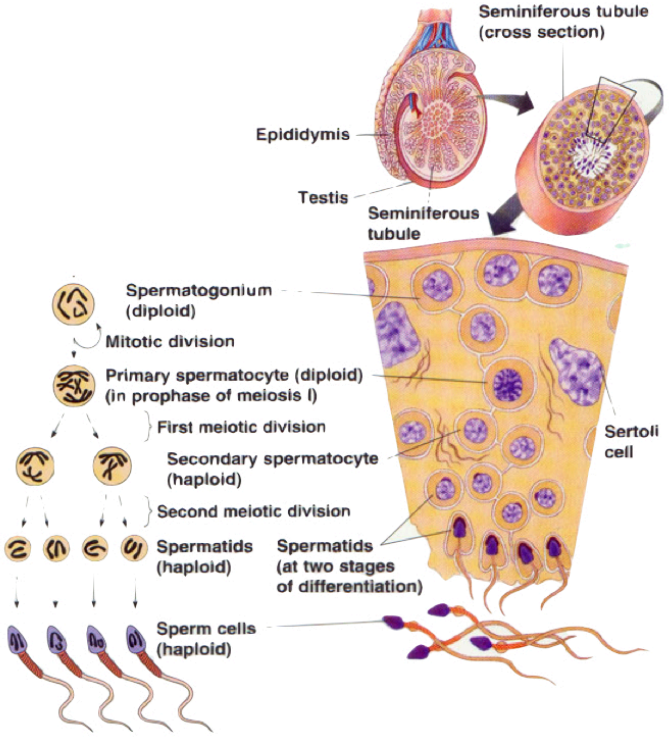


Figure 2.2. Spermatogenesis is the process of sperm production in the testis. A cross section of a testis is shown in the top of the figure with the epididymis on the outside. Spermatogenesis takes place in the seminiferous tubules where the germ cells move from the outer part close to the basement membrane towards the adluminal part during development. Haploid spermatids are released into the lumen and subsequently called spermatozoa. Spermatozoa undergo further differentiation during their passive transport out of the testis to the epididymis through rete testis. The differentiation stages of germ cells and their ploidity are outlined to the left in the figure (31).

2.3 Testis cancer

Testis cancer accounts for 1% of male cancers worldwide (36). Testis cancer incidence is gradually increasing especially in the developed countries. Incidence rates are highly variable between countries and with 15.2 cases per 100,000 inhabitants, Denmark is the country with the highest incidence rate in the world (1) (Figure 2.3). Similar trends are seen for other TDS phenotypes and this relatively rapid increase in incidence strongly points to the possible role of environmental and lifestyle factors in the etiology (37; 38).

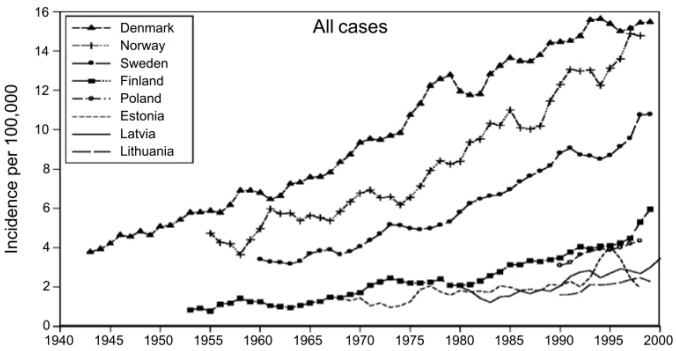


Figure 2.3. Testis germ cell tumor incidence by year of diagnosis and country (1).

Testis cancers can be divided in testicular germ cell tumors (TGCTs) and non-germ cell tumors. The vast majority of testis cancer patients are diagnosed with TGCTs (around 95%) (39), which most often occur in young men between 17 and 45 years of age (40). TGCTs are divided into two histological types: seminoma (SEM) and non-seminoma (NSEM). SEM are very uniform tumors that resemble early fetal germ cells, whereas NSEM are very heterogeneous and more malignant than SEM (41). Both SEM and NSEM arise from a common precursor cell, the carcinoma *in situ* (CIS) cell (42; 43). CIS cells are thought to be derived from gonocytes that did not fully differentiate during embryonal development (42; 43). Histologically CIS cells resemble gonocytes (44) and occupy the place of spermatogonia in the postpubertal testis (40). CIS cells also resemble gonocytes immunohistochemically and express several factors associated with stem cells, such as the KIT receptor (45; 46) and *octamer-binding transcription factor 4 (OCT4)* (47).

CIS cells are most likely generated because of a developmental arrest of germ cell differentiation. The malignant transformation is properly caused by a disturbance in the microenvironment of the differentiating fetal germ cells, a process that is strictly regulated and very sensitive to disturbances by hormonal and paracrine factors. The factors might not target the germ cells directly, but can be an indirect effect via disturbed somatic cell function (48). Androgens are important players in regulation of the cellular microenvironment and essential for proper male development and often androgens are also targets for environmental chemicals (40). Differentiation of PGCs into early spermatogonia is a long and slow process, which is the window where the hormonal disturbances are believed to cause the generation of CIS cells (48).

The exact mechanisms whereby CIS cells transform into malignant cells are poorly understood. CIS cells likely functions as cancer stem cells that drive tumorigenesis by the abilities of self-renewal and pluripotency (49).

Male germ cells lie dormant in the testis until puberty where spermatogenesis is initiated by a boost in testosterone (26). The increased testosterone level at puberty activate the somatic cells in the testis, which induce the spermatogonia to start dividing and go into meiosis. These signals are presumably misunderstood by the gonocyte-like CIS cells as a signal to proliferate (44). CIS cells either fills the tubules and form solid SEM or develop and give rise to NSEM with a higher cellular variety (50).

2.3.1 Cancer cell biology

Cancer typically develops from cells that acquire genomic alterations and the corresponding changes in gene expression modify normal growth control and survival pathways. The genetic changes include all sorts of alterations such as altered karyotype, gene copy number (GCN) variations, single nucleotide polymorphism (SNPs) and epigenetic changes (51). Since cancer arises from genetic changes, no patient suffers from the exact same disease, but cancers show some general features. Hanahan and Weinberg describes that cells need to acquire six distinct features in order to become cancerous: self-sufficient in growth signals, limitless replication potential, apoptosis resistance, sustained angiogenesis and tissue invasion and metastasis (52).

Normal cells have checkpoints and several repair mechanisms that prevent damaged DNA to be incorporated in the genome and passed on to daughter cells. Yet, mutations accumulate with age and increase cancer risk (53). Cancers arise both from acquired and inherited genetic changes and many cancers have some known germline mutations that predispose for the disease (54). However, most cancers do not have a clear hereditary link (55) and develop slowly over many years from pre-malignant tumors to benign to malign tumors that are able to spread though invasion and metastasizing.

Cancerous gene aberrations can be divided into two types: gain-of-function mutations in oncogenes and loss-of-function mutations in tumor suppressor genes (52). Most proteins encoded by oncogenes are components of signaling pathways that promote cell proliferation and survival and can have various roles in signaling cascades, such as growth factors and transcription factors or regulators of cell cycle or cell death (56). Gain-of-function mutations are often dominant and promote inappropriate high proliferation and survival of cells (57).

Tumor suppressors are genes whose loss or inactivation also can promote cancer development. Loss-of-function mutations in tumor suppressor genes are most often recessive where both alleles need to have loss-of-function mutations before the function of the gene is lost (58). Tumor suppressors consist of two subgroups: caretakers and gatekeepers. Proteins of caretaker genes are involved in DNA maintenance and repair and loss-of-function mutations lead to decreased protection of the genomic integrity. Examples of gatekeeper proteins are regulators of the cell cycle, checkpoint-proteins that arrest cell cycle upon DNA damage and receptors or signals transducers for signals that inhibit cell proliferation and apoptotic proteins (57).

Tumor cells are continuously evolving through genetic aberrations. A permanent selective pressure of tumor cells exists for the most competitive cells for tumor growth and dissemination. Initially, chromosome changes lead to small premalignant, hyper-proliferative lesions. As the benign cells gain more genetic aberrations, the lesion becomes more compatible and can evolve to more advanced tumors (59).

2.3.2 microRNAs

microRNAs (miRNAs) are small non-coding RNAs with an approximately length of 22 nucleotides (nts). miRNAs regulate translation of messenger RNAs (mRNAs) by binding to complementary sequences in the 3'-untranslated region (UTR). Most miRNAs repress translation of the bound mRNA (60; 61), while others enhance translation (62). Most often miRNA binding also reduce the level of the target mRNA (63). miRNAs play important roles in carcinogenesis and have both tumor suppressor and oncogenic effects (64) and miRNA target genes that are involved in angiogenesis, apoptosis, cell migration and invasion (65).

miRNAs also have important roles for testis biology since miRNAs are essential for proper development of PGCs and male germ cell differentiation (66; 67). miRNAs have also shown to be involved in testis carcinogenesis: miR-17-92 might promote tumor development through inhibition of apoptosis by inhibiting E2F transcription factor 1 (E2F1) protein synthesis in CIS cells (Novotny et al., 2007b). miR-372 and -373 can bypass the Tumor protein p53 (TP53) cell cycle checkpoint and these miRNAs have been shown present in TGCTs (68; 69). Also, other miRNA clusters have shown differential expression between normal tissue and germ cell tumors (70). We have investigated miRNA profiles in testis CIS cells and tumors in the project presented in chapter 9.

2.3.3 Testis cancer diagnosis and treatment

There are currently no imaging or serological methods in clinical use for diagnosis of CIS cells. Open surgical biopsy followed by immunohistochemistry (IHC) staining is the only way to diagnose CIS (44). Therefore, testis cancer is usually detected as a swelling of the testis where 5-10% of TGCTs have spread extragonadally. Diagnosis is confirmed by ultrasound, tumor markers and imaging to detect any metastases. The used serum markers are HCG, lactate dehydrogenase (LDH) and alphafetoprotein (AFP) levels, which are all SEM and/or NSEM markers (71). Testis cancer is staged in three stages: *(1)* within the testis, *(2)* spread locally to pelvis or abdomal lymph nodes, and *(3)* spread to distant lymph nodes. The staging is an estimate of how far the cancer has spread and dictates treatment in general. Testis cancer has a very good survival rate, testis cancer patients have a cure rate of 90% (72) and for early testis cancer without metastases the cure rate is about 99% (73).

Radiotherapy

Radiotherapy has been shown to be very useful in treatment of testis cancer (74; 75). Radiation dosages used in Denmark lie in the range of 16-20 Gray (Gy) (76). Between 5 to 15% of testis cancer patients have CIS cells present in the contralateral testis (77) and one-sided TGCTs carry an approximately increased risk of 25- to 50-fold of developing a contralateral tumor (78; 79). Therefore, CIS and cancer cells in both testis should be killed by therapy (44). Radiotherapy destroys the rapidly dividing CIS, cancer and germ cells, whereas the more slowly dividing somatic testis cells are preserved. Adjuvant radiotherapy is offered to stage 1 and in some cases to stage 2 testis cancer patients (72).

The cellular effects of testis irradiation are still not fully described. Patients treated with irradiation (20 Gy) showed Sertoli-cell-only (SCO) syndrome with all CIS and germ cells eradicated (80). Patients often have increased HCG and LH levels and decreased testosterone levels after radiotherapy (75; 80). Thus, Leydig cells seem to be affected by irradiation to some extent. This is further investigated in the project presented in Chapter 8.

Cisplatin-based chemotherapy

Cisplatin-based chemotherapy is offered to some stage 2 patients and all stage 3 patients that have disease spread beyond the testis (72). The high cure rate of testis cancer is to a large extent due to Cisplatin-based chemotherapy that cures more than 80% of TGCTs (72). Cisplatin becomes positively charged once it enters cells and its main anti-cancer effect is through DNA binding that hinders transcription and DNA replication that lead to cell cycle arrest and apoptosis (81).

Chemotherapy often causes azoospermia with no measurable sperm in the semen, however patients may recover after treatment (82). Yet, up to 60% of testis cancer patients already have low sperm counts at diagnosis (83) and sperm banking is an option for patients with a reasonable sperm count before both radiotherapy and chemotherapy (84).

Chapter 3

The breast

The female mammary glands develop to fulfill their primary function, which is to synthesize milk to nourish infants (85). Anatomically, the human breast is segmentally divided into 15-20 glandular lobes. Each lobe is composed of numerous smaller lobules that contain the mammary glands, which produce milk at lactation. The milk is drained from the site of synthesis via small branching ducts that lead to the terminal duct lobular units (TDLUs) (86). The ductal network further carries the milk into the lactiferous duct that leads to the nipple (87). The milk ducts consist of an inner layer of luminal epithelial cells and an outer contractile layer of myoepithelial cells. The lobes and milk ducts lie within fat pads of encapsulated adipocytes and connective tissue, which also contain the vascular and lymphatic system (Figure 3.1).

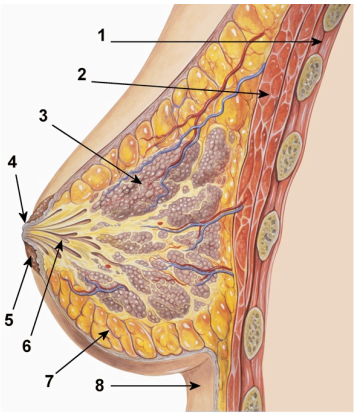


Figure 3.1. Anatomy of the adult human female breast. The arrows point at the following: 1) chest wall, 2) pectoralis muscles, 3) lobules, 4) nipple, 5) areola, 6) lactiferous duct, 7) fatty tissue, and 8) skin (88).

3.1 Breast cancer

Breast cancers mostly arise from the epithelial cells lining the TDLUs. As long as cancer cells remain within the basement membrane of the TDLUs and the draining ducts the cancer is called *in situ*. Once the cancer cells disseminate to outside the basement membrane into the surrounding tissue it is classified as invasive (89). Breast cancer is the most common type of cancer among women and is most frequent in developed countries such as Europe, Australia and Northern America (Figure 3.2) (90). The lifetime risk of developing breast cancer is about 12% in the United States, which is the lifetime risk in other western countries as well (3).

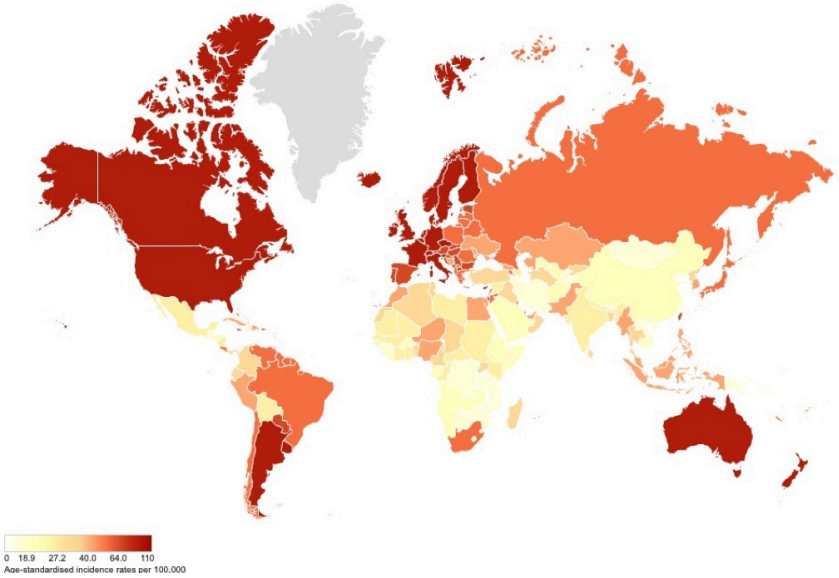


Figure 3.2. Estimated breast cancer incidence worldwide in 2008 (90).

Risk factors

Breast cancer is a multifactorial disease. Hereditary predisposition to breast cancer is an important risk factor with 5-10% of newly diagnosed breast cancers having a positive family history. The most common germ line mutations include mutations in *Breast Cancer 1 (BRCA1)*, *Breast Cancer 2 (BRCA2)* and *TP53*. Mutations in these high risk breast cancer genes considerably increase lifetime risk especially in young women (91).

Lifestyle has huge impact on breast cancer development. Alcohol consumption, smoking and physical inactivity are factors known to increase breast cancer risk (92). Late full-term pregnancy also increases breast cancer

risk, because pregnancy causes cell differentiation in the breast as preparation for lactation, which makes the breast less susceptible to carcinogenesis (93). Prolonged exposure to estrogens is strongly associated with an increased breast cancer risk. Natural estrogen risk factors are early age of menarche and late age of menopause, whereas exogenous estrogen risk factors include use of oral contraceptives and hormone replacement therapy (94). The adipose tissue serves as site for synthesis of hormones and obesity increases breast cancer risk, especially in postmenopausal women where the adipose tissue becomes the major source of estrogens (95).

In addition to abovementioned factors, the environment can also promote breast carcinogenesis. An example is that environmental estrogens can cause neoplastic transformations in breast cells (96). The interactions between environmental factors and genetic predispositions are also of increasing interest. Low penetrance susceptibility genes have little impact on overall breast cancer risk, but might have high penetrance in combination with specific environmental factors (92). For a more detailed review of cancer biology see section 2.3.1.

3.2 Diagnosis and prognosis

Breast cancer is usually detected as a lump in the breast, other signs can be unexplained breast swelling, thickening, skin irritation and tenderness (97). Breast cancer diagnosis and treatment in Denmark follows the guidelines defined by the Danish Breast Cancer Cooperative Group (DBCG)[1]. Mammography screening is used to detect breast cancer as early as possible, even earlier than the lump is clinically detectable by palpation (98) and mammography screening is offered to women of 50-69 years. Mammography is not accurate enough to prove the presence or absence of cancer by itself, but in combination with clinical examination, biopsy, and other imaging techniques, a more definitive diagnosis can be determined (99).

Several histopathological characteristics of the tumor are determined at diagnosis to provide prognostic information. Prognosis is an estimate of patient outcome at the time of diagnosis independent of systemic treatment and has high impact on the choice of treatment allocated the patient. The tumor characteristics used in the clinic in Denmark are: tumor size, tumor type, tumor grade, number of involved axillary lymph nodes and metastatic disease (99). Pathologists divide breast tumors into different types because different tumor types have different prognosis. Ductal and lobular carcinomas are the most common. They arise from luminal epithelial cells lining the ducts and from lobules of the milk glands, respectively (100). Tumor grade is an estimate of tumor differentiation. Tumors are divided into three different histological grades based on the amount of tubular formation in the tumor, the phenotypic characteristics of the cells and mitotic counts. Tumors that are poorly differentiated (high grade) have worse recurrence-free and overall

[1] http://www.dbcg.dk

survival (101). The spread of the disease is evaluated by spread to axillary lymph nodes and to other organs (102).

Biomarkers are also used for breast cancer prognosis, today the panel includes the estrogen receptor alpha (ERα), the progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and *topoisomerase (DNA) II alpha (TOP2A)*. These factors are all used to assess prognosis and to predict optimal adjuvant therapy for breast cancer patients. ERα and PR are together called hormone receptors (HRs) and protein expression of both are determined by IHC (99). HR-status is a prognostic marker and HR-positivity of a tumor is predictive for endocrine therapy (103). Endocrine therapy targets the estrogen receptors (ERs) and PR serves as a functional assay to indicate that the ERα signaling pathway is intact, even if the tumor is reported ERα-negative (104). Approximately 75% of all breast cancers are HR-positive (105) and these patients have a more favorable prognosis compared to HR-negative patients. This may be due to a prolonged disease course as these patients in general metastasize later than HR-negative breast cancers (106).

HER2 is a membrane tyrosine kinase and is overexpressed and/or gene amplified in about 20% of breast cancers (107). Overexpression of HER2 is determined by IHC. Overexpression of HER2 is often caused by amplification of the coding gene *v-erb-b2 Avian Erythroblastic Leukemia Viral Oncogene Homolog 2 (ERBB2)* and IHC staining of HER2 can be confirmed by determination of the GCN of *ERBB2* by fluorescence *in situ* hybridization (FISH) (108). Both HER2 overexpression and *HER2* amplification have prognostic and predictive value and are associated with a higher proliferation rate, higher resistance to apoptosis, and in general more aggressive cancers (107). HER2-positive tumors are recommended anti-HER2 therapy with trastuzumab, an antibody described in the next section (108).

TOP2A is the fourth biomarker used at present in clinics in Denmark. TopoIIα, encoded by *TOP2A*, is a vital enzyme in all cells where it functions to maintain DNA structure (109). *TOP2A* has GCN changes in almost 90% of tumors with *HER2* amplification and in 10% of tumors that do not show *HER2* amplification (110). GCN of *TOP2A* can be determined by FISH and both amplifications and deletions are predictive for antracycline-based therapy (103; 111), described in the next section.

3.3 Treatment

The primary treatment of breast cancer is most often surgery, mastectomy or breast conserving surgery, lumpectomy. The sentinel node technique is used to determine the involvement of lymph nodes and verify the amount of operation needed. Neoadjuvant therapy is administration of systematic treatment before main treatment and in some cases it is allocated patients before surgery to reduce tumor size and increase the possibility of successful breast conserving surgery (112). Postoperative radiotherapy is offered to

all patients following lumpectomy and to patients with node positive disease, which is when the tumor cells have spread to the draining lymph nodes (113).

In general, primary operable breast cancer has a 40% risk of recurrence after surgery (114). For some patients surgery, perhaps followed by radiotherapy, is enough treatment to cure their cancer. These patients should not be offered systematic adjuvant therapy if possible, as these patients do not benefit from the treatment. Other patients have more aggressive cancers and therefore need additional treatment to either cure their cancer or prolong and improve their lives if curation is not possible. Systematic adjuvant therapy is offered to high risk patients with a 10 years recurrence risk higher than 10% and a mortality risk higher than the population in general, called the high risk group (99; 115). Breast cancer patients are divided into a low risk and a high risk group based on the histopathological characteristics and prognostic markers of their tumor (Table 3.1). Systematic treatment reduces the risk of recurrence by 40-50% (116) and is allocated to all high risk patients. Patient has to fulfill all criteria in the low risk group to be classified as a low risk patient (99).

Table 3.1. Prognostic factors used to classify breast cancer patients into a low risk and a high risk group (99; 103).

Risk Factor	Low Risk	High Risk
Age	$\geq$ 50 years	$<$ 50 years
Lymph node status	Negative	Positive
Tumor size	$\leq$ 10mm	$>$ 10mm
IDC malignancy grade	I	II-III
ILC malignancy grade	I-II	III
HR status	Positive	Negative
HER2 status	Negative	Positive
TOP2A status	Normal or unknown	Deleted or amplified

IDC: Invasive ductal carcinoma; ILC: Invasive lobular carcinoma

Endocrine therapy

Breast cancer patients with HR-positive tumors in the high risk group are offered endocrine therapy (99). HR-positive tumors are estrogen-dependent for growth and endocrine therapy is prescribed to reduce and/or stop growth of the tumor cells. Three different classes of endocrine therapies exist with different working mechanisms: selective estrogen receptor modulators (SERMs), selective estrogen receptor downregulators (SERDs) and aromatase inhibitors (AIs). SERMs and SERDs both work by competing with estrogen for binding to the ERs. Tamoxifen is a SERM, a partial ER antagonist that acts as an antagonist in breast tissue and as an agonist in other tissues e.g. the uterus, whereas fulvestrant is a SERD and a pure ER antagonist in all tissues.

AIs inhibit estrogen synthesis by inhibiting aromatase that converts androgens to estrogens. Treatment with AIs reduces circulating estrogen, which limits available estrogen that can stimulate tumor growth (117). Ovarian suppression or ablation are alternative ways to reduce the level of circulating estrogens. Ovarian suppression is done by luteinizing-hormone-releasing hormone (LHRH) treatment, which leads to low LH and further low estrogen production. Ovarian ablation is permanent either done by oophorectomy or ovarian radiation, which hinders the major production site for estrogens in producing the hormone (118).

Anti-HER2 therapy

Patients with HER2-positive tumors are offered anti-HER2 therapy with trastuzumab (99). Trastuzumab is a humanized monoclonal antibody that binds the extracellular domain of HER2 and thereby inhibits downstream signaling of the membrane tyrosine kinase (107). The exact mechanism of action is not known, but trastuzumab might induce immune cells to kill the cells bound by the antibody (119).

Chemotherapy

Endocrine therapy and anti-HER2 therapy both target specific molecules and pathways in tumor cells. Chemotherapy does not target specific molecules, but is cytotoxic to fast proliferating cells. In general, cancer cells have a higher proliferation rate than other somatic cells and are therefore more sensitive to anti-proliferative agents. In Denmark, chemotherapy is offered to patients with HR-negative tumors, HER2-positive tumors in addition to trastuzumab and to patients with *TOP2A*-positive tumors (99). Chemotherapy is most often given in a regime of three drugs with different working mechanisms, because it has proven more efficient than single-agent treatment (116). Some of the most used chemotherapies are anthracyclines and taxanes. GCN changes of *TOP2A* are predictive for antracycline-based therapy (120). Anthracyclines have multiple antitumor activities such as interference with DNA synthesis, production of free radicals that damage DNA, and induction of apoptosis (121). The exact function of taxanes is not fully known, but taxanes are thought to cause apoptosis in proliferating cells by stabilizing the microtubules in the spindle assembly check point in mitosis whereby the cell is hindered in finishing mitosis (122).

3.4 Further subclassification

Dividing breast cancers by histopathological characteristics and biomarkers and tailor adjuvant treatment thereafter is beneficial for the course of disease and survival of the patients. Clinical trials have proven significant benefit of tailored systematic adjuvant therapy for specific patient groups: HR-positive breast cancer patients allocated the recommended 5-year endocrine therapy

have a reduced risk of mortality of 11.8% compared to patients that were not allocated endocrine therapy (116). HER2 is a promising molecule for targeted anticancer therapy and trastuzumab therapy allocated to patients with HER2-positive patients reduce their risk of disease recurrence by 50% and the risk of death by about 33% (123). *TOP2A*-positive patients are more responsive to antracycline-based adjuvant therapy (124). In general, six months of anthracycline-based polychemotherapy reduces the annual death rate with 38% for patients younger than 50 years and by 20% for patients between 50-69 years at diagnosis (116).

Yet, not all patients that have tumors with predictive markers for a therapy benefit from the treatment. Only 50-60% of HR-positive tumors respond to endocrine therapy (125). Response rates to anti-HER2 treatment with trastuzumab as a single agent lie between 15% to 30% and in combination with other chemotherapies from 50% to 80% (126). Response rates to chemotherapies lie in the range of 11.5% to 57% (127). Thus, the markers used in the clinic today are not efficient enough in themselves as prognostic markers or as predictors of benefit from anti-cancer treatment.

In addition to *de novo* resistance, acquired resistance is also a big problem for breast cancer patients. Some patients that in the beginning respond to treatment later develop resistance. Acquired resistance to endocrine therapy is mostly due to development of anti-estrogen unresponsiveness as absence of ER is an uncommon phenomenon in acquired resistance. Almost all patients with metastatic disease and many that receive tamoxifen as adjuvant therapy eventually experience tumor tamoxifen resistance. Few of these experience a further switch to a tamoxifen-stimulated growth (128).

Development of chemotherapy resistance is also a big problem in treatment of breast cancer patients. Most anticancer therapies do change tumor growth, but often the response is not long lasting and in some cases, a more aggressive cancer and drug resistance develops. Around 30% of women diagnosed with early-stage disease progress into metastatic breast cancer. Response rates for these patients to chemotherapy lie between 30%-70%. Often the response is not durable and once resistance develops, treatment options are limited (129).

Thus, a more detailed diagnosis, prognosis and development of new anticancer treatments are needed. A research project that investigates the predictive value of a potential new marker, Tissue metallopeptidase inhibitor-1 (TIMP-1), is presented in chapter 10.

Chapter 4

Systems biology

4.1 Integrative systems biology

Throughout the 19th and 20th centuries, classical molecular biology was the dominant approach in the study of the complexity of cells (130). The discipline is described as reductionist as it assumes that the behavior of a biological system can be explained by the properties of its constituents. Hence, the behavior of cells can be understood by the study of genes, proteins etc. on a single molecule level (131). Molecular biology has successfully identified a lot of the components and many of the interactions in cells (132). This is in part due to the development of analytical methods that are able to measure most of the cellular components at cellular level. Yet, the characteristics of the individual molecules do not lead to a functional understanding of a system, but the methods have increased the possibility of describing the complete system (133) and understand the components in more precise detail.

Systems biology is an alternative approach to study and understand how complex systems underlie life. Systems biology is often described as holistic as it believes that a biological system cannot be explained by its components alone but only via the study of the system as a whole (131). The systems biology approach has evolved during the last 50 years and has rapidly increased during the last decade and become a mature field in biological research (134). Although many interpretations of the phrase 'systems biology' exist, two general approaches are commonly used: hypothesis-driven and data-driven. Hypothesis-driven research starts out with a theoretical model of a biological process based on assumptions and hypotheses, which then need to be underpinned via experimental data and analysis. Data-driven research is the unbiased experiments and data analysis not guided by previous knowledge or theories.

A common aspect of most systems biology approaches is the integration of heterogeneous data types, which combines experimental biology with theoretical modeling. Complex biological networks are thought to give better understanding as a whole by observing multiple components simultaneously by data integration of quantitative measurements (135). Multiple data types and resources can be used as evidence layers such as omics data, chromatin immunoprecipitation (ChIP)-chip data and protein-protein interaction (PPIs). Some of the data types used in integrative systems biology and also used in my PhD projects will be described next. The systems biology approach will be illustrated in the end of this chapter with a data analysis of the role of OCT4 in embryonic stem cell (ESC) pluripotency and differentiation. This is an analysis for a future paper described in an abstract in chapter 6.

4.1.1 Omics data

The ability to obtain, integrate and analyze complex datasets is one of the cornerstones in systems biology. Omics refers to the study of omes, e.g. genomes and proteomes, through high-throughput screening techniques that collect data for the study of interactions and regulatory mechanisms at a systems level (136). The number of 'omes' referred to today has grown from 'genome' and 'proteome', to microbiome, glycome, transcrptome and metabolome to name a few, all enabled through the development of particular technology platforms.

The DNA sequence, the genome, includes protein-coding genes, non protein coding genes, regulatory elements and noncoding sequences. Genomics is the study of the organismal genome and its variations, which encompasses SNPs, somatic mutations as well as genomic rearrangements and aneuploidies. Transcriptomics is the study of all RNA molecules produced in one cell or in a cell population (136; 137). Proteomics is the measurement of proteins and peptides on organismal, tissue, or cellular level and the study of the structure, expression pattern and function. Proteomics also includes the study of posttranslational chemical modifications of proteins and peptides such as phosphoproteomics and glycoproteomics (136).

The omics fields have lead to a drastic increase in the level of detail of cells in form of identification and quantification of its components (133). The methods used to study omics vary in different fields from microarrays to sequencing to affinity purification and mass spectrometry (MS). Of the mentioned omics areas, I have worked thoroughly with transcriptomics both global gene expression conducted by microarrays and RNA sequencing (RNA-seq). I have especially worked with gene expression microarray data, which is part of the projects presented in chapter 6, 8, 9 and 10. Therefore, the next chapter in the thesis focuses on transcriptomics, the methods of gene expression microarrays and RNA-seq and their data analysis (chapter 5).

4.1.2 ChIP-chip

ChIP-chip is an experimental method to study interactions between DNA and proteins. This *in vivo* technique isolates cross-linked protein-DNA molecules with an antibody specific to the protein of interest. The DNA molecules bound by the protein are then identified by the DNA microarray technique (the chip) (138), Figure 4.1.

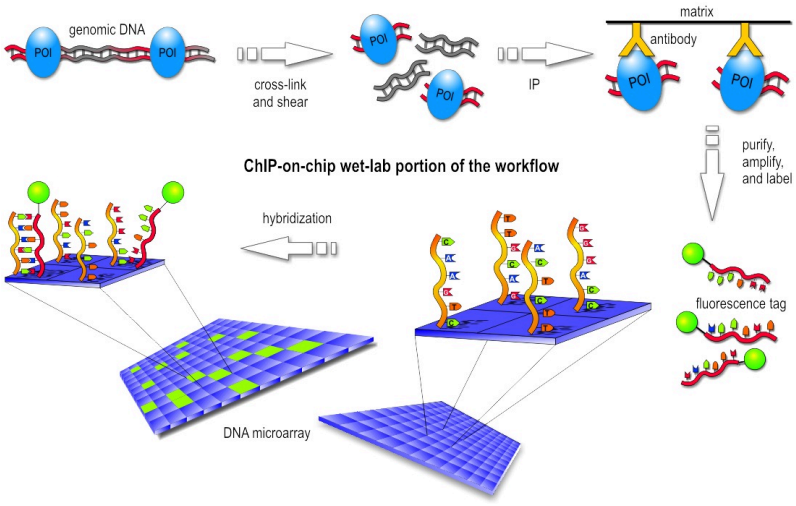


Figure 4.1. An overview of the wet-lab part of a ChIP-chip experiment. DNA and proteins are covalently cross-linked and the chromosomal DNA is subsequently fragmented. Specific antibodies immunoprecipitate the target protein with the cross-linked DNA fragments. The unbound molecules are washed away and the bound DNA fragments are released, labeled and identified by the DNA microarray technique. POI: protein of interest (138; 139).

The method is used widely for identification of gene expression control elements bound by transcription factors (139) and in determination of the genome-wide transcriptional regulatory network. Transcription factors play a central role in regulation of gene expression and have great impact on cell function, because transcription factors often bind and influence the expression of a lot of genes, either increasing or decreasing (140). The bound transcription factors bring the regulatory elements into proximity with the target gene and interact with the DNA polymerase complex directly or via other proteins or complexes that mediate polymerase interaction, i.e. chromatin remodelers and modifiers (141).

The unique strength of ChIP is that it captures protein-DNA interactions *in vivo* and that the interactions can be studied under different cellular conditions and time points. A disadvantage is that the method requires

an antibody specific for the protein of interest. Also, the method does not discriminate between proteins that bind directly to DNA and proteins that bind to DNA via other proteins (138). ChIP-chip is being rapidly replaced by chromatin immunoprecipitation-sequencing (ChIP-seq), because of the higher resolution of sequencing (142).

4.1.3 PPI networks

A way to study functional modules in cells is through PPI networks, which is a useful tool to understand the interactome at a molecular level (143). Protein complexes are dynamic and change configuration and interactions in a time- and space dependent fashion in order to take part in signaling pathways and contribute to cellular architecture. Thus, the study of protein complexes is useful to understand the molecular nature of cellular programs (144).

Experimental methods to identify PPIs

Advances in experimental biology have enabled large quantities of interaction data between genes and proteins. High throughput screens have yielded identifications of protein complexes with methods such as the yeast two-hybrid (Y2H) assay and tandem affinity purification (TAP) followed by MS, together called TAP-MS (145). The Y2H assay identifies interactions of a target protein through the expression of a downstream reporter gene that is transcribed upon activation by binding of a transcription factor to an upstream activation sequence (UAS). The transcription factor is split into two fragments: the DNA binding domain and the activation domain. The binding domain is fused to the known target protein ('bait') and the activation domain is fused to other proteins maybe a whole library of proteins ('prey') that potentially interact with the bait. The binding domain with the bait binds the UAS and if a prey binds the bait the activation domain initiates transcription. The PPI is can then be detected as the expressed reporter gene, see Figure 4.2 (146; 147).

An alternative method to identify PPIs is TAP-MS, which is a two-step purification step of a protein of interest, bait, together with its interacting proteins in complex, preys. The bait protein is expressed *in vivo* together with the TAP tag that consists of a calmodulin binding protein (CBP) closest to the bait followed by Protein A. Protein A binds immunoglobulin G (IgG) and in the first round of purification the tagged bait is purified by IgG-agarose beads in a column. The column is washed, the TAP-tagged bait is released from the column along with its protein complexes, and Protein A is cleaved off. In the second purification, the baits are purified via their CBP tag on calmodulin beads. The column is washed and the protein complexes are released. The proteins in the purified complexes are separated by one-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS/PAGE) and characterization by MS (144).

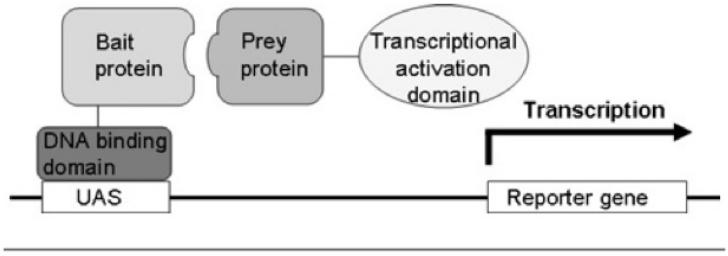


Figure 4.2. The Y2H assay. The protein of interest ('bait') is fused to the DNA binding domain of a known transcription factor that binds and activates a reporter gene. The other half of the transcription factor that contains the transcriptional activation domain is fused to proteins that potentially interact with the bait protein ('prey'). The binding domain binds an upstream activation sequence of a reporter gene and if a prey interacts with the bait, the reporter gene is transcribed and the interaction is thereby detected. UAS: upstream activation sequence (147).

Y2H screens have been widely used and have identified a large number of binary interactions. Yet, the method has received criticism as it tends to identify a large proportion of false positives (147). TAP-MS is an *in vivo* method that is able to purify low-abundance proteins still attached to other proteins as a complex. Not all proteins in a protein complex interact (148) and from TAP-MS it is not possible to identify which proteins that interact. The method looses some week interactions in the complex during the purification steps and the method also generates some false positives (147). The data quality from both experimental methods must therefore be considered when analyzing the data.

InWeb

PPI data is publicly available in numerous databases with some of the most common being BIND (149), DIP (150), GRID (151), HPRD (152), IntAct (153), KEGG (154), MINT (155), MPact (156) and Reactome (157). The knowledge in the databases is extracted from published data either by textmining, manual curation or a combination of the two (148).

CBS has its own in house human PPI network, InWeb, which is updated on a regular basis. The database is built from a selection of interactions from the public available PPI databases mentioned above. The public available human PPI data is limited compared to data from model organisms. Many PPIs are conserved across species (158) and high throughput PPI data from model organisms have been incorporated in InWeb by mapping proteins to human orthologues by use of InParanoid database in order to increase coverage (159).

The reliability of the interactions in InWeb are confidence scored to filter out spurious interactions that are likely false positives (160). First, a topology score is calculated for each interaction in the network based on the topology of the local network (161) and also taken into account the experiment that reported the interaction and the number of experiments that confirm it. The topology score is then converted into an evidence score by comparison against a benchmark set of high-confidence interactions. InWeb is the protein interactome used in the analyses presented in the next section.

4.2 OCT4-regulatory network

The project presented here is in collaboration with The Max Planck Institute for Molecular Genetics, Berlin, Germany, from where the data was received after a preliminary analysis. I have performed the integrative systems biology analysis at CBS that is described in the following.

OCT4 is a transcription factor that is an important player in ESC pluripotency and differentiation (162). For further introduction see chapter 6. The aim of this study was to identify OCT4-regulated genes important for ESC differentiation. The analysis is included in this chapter to illustrate the systems biology approach to gain knowledge of functional modules in cells. The analysis integrates three datasets generated on human cells: OCT4 ChIP-chip data from ESCs, gene expression data from an OCT4 silencing (72h) study from ESCs, and gene expression data from a time series of embryoid body (EB) differentiation. EB arise from culturing ESCs derived from blastocysts (163). EB differentiation begins by formation of an aggregate of ESCs and the cells then start to differentiate to all three germ layers: endoderm, mesoderm and ectoderm (164). Thus, EBs provide a good *in vitro* culture for the study of early differentiation events (165).

The ChiP-chip data and silencing data was first compared to previously published ChIP-chip (166; 167; 168; 169; 170; 171) and silencing studies (169; 171; 172; 173; 174) to find out whether agreement among published studies of OCT4 targets exists. If so, the overlap is a good estimate for the most potential OCT4-regulated genes. However, very little overlap was found between both OCT4 ChIP-chip studies and silencing studies. Therefore, the conclusion was that no benefit was gained from including the previously published OCT4 studies and the further analysis was performed using the data received from the Max Planck Institute alone.

A group of 139 genes was identified from the human ESC ChIP-chip study to be bound by OCT4, thus primary targets. To identify the targets involved in ESC differentiation, we focused the further analysis on the OCT4 targets that changed in gene expression during EB differentiation, which were 94 of the 139 primary OCT4 targets. We generated a PPI network based on the 94 genes and set a cutoff for the confidence score of 0.20, which gave us a network with 118 proteins in total whereof 70 proteins were from the input set and 48 were added interacting proteins.

Gene Ontology analysis was performed on the proteins in the network to gain knowledge of their roles in cells. The Gene Ontology Consortium[1] provides a knowledge database of the role of eukaryotic genes and proteins in three independent ontologies: biological process, molecular function and cellular component (175). Gene Ontology analysis is an easy obtainable way to gain a broad functional overview of a subset of proteins.

The Gene Ontology analysis of the 118 proteins in the PPI network confirmed that the proteins in the network were highly involved in ESC differentiation. Two of the most statistically significant biological process categories were multicellular organismal development and anatomical structure morphogenesis, all enriched categories are listed in Table 4.1.

Table 4.1. Gene Ontology biological process categories enriched in the proteins in the OCT4-regulated network.

Category	No.	Bonferroni *p*-value
Multicellular organismal process	55	1.18^{-9}
Multicellular organismal development	38	3.33^{-5}
Plasma lipoprotein particle assembly	5	3.39^{-5}
Protein-lipid complex assembly	5	3.39^{-5}
Phospholipid transport	6	1.32^{-4}
Organ development	27	3.49^{-4}
Blood coagulation	8	0.001
Coagulation	8	0.002
Phospholipid efflux	4	0.002
Developmental process	42	0.002
System development	30	0.002
Hemostasis	8	0.002
Anatomical structure morphogenesis	22	0.004
Organ morphogenesis	15	0.005
Platelet activation	5	0.008
Regulation of body fluid levels	8	0.008
Anatomical structure development	31	0.009
Reverse cholesterol transport	4	0.01
Sterol transport	5	0.01
Cholesterol transport	5	0.01
Cholesterol homeostasis	5	0.01
Sterol homeostasis	5	0.01

Each of the 118 proteins in the network was statistically tested for whether the protein had more interacting proteins differentially regulated during EB differentiation than expected by random. The more interconnected the proteins in the PPI network were with OCT4-regulated proteins, the more important they appeared for ESC differentiation. We did not require that the node itself was differentially expressed during EB differentiation. After multiple testing correction, 32 genes were statistically significant highly interconnected with regulated genes in the network. A gene list was prioritized according to their interactions with regulated proteins during EB differentiation.

[1] http://www.geneontology.org/

A third layer of evidence was added in the form of whether a gene changed in gene expression after OCT4 silencing. RNA silencing is an experimental method to inhibit translation of a target gene transcript (176). It is a useful method to study cellular changes, in this case gene expression changes, caused by the lack of a protein. Compared to ChIP-chip studies that study primary targets, RNA silencing studies all molecules affected by the function of the target molecule (177).

The most potential OCT4-regulated candidates from this data integration analysis (1) were the primary targets of OCT4 identified from the ChIP-chip study that (2) were highly interconnected with genes that changed in gene expression during EB differentiation and also (3) were affected by OCT4 silencing. Eight genes were identified as top candidates of OCT4-regulated genes important for ESC differentiation as they fulfilled all three criteria (Table 4.2).

Table 4.2. Top candidate genes in the OCT4-regulated network in ESCs

Gene	Gene *reg.	*Reg. interactions	Total interactions
APOA1	yes	10	18
APOA2	yes	9	11
AHSG	yes	10	12
FGB	yes	4	7
IGFBP4	no	2	2
APOF	no	2	2
IGFBP6	no	2	2
THBD	no	2	3

AHSG: Alpha-2-HS-glycoprotein; APOA1: Apolipoprotein A-I; APOA2: Apolipoprotein A-II; APOF: Apolipoprotein F; FGB: Fibrinogen beta chain; IGFBP4: Insulin-like growth factor binding protein 4; IGFBP6: Insulin-like growth factor binding protein 6; THBD: Thrombomodulin. *Reg: Regulated during EB differentiation

A further step in the analysis was to choose OCT4-regulated gene candidates for experimental validation at The Max Planck Institute. Secreted proteins were preferred for validation identified as part of the extracellular matrix. The enriched Gene Ontology cellular components categories (Table 4.3) showed that four of the eight OCT4-regulated gene candidates were part of the extracellular matrix: Apolipoprotein A-I (APOA1), Apolipoprotein A-II (APOA2), Alpha-2-HS-glycoprotein (AHSG) and Fibrinogen beta chain (FGB). These four proteins are therefore the most potential candidates for experimental validation.

The full PPI network of the OCT4-regulated genes are shown in Figure 4.3 that display that the most potential OCT4-regulated genes are highly interconnected in the network.

This analysis of the OCT4-regulatory network important for ESC differentiation is a good example of how the integrative systems biology approach can add to the enlightenment of functional modules in cells. Also, that this approach gives a better overview of the functional cell as a system compared to the study of molecules at a single molecular level.

Table 4.3. Gene Ontology cellular component categories enriched in the proteins in the OCT4-regulated network.

Category	No.	Bonferroni p-value
Extracellular space	37	2.57^{-22}
Extracellular region part	37	4.57^{-21}
Extracellular region	43	2.03^{-14}
Extracellular matrix	14	2.23^{-6}
Proteinaceous extracellular matrix	13	9.58^{-6}
Chylomicron	4	3.75^{-4}
Protein-lipid complex	5	7.32^{-4}
Plasma lipoprotein particle	5	7.32^{-4}
Very-low-density lipoprotein particle	4	0.001
Triglyceride-rich lipoprotein particle	4	0.001
Fibrinogen complex	3	0.005
High-density lipoprotein particle	4	0.006
Spherical high-density lipoprotein particle	3	0.008

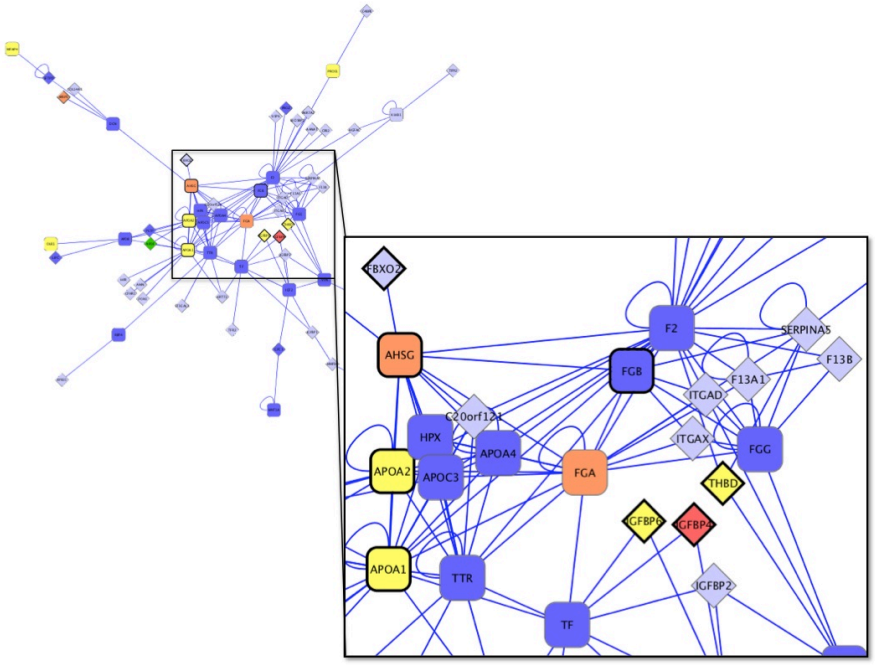


Figure 4.3. **The PPI network based on the primary OCT4-regulated genes changed in expression during EB differentiation. The top eight OCT4-regulated genes are marked with black bold node borders. The four proteins chosen for validation are APOA1, APOA2, AHSG and FGB. Purple: No change in gene expression during EB differentiation or after OCT4 silencing; Blue: Downregulated after EB differentiation, but no significant change after OCT4 silencing; Yellow: Downregulated during EB differentiation and upregulated after OCT4 silencing; Orange: Downregulated during EB differentiation and after OCT4 silencing; Red: No significant change during EB differentiation and upregulated after OCT4 silencing; Green: No significant change during EB differentiation, but downregulated after OCT4 silencing.**

Chapter 5

Gene expression profiling

The transcriptome can be divided into two types of RNAs: the coding mRNAs that are translated into proteins and the non-coding RNAs that have regulatory and structural roles without being translated to proteins (178). Gene expression microarrays are today used routinely to determine gene expression values for thousands of transcripts in a single experiment (179). Yet, next generation sequencing (NGS) provides an alternative technology, RNA-seq (180), and this technology finds a very rapid early adoption globally that might overtake microarrays in the not too distant future.

Many of the projects in this thesis include gene expression profiling. I have been involved in all elements of the process of gene expression profiling by microarrays from the process of extracting RNA, labeling, hybridization and especially in the data analysis. I have also analyzed RNA-seq data that was generated from some of the same MCF-7 cell clones that were gene expression profiled by microarrays (study published in chapter 10). The sequencing data is not included in that paper and not published yet. The methods of both gene expression microarrays and RNA-seq will be described in this chapter as well as the analysis of data generated by the two methods.

5.1 Gene expression microarrays

The development of the DNA microarrays started back in the 1970s, where spotted arrays were developed from growing bacterial colonies with genomic inserts on filter paper, that was then hybridized with radioactive labeled targets (181). The first high-throughput microarray was developed in 1995 (182). Today competing vendors have evolved multiple technologies and microarray platforms, which differ on areas such as probe implementation on the array, probe design, density of the probes on the array and RNA isolation

and labeling (183). Since I have worked mostly with Agilent gene expression arrays, I will focus on their technology in the following section.

5.1.1 The microarray platform

Agilent produces long oligonucleotide (60-mers) arrays where the probes are synthesized *in situ* by inkjet printing using phosphoramidite chemistry (184). The longer probes compared to the ones used by other technologies (e.g. Affymetrix uses 25-mers) provide higher hybridization property, but are also more tolerant to mismatches (179). The inkjet approach has the advantage of allowing customization of probe sequences, but it requires more space compared to alternative methods such as photolithography and therefore Agilent arrays have lower probe density compared to many other platforms (184).

There are several protocols for preparation of the target library that is applied to the array. In general, the first step is to isolate total RNA or mRNA from the sample of interest. In gene expression profiling the focus is to quantify the expression of protein-coding mRNAs, which in eukaryotic cells carry a polyadenylated (poly(A)) tail at the 3' end. Most RNAs in eukaryotic cells are ribosomal RNA (rRNA) and the poly(A) sequence is often used to enrich for mRNA transcripts (183), which can be done by use of oligo(dT) primers for complementary DNA (cDNA) synthesis (185).

Several amplification steps are often included in the preparation of the library to lower the need for RNA and to ensure enough hybridization signal (181). Nucleic acid amplification is often accomplished by a combination of reverse transcription polymerase chain reaction (RT-PCR) and polymerase chain reaction (PCR). The library that is applied to the array consists of cDNA or complementary RNA (cRNA) that is labeled during amplification or after hybridization with the probes on the array. During hybridization, complementary sequences gradually find each other preferentially over mismatched pairings. Unspecific bound and unbound molecules are washed away after hybridization and the array is then scanned to detect the amount of labeled targets attached to the specific probes on the array (183).

5.1.2 Data analysis

Gene expression profiling is most often performed as case-control comparisons, such as healthy versus disease and before and after treatment. Thus, data from several arrays are compared in the data analysis. Small differences in the steps of preparation of the library can add technical noise and global differences between the signals on the arrays. Hence, microarray data has to be normalized between arrays to remove systematic biases, so that the following statistical test most likely detects biological differences. One approach is linear normalization where the intensities on the arrays are scaled according to a chosen constant, such as housekeeping genes or an array constant like the median of all array intensities. Linear normalization assumes that variations are linear, which are often not the case. An alternative approach

is non-linear normalization (186), e.g. quantile normalization that makes the distribution of probe intensities the same across arrays. The idea behind this is that the total amount of fluorescence intensities are the same across arrays, because most of the genes in the different samples are equally expressed and only a small number of genes are differentially expressed (187).

5.1.3 Statistical testing

Gene expression microarrays are high throughput experiments that measure mRNA levels of thousands of transcripts in parallel and statistical testing is needed to identify genes that are differentially expressed between samples. In two-condition cases with a proper number of replicates, assigning significance with students *t*-test can be a powerful method. In cases with two or more conditions, one-way analysis of variance (ANOVA) can compare the means in the groups, whereas a two-sided ANOVA can identify differentially expressed genes affected by two variables, such as disease and treatment. These methods are all parametric and make some assumptions about the data, such as that the data is normally distributed and the variables are independent. Alternative non-parametric tests use rank-based approaches and are more robust than the parametric tests, an example of which is the Wilcoxon rank-sum test (188).

Alternative experimental set ups exist such as time series. Several approaches exist to analyze time series data of which one of the most used are clustering algorithms (189). In the study presented in chapter 10, I analyzed a series of gene expression data from ten MCF-7 breast cancer cell clones with different expression of TIMP-1. I used linear regression analysis to identify transcripts whose expression correlated with changing TIMP-1 expression. Linear regression is a statistical method that enables identification of a linear relationship between a dependent variable and one or more variables (136). From the linear regression analysis I identified genes that correlated both positively and negatively with TIMP-1 expression (see Figure 5.1).

In statistical tests, one hypothesis is typically set up for each gene detected, so each gene is statistically tested separately for differential expression between samples. A widely accepted statistical caution states that by testing tens of thousands of hypotheses, one will, by chance, identify a subset of transcripts that show statistical significance. One approach to compensate for this randomness is to correct for multiple testing of which there are several approaches. One is to apply Bonferroni correction on the *p*-values by multiplying by the number of performed tests (190). This will limit the number of false positives, but also remove some of the true positives. An alternative approach is to calculate the false discovery rate (FDR) by random tests of the data and reset the *p*-value cutoff to an acceptable level of false-positives (191).

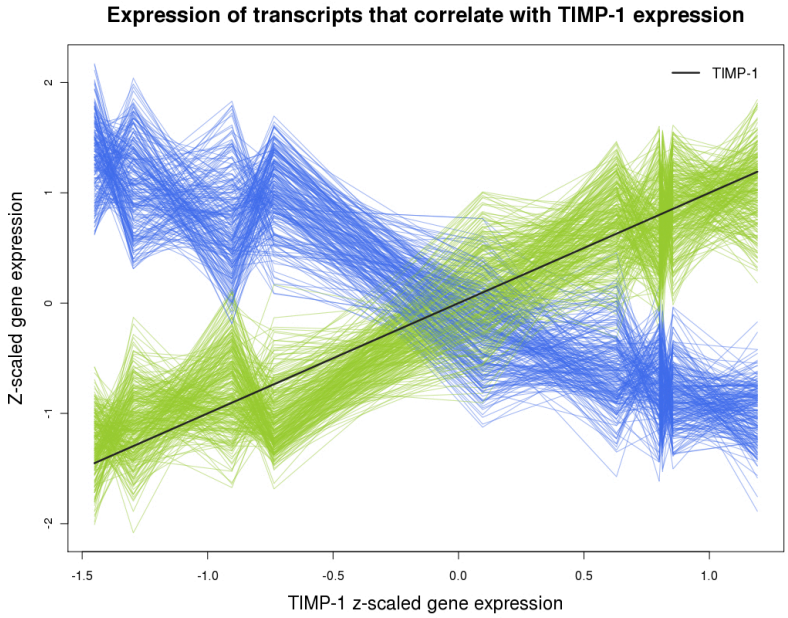


Figure 5.1. Expression of transcripts (n=508) that correlated with TIMP-1 expression in MCF-7 breast cancer cell clones identified by linear regression analysis. TIMP-1 expression is marked by black and the positively and negatively correlated genes are marked by green and blue, respectively.

5.2 RNA sequencing

The development of genomic sequencing technologies started back in the 1970s (192) and the first automated DNA sequencer was introduced in 1986 (193). First generation sequencers sequenced individual DNA clones, generated long continuous reads (>800 nts) (178) yet was a very expensive and time consuming procedure (194). Automization in sequencing has been a revolutionary technology in genomics and was used to sequence the first human genome published in 2003 (195; 196). The second generation sequencers were introduced in 2005 and sequence up to tens of millions molecules in parallel, but with shorter read length around 25-400 nts (178). The introduction of the second generation sequencers lowered the expenses significantly and the sequencing technology became more competitive with the microarray technology.

The development of high-throughput DNA sequencing has also provided a method for mapping and quantifying the transcriptome. Both mRNA and small RNAs can be sequenced depending on the protocol for library preparation and the sequencing method. There are several applications for RNA-seq:

quantification of gene expression and of alternative splicing, discovery of novel transcripts, discovery of transcribed SNPs and somatic mutations (197). I have analyzed paired-end RNA-seq data generated by Illumina sequencing[1] on mRNA from four of the ten MCF-7 breast cancer cells clones that were gene expression profiled on microarrays (chapter 10). The analysis is ongoing and the main focus of the analysis has been to quantify gene expression levels from the RNA-seq data and compare with the microarray data. Next, I will describe the method of Illumina library construction and sequencing and describe the analysis of the data as I have performed it with the challenges that I have experienced.

5.2.1 The sequencing method

The first step in RNA-seq is to prepare the cDNA library from the RNA of interest. Total RNA is isolated and enriched for mRNA by purification on oligo(dT) beads. NGS still has limitations on read length and mRNAs are fragmented into 200-700 nt fragments (198) that are used as templates for first-strand cDNA synthesis. After second-strand cDNA synthesis, the short cDNA fragments are isolated and the ends of the fragments are repaired. The poly(A) is subsequently added to the 3' ends to which adapters are ligated. The adapters provide sequence specificity for PCR primers and is used during sequencing to ligate the targets to the flow cell. The ligated cDNA molecules are purified, amplified by PCR and size selected on a gel (199).

The cDNA library is denatured and ligated randomly to the surface of the solid flow cell by the adapters. The ligated cDNAs undergo bridge amplification by PCR to form a cluster of identical cDNA fragments in proximity, illustrated in Figure 5.2. Reaction mixture is added to the flow cell including the four nts labeled with different fluorescent dyes. Sequencing is performed in cycles. For each cycle the complementary nt is bound to the cDNA fragments, the unbound nts are washed away and the position and dye is registered by a charged-coupled device (CCD) camera. The cycles are repeated a number of times and a base-calling algorithm assigns sequences and associated quality values for each read. Paired-end reads are produced from sequencing each cDNA from both ends and thus yields two reads per cDNA molecule. After completion of the first read, the templates can be regenerated *in situ* for sequencing of the second read from the complementary strand (197). Analyses have indicated that paired-end sequencing yields more information about alternative exons and isoforms compared to single-end sequencing (200).

5.2.2 Data analysis

Quality control

The raw reads are received in the fastq format that represents each read with four lines: *(1)* sequence identifier, *(2)* raw sequence letters, *(3)* optional

[1] www.illumina.com

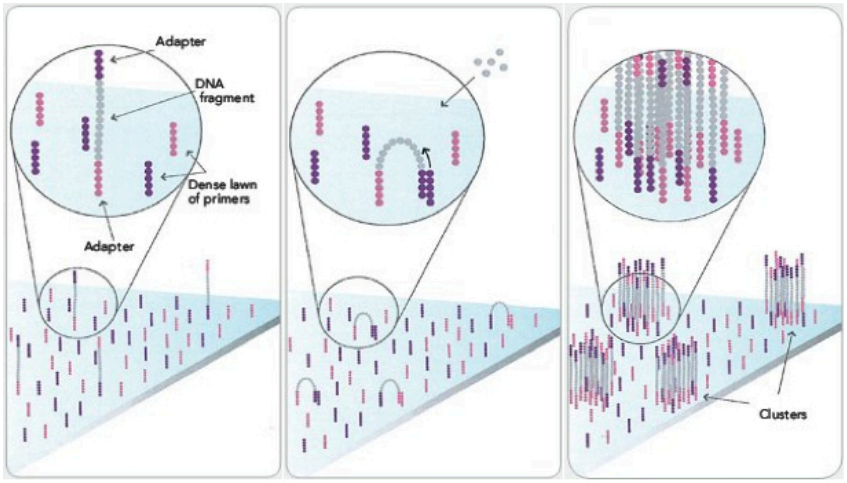


Figure 5.2. Illumina RNA-seq flow cell preparation. Left: The single-stranded cDNA fragments from the library are attached randomly to the flow cell by the adapter sequence. Middle: The fragments undergo multiple bridge amplifications on the flow cell to generate clusters of similar cDNA fragments. After each amplification, the double-stranded cDNAs are denatured to single-stranded cDNAs to function as templates for yet another amplification. Right: Sequencing is performed on single-stranded cDNA clusters. Picture modified from www.illumina.com.

repeat title line, and *(4)* phred score quality values, one for each nt in the read. The phred quality score is defined in Equation 5.1, where P is the base error probability (201).

$$\mathrm{Q} = -10 \times \log_{10}(P) \tag{5.1}$$

It is useful to calculate the sequence coverage and depth of the data. Sequence coverage is the average number of reads that cover a nt in the coding transcriptome (202), whereas the depth is the number of times the nt is covered by reads. This gives an idea of the quality of the data. The depth needed for proper results of the analysis depends on the purpose of the analysis, the more rare splice variants one wants to detect the higher sequencing depth is needed (198). At the single read level, some important parameters to check are whether adapter sequences are left in the reads, the quality of the reads and the amount of clonal reads in the data. NGS is based on PCR amplification than can introduce some base sequence errors and favor certain sequences over others, thus changing the relative frequency and abundance of various cDNA fragments than existed before amplification (203). There are several publicly available tools that provide an overview of the raw read quality, I used one called FastQC[2] that reports numerous quality

[2] www.bioinformatics.bbsrc.ac.uk/projects/fastqc/

parameters such as per base sequence quality, sequence length distribution and overrepresented sequences, e.g. adapters.

Some of the raw reads contain adapter sequences and other exogenous contents. Some times adapters are even sequenced due to operational errors and other unknown reasons. Adapter sequences are not biological sequences and if a read is mapped containing an adapter sequence, the adapter part will not map and the whole read is discarded. The quality of the reads also becomes lower towards the end of the reads and base errors can cause mapping errors and false positives during mapping. Therefore, both adapters and bad quality nts need to be trimmed out before alignment to improve correct mapping (204). Bad quality reads are either trimmed or removed completely based on the phred quality score (Equation 5.1). I have used a tool called Fastx trimmer[3] designed to handle paired-end reads. An example of per base sequence quality of the reads before and after trimming, generated by FastQC, is illustrated in Figure 5.3.

Alignment

Alignment of RNA-seq data has several challenges: Single base-call errors do not influence alignment in a big way since most mapping algorithms are tolerant to a few mismatches especially over a read as long as 90 nts. However, resolving larger differences will require higher sequencing depth (198). Some reads span exon-exon junctions as the library is made from mRNA (178). There are two different types of aligners that can handle exon-exon splice sites: unspliced and spliced aligners. Unspliced aligners map reads against a reference cDNA library and do not allow gaps. Spliced aligners align reads against a reference genome and allow large mapping gaps for proper placement of reads that span exon-exon junctions (205). I have used a spliced aligner called TopHat[4] that align the reads in two rounds. First, all unspliced reads are mapped using an unspliced aligner called Bowtie[5]. Second, all unmapped reads are split into shorter segments that are aligned independently. The genomic regions where the segmented reads map are then searched for possible splice connections (206).

At aligning a mapping quality (MAPQ) is assigned to the aligned reads, which is the phred-scaled probability (201) that a read alignment may be wrong. Equation 5.2 outline the calculation, where P_r is the probability that a read is wrongly mapped. Thus, MAPQ=30 implies that there is a probability of 1 in a 1000 that the read is incorrectly mapped (207). In Figure 5.4 is mapping statistics calculated by SAMstat[6](208) on the mapping of one of the RNA-sequenced MCF-7 cell clones.

$$\text{MAPQ} = -10 \times \log_{10}(P_r) \tag{5.2}$$

[3]http://hannonlab.cshl.edu/fastx_toolkit/
[4]http://tophat.cbcb.umd.edu/
[5]http://bowtie-bio.sourceforge.net
[6]http://samstat.sourceforge.net/

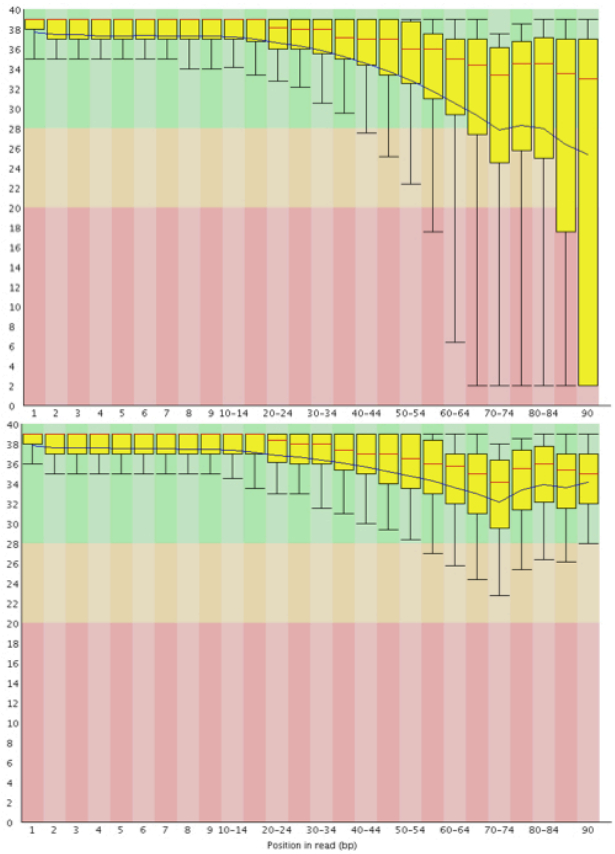


Figure 5.3. FastQC reports on average per base sequence quality. On the *X*-axis is the nts in the 90 base pair reads and on the *Y*-axis is the quality distribution for each nt. The yellow boxes are the interquartile range (25-75%), the red line is the median and the blue line is the mean quality. The black whiskers go from the 10th to the 90th percentile. In the top is the read quality distribution of the raw reads and below is the quality distribution of the trimmed reads. The reads were trimmed with the criteria: a minimum phred score of 20 and minimum read length of 25 nts. A phred score of 20 corresponds to a base call accuracy of 99%. (Equation 5.1.)

Transcript assembly

Mapped reads are assembled into transcripts for quantification and comparison across samples. Read alignment and transcript assembly are complicated by numerous factors: The cDNA library is build from both mature and incomplete spliced mRNA, which makes it difficult to identify the mature sequences. Between 15 and 20% of the reads in the human genome cannot

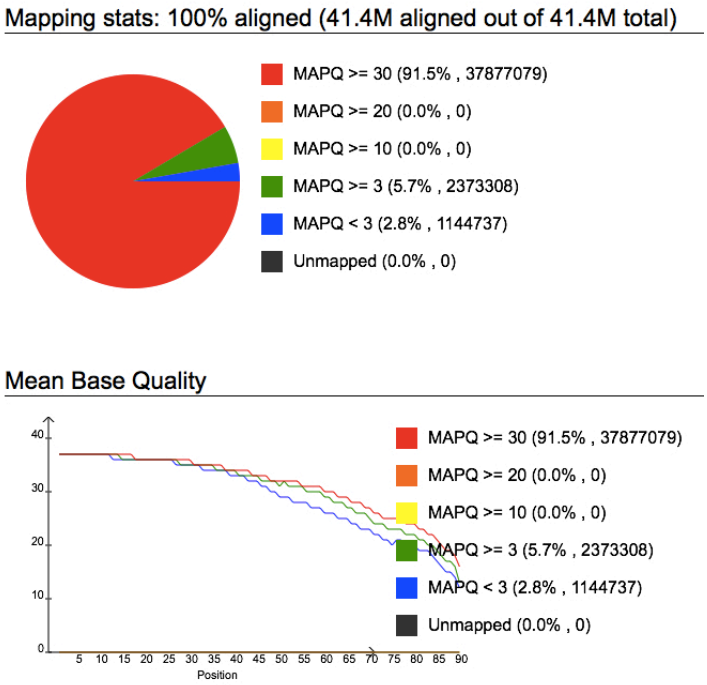


Figure 5.4. Mapping statistics. In the top is the number of alignments and below is the mean base quality of the reads, both in various MAPQ intervals. Percentage and the number of alignments in each category is given in brackets in both plots.

be unambiguously mapped to a single location (209), whereof pseudogenes and gene families account for many repetitive elements around the genome. Overlapping genes also complicate assembly, because it is not possible to identify from which sequenced mRNA a read belongs to unless reads only map to exon positions that belong to only one of the overlapping genes. Yet, this is still complicated as most transcript and exon boundaries are poorly defined (198). Fusion genes also complicates read alignment and transcript assembly. Fusion genes are chromosome abnormalities that arise from rearrangements in the genome. RNA-seq have shown useful to identify fusion genes (210; 211). Yet, one has to be aware of fusion genes when analyzing RNA-seq and separate expression of each of the two genes with the expression of the fusion gene.

Many genes have isoforms that share multiple exons (212) and it is difficult to estimate the abundance of each isoform as reads may map to shared exons. Several approaches have been suggested to detect and quantify the correct splice variants from RNA-seq data. One is to only count the reads that map to unique exons for each single isoform (213). Yet, this approach does not

work for genes that do not have exons unique to its splice variants (205). An alternative approach is the "isoform-expression method" that handles uncertainties by constructing a likelihood function and assign abundances of splice variant estimates based on the maximum likelihood that best explain the reads obtained in the experiment (205).

Transcript assembly can be performed genome-independent without a reference genome or genome-guided. There are pros and cons of both approaches and the choice depends on the biological questions that one wants answered. The genome-independent is preferable if one wants to identify new exons or new splice variants. The approach is also necessary in cases with no annotation. An advantage of the genome-guided method is that it captures all known information in cases of well annotated organisms, but it is not able to identify unknown splice variants and might struggle with samples with a lot of rearrangement events, such as cancer cells (205). I have used a genome-guided algorithm, Cufflinks[7], that assemble reads aligned by a spliced aligner, such as TopHat, and uses the "isoform-expression method" to handle splice variants. Cufflinks assembles reads that are aligned in close genomic proximity into a graph and reports the minimal number of compatible isoforms chosen so that all reads are included in an isoform. Cufflinks reports the set of splice variants that have the highest read coverage, which are the isoforms most likely present in the sequenced sample (214). I experienced that Cufflinks had large variations in the splice variants that it reported expressed in the four breast cancer cell clones and it seems unlikely that four cell clones should have large differences in the splice variants expressed. I found it difficult to interpret the biological relevance of the splice variants and identify the ones that were actually expressed in the cell clones.

The analysis of the RNA-seq data from the MCF-7 cell clones are still ongoing. At present the aim is to get one gene expression estimate per gene to compare with gene expression microarray estimates and thereby get an idea of how well the two methods agree. Also, we are working on a method to include the considerations regarding pseudogenes, gene families, overlapping genes and fusion genes in calculation of the abundances. In the long run we also aim at identifying the splice variants expressed and we hope that the analysis can lead to a pipeline for RNA-seq analysis for general use at CBS in the future.

Quantification and comparison

Normalization of RNA-seq data is as important as for microarrays. Systematic variability occurs primarily because fragmentation during library construction causes longer transcripts to produce more reads than shorter transcripts at the same expression level. Fluctations in the number of reads can also be an artifact of the sequencing technology (215). The read count is usually given as reads per kilobase of transcript per million mapped reads

[7] http://cufflinks.cbcb.umd.edu/

(RPKM) that is normalized by transcript length and total reads mapped in the sample (216).

$$\text{RPKM} = \frac{\text{total exon reads}}{\text{mapped reads (millions)} \times \text{exon length (kilobases)}} \tag{5.3}$$

Abundances for paired-end data are given in the unit fragments per kilobase of transcript per million mapped reads (FPKM), which is analogous to RPKM but takes into account the dependency of the paired-end reads (214). Once the RNA-seq data is normalized and quantified many of the considerations regarding statistical testing and multiple testing correction are the same as for gene expression microarrays (section 5.1.3).

5.3 Microarrays versus RNA-seq

By now, microarray technology is well established and used routinely. How well gene expression microarrays detect the actual levels of mRNAs is not well studied, because much of the generated data is published without validation by other methods such as quantitative reverse transcription polymerase chain reaction (QRT-PCR) and Northern blot analysis. Yet, since gene expression profiling is most often performed as a comparison between conditions the data should give a good estimate of the differences (179). Limitations of the microarrays are that only a limited number of probes fit on a microarray and the resolution is therefore limited (178). Also, microarrays only measure cDNA that is complementary to the probes on the array and there are limitations of the dynamic range of gene expression values detected because of probe saturation (138).

RNA-seq is a more unbiased analysis compared to microarrays as gene expression estimates are not dependent or restricted by annotation of the organism and a predetermined probe set (138; 142). Sequencing has no upper limit for quantification and theoretically a larger dynamic range than microarrays (217). Second generation sequencing is based on PCR amplification that can introduce errors and biases in the library (203). This problem can be avoided if sequencing could be performed directly on the single molecule without the need for amplification, which is one of the aims of third generation sequencing (194). Beneficial improvements would also be longer read lengths so the method would allow sequencing of full length mRNAs and methods to sequence all RNAs in one experiment. This will provide further insight into the transcriptome, regulation of expression and alternative splicing (178).

Part II

Projects

Chapter 6

OCT4-regulatory network

6.1 Prelude

The following paper is in an early stage of preparation and therefore only included as an abstract. I have chosen to include this project in the thesis, because it exemplifies the integrative systems biology approach very well in the aim to gain knowledge of functional modules in cells.

The data used in the project was generated at The Max Planck Institute for Molecular Genetics, Berlin, Germany. I received the data after a preliminary analysis and performed the integrative systems biology analysis. In the following section is a short abstract of the project and the data analysis is outlined in section 4.2.

6.2 Abstract

The human body consists of several hundred different cell types that carry out the biological functions necessary for its existence (218). The human body develops from ESCs, which are found in the inner cell mass of the blastocyst that is present a few days after fertilization of the egg. ESCs are pluripotent cells with the potential of differentiating to any of the three embryonic germ layers: endoderm, mesoderm and ectoderm (164). An essential factor for ESC pluripotency is the transcription factor OCT4, ESCs without OCT4 lack pluripotency (162). Yet, the regulatory network of OCT4 and its regulatory roles in ESC differentiation are still relatively unknown.

The aim of the present project was to identify the genes and processes that are regulated by OCT4 in ESCs based on an integrative systems biology analysis of three datasets generated on human cells: OCT4 ChIP-chip data from ESCs, ESC gene expression data from an OCT4 silencing (72h)

study, and gene expression data from a time series of EB differentiation. EBs arise from culturing ESCs derived from blastocysts (163). EB differentiation begins by formation of an aggregate of ESCs whereafter the cells start to differentiate to all three germ layers. Thus, EBs provide a good *in vitro* culture to study early differentiation events (165).

From the analysis described in detail in section 4.2, we identified a highly interconnected PPI network of OCT4-regulated proteins that appear to be important for ESC differentiation. The network was based on 94 primary targets of OCT4 identified by a ChIP-chip study, which also were differentially expressed during EB differentiation. The PPI network generated by InWeb (described in section 4.1.3) contained 70 of the input proteins and 48 added interaction partners. Gene Ontology analysis confirmed that the proteins in the network were relevant for ESC differentiation as the most statistically significant biological process categories were multicellular organismal development, developmental process and anatomical structure morphogenesis.

Each of the proteins in the network were tested for whether they had more interactions with proteins encoded by genes that changed gene expression during EB differentiation than expected by random. After multiple correction testing, 32 proteins were significantly associated with proteins encoded by regulated genes. Eight of the 32 proteins in the network were found of particular interest as they also changed in gene expression after OCT4 silencing, a dataset that was used as a third layer of evidence.

Some of the proteins in the network will be experimentally validated at The Max Planck Institute for Molecular Genetics. Their preferred candidates were soluble factors, which were identified by Gene Ontology analysis as proteins that were part of the extracellular matrix. Four of the eight most interconnected and differentially regulated genes that encode proteins in the network was found as part of the extracellular matrix: APOA1, APOA2, AHSG and FGB. These targets are the top candidates for experimental validation. The project will be published once the study has been validated.

Chapter 7

Androgen receptor signaling in the testis

7.1 Prelude

Testosterone is produced by Leydig cells already from prenatal Leydig cells during embryogenesis and through adulthood (20). Testosterone is essential for proper testis development and spermatogenesis and works via its receptor, AR, that is expressed predominantly in Sertoli, PTM, Leydig and perivascular smooth muscle cells (12). In the project presented next, we investigated the role of AR signaling in PTM cells and its importance for proper Leydig cell development and function. We found that PTM AR signaling is essential for normal Leydig cell development and that lack of PTM AR signaling led to development of two different subpopulations of Leydig cells: one 'normal' adult Leydig cell pool and another Leydig cell pool that seemed arrested in development.

The project was carried out during my external stay at MRC Human Reproductive Sciences Unit, Centre for Reproductive Biology, The Queen's Medical Research Institute, Edinburgh, UK. The basis for the study was knock out mice that had their AR knocked out specifically in their PTM cells. Hence, the AR and testosterone action are normal in other testis cells and the effects we detected were solely caused by disturbed AR signaling in PTM cells. The study was exclusively experimental and I participated in mice handling, the IHC analyses and QRT-PCR experiments.

The manuscript was published in *International journal of andrology* in 2011.

international journal of andrology ISSN 0105-6263

ORIGINAL ARTICLE

Androgen receptor signalling in peritubular myoid cells is essential for normal differentiation and function of adult Leydig cells

M. Welsh,* L. Moffat,* K. Belling,† L. R. de França,‡ T. M. Segatelli,‡ P. T. K. Saunders,* R. M. Sharpe* and L. B. Smith*

*MRC Human Reproductive Sciences Unit, Centre for Reproductive Biology, The Queen's Medical Research Institute, Edinburgh, UK, †Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Kemitorvet, Lyngby, Denmark, and ‡Laboratory of Cellular Biology, Department of Morphology, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil

Keywords:
androgen receptor, hormone receptors, hormones, Leydig cell, peritubular cells, steroidogenic enzymes, testosterone

Correspondence:
Lee B. Smith, MRC Human Reproductive Sciences Unit, Centre for Reproductive Biology, The Queen's Medical Research Institute, 47 Little France Crescent, Edinburgh EH16 4TJ, UK. E-mail: l.smith@hrsu.mrc.ac.uk

Received 9 September 2010; revised 6 January 2011; accepted 11 January 2011

doi:10.1111/j.1365-2605.2011.01150.x

Summary

Testosterone synthesis depends on normal Leydig cell (LC) development, but the mechanisms controlling this development remain unclear. We recently demonstrated that androgen receptor (AR) ablation from a proportion of testicular peritubular myoid cells (PTM-ARKO) did not affect LC number, but resulted in compensated LC failure. The current study extends these investigations, demonstrating that PTM AR signalling is important for normal development, ultrastructure and function of adult LCs. Notably, mRNAs for LC markers [e.g. steroidogenic factor 1 (*Nr5a1*), insulin-like growth factor (*Igf-1*) and insulin-like factor 3 (*Insl3*)] were significantly reduced in adult PTM-AR-KOs, but not all LCs were similarly affected. Two LC sub-populations were identified, one apparently 'normal' sub-population that expressed adult LC markers and steroidogenic enzymes as in controls, and another 'abnormal' sub-population that had arrested development and only weakly expressed INSL3, luteinizing hormone receptor, and several steroidogenic enzymes. Furthermore, unlike 'normal' LCs in PTM-ARKOs, the 'abnormal' LCs did not involute as expected in response to exogenous testosterone. Differential function of these LC sub-populations is likely to mean that the 'normal' LCs work harder to compensate for the 'abnormal' LCs to maintain normal serum testosterone. These findings reveal new paracrine mechanisms underlying adult LC development, which can be further investigated using PTM-ARKOs.

Introduction

Testosterone is essential for normal male fertility, controlling development of the male reproductive system and the later initiation and maintenance of spermatogenesis (reviewed in Sharpe, 1994; McLachlan *et al.*, 2002), but how this is effected remains largely unknown. Testosterone is produced by the testicular Leydig cells (LC) and binds to the androgen receptor (AR) to modulate gene transcription in target cells (Quigley *et al.*, 1995). Postnatally, testosterone production is driven by luteinizing hormone (LH), which binds to the LH receptor (LHR) on the LCs (Huhtaniemi & Toppari, 1995). Testosterone is synthesized from cholesterol, which can be imported into the cell from the circulation or made de novo (Scott *et al.*, 2009); these processes involve scavenger receptor b1 (Scarb1) and steroidogenic acute regulatory protein (StAR) or 3-hydroxy-3-methylglutaryl-coenzyme A synthase 1 (HMGCS1) and reductase (HMGCR1), respectively. Cholesterol is converted into pregnenolone and then into 5-dehydroepiandrosterone (DHEA); this process is controlled by StAR, p450 side-chain cleavage (termed P450scc or cyp11a1) and p450c17/CYP17A1 (Miller, 1998; Handelsman, 2008). Within the gonads, DHEA is then converted into testosterone or oestradiol by 3β-hydroxysteroid dehydrogenase (3β-HSD) and 17β-HSD or aromatase, respectively.

Testosterone synthesis depends on the normal development and differentiation of the LCs, but this process is poorly understood. There are two phases of LC development, which produce two generations of LCs, foetal and adult. Foetal LCs develop in embryogenesis and persist into early postnatal life when they become inactive and involute (Zhang *et al.*, 2001; Haider, 2004). Adult LCs develop from LC precursor stem cells around postnatal d7–10 in mice, when they begin to proliferate and subsequently transform into progenitor cells (Wu *et al.*, 2010; Vergouwen *et al.*, 1991; Nef *et al.*, 2000). At this stage, precursor cells lose their spindle shape and proliferative capacity and enlarge. These progenitor cells then differentiate beyond d21 in mice (Wu *et al.*, 2010) and increase expression of 3β-HSD first, and latter cyp17a1 and cyp11a1 (Zhang *et al.*, 2004). Onset of these steroidogenic genes is regulated by steroidogenic factor 1 (SF-1; Ikeda *et al.*, 1994; Hiroi *et al.*, 2004) and occurs by d35 in mice when immature LCs can be identified (Wu *et al.*, 2010; Hardy *et al.*, 1989; Baker & O'Shaughnessy, 2001b; Habert *et al.*, 2001). Unlike mature adult LCs, immature LCs have numerous cytoplasmic lipid droplets. Immature LCs then undergo a final cell division around d45 to become mature adult LCs (Wu *et al.*, 2010). A few precursor LCs are thought to persist in normal adult testes but, unlike mature adult LCs, they do not express functional LHRs (Shan & Hardy, 1992).

The mechanisms controlling adult LC differentiation are poorly understood, but several factors have been suggested to be involved. These include factors such as desert hedgehog (*Dhh*; Clark *et al.*, 2000; Yao *et al.*, 2002), insulin growth factor 1 (*IGF-1*; Khan *et al.*, 1992), and platelet-derived growth factor alpha (*PDGFα*; Gnessi *et al.*, 2000), and hormones such as insulin-like factor 3 (INSL3; Ferlin *et al.*, 2009) and LH (Baker *et al.*, 2003). For example, initiation of adult LC differentiation is LH-independent, but LH is required later for the steps beyond progenitor cell differentiation (Baker *et al.*, 2003; Mendis-Handagama *et al.*, 2007). Androgens are also important for normal LC development (Murphy *et al.*, 1994); for example, LC numbers are reduced in adult AR knockout mice and their function is impaired (O'Shaughnessy *et al.*, 2002; De Gendt *et al.*, 2005). LC number is also reduced in sertoli cell (SC)-specific AR knockout (SCARKO) mice (De Gendt *et al.*, 2005), which highlights a role for SC-produced paracrine signals in LC development. However, LC function is not impaired in SCARKO mice (De Gendt *et al.*, 2005) raising the possibility that androgens act via cells other than SCs to regulate LC function. We recently demonstrated that AR ablation from some peritubular myoid cells (PTM-ARKO) had no effect on LC number, but resulted in compensated LC failure, despite normal SC and LC AR expression, and high intra-testicular testosterone (Welsh *et al.*, 2009). The aim of the current study was to investigate LC development and functional differentiation in these PTM-ARKO mice to identify the basis of this altered LC function. These studies uncovered evidence for profound effects of PTM cell AR signalling on postnatal LC development.

Materials and methods

Breeding of transgenic mice

Mice in which the AR was selectively ablated from the PTM cells were previously generated using *Cre/loxP* technology (Welsh *et al.*, 2009); male mice heterozygous for *Cre* recombinase under the control of a smooth muscle myosin heavy chain (MH; Xin *et al.*, 2002) promoter were mated to female mice homozygous for a floxed AR (De Gendt *et al.*, 2004). The *Cre*-positive (AR^{flox} positive) male offspring from these matings are termed PTM-ARKO, whereas the *Cre*-negative AR^{flox} positive littermates were used as controls. All mice were bred under standard conditions of care and use under licensed approval from the UK Home Office. Mice were genotyped from ear or tail DNA for the presence of *Cre* using standard PCR (http://jaxmice.jax.org/pub-cgi/protocols/protocols.sh?objtype=protocol&protocol_id=288); all male offspring were hemizygous for X-linked AR^{flox}.

In vivo treatments

Exposure to exogenous testosterone has been previously shown to inhibit endogenous testosterone production by suppressing LH secretion, leading to involution of LCs. Adult (d100) male PTM-ARKO and control mice ($n = 5$) were implanted subcutaneously in the upper back with 1 cm silastic implants filled with testosterone (or 'empty' sham implants as controls) to investigate the LC involution response in PTM-ARKO testes. Implants were left in situ for 13 weeks, then mice were culled and reproductive tissues and blood were collected for analyses as detailed next.

Recovery of testes

Male mice were culled at various postnatal ages (d12–d300) by inhalation of carbon dioxide and subsequent cervical dislocation. Testes were removed from the mice, weighed and either snap-frozen for subsequent RNA analysis or fixed in Bouins for 6 h. Bouin-fixed tissues were processed and embedded in paraffin wax and cut into 5-μm sections for histological analysis as reported previously (Welsh *et al.*, 2006). Sections of testis were stained with haematoxylin and eosin using standard protocols and examined for histological abnormalities. Testes were

recovered also from adult PTM-ARKO and control males (d160–200, $n = 3$) and fixed for 24 h in 4% paraformaldehyde for Oil Red O staining for lipid droplets. Testes were washed in water before cutting 10-μm frozen sections on a cryotome. Sections were mounted on slides and stained for Oil Red O using standard staining protocols.

Hormone analysis

Immediately after culling, blood was collected from testosterone-treated and 'empty' implant mice by cardiac puncture. Sera were separated and stored at −20 °C until assayed. LH and testosterone were measured using previously published assays (Corker & Davidson, 1978; McNeilly *et al.*, 2000). Intra-testicular testosterone concentrations were measured as published previously (Fisher *et al.*, 2003). All samples from each mouse were run in a single assay for each hormone, and the within-assay coefficients of variation were all <10%.

Determination of testicular cell composition

Standard stereological techniques involving point counting of cell nuclei were used as described (De Gendt *et al.*, 2004), to determine the nuclear volume per testis of each population of LCs, namely normal adult LCs strongly immunostained for 3β-HSD or abnormal LCs weakly positive for 3β-HSD. Briefly, cross-sections of testes from four to six KO or control mice at d12 and d100 were stained for 3β-HSD (detailed next) and examined using a Leitz 363 Plan Apo objective (×63) fitted to a Leitz Laborlux microscope (Leica Microsystems, Wetzlar, Germany) and a 121-point eyepiece graticule. For each animal, 32–64 microscopic fields were counted, and values for percentage nuclear volume were converted into absolute nuclear volumes per testis by reference to testis volume (=weight). LC nuclear size was determined using an Olympus Optical BH-2 microscope fitted with a Prior automatic stage (Prior Scientific Instruments, Cambridge, UK) and Image-Pro Plus version 4.5.1 with Stereologer-Pro 5 plug-in software (Media Cybernetics, Bethesda, MD, USA). Data were used to determine the nuclear volumes of and number of LCs per testis at d12 and d100.

Immunohistochemical analysis

Three immunohistochemical detection methods were used: (i) fluorescent immunostaining, (ii) colorimetric staining with streptavidin-HRP and DAB or (iii) colorimetric staining using a Bond-X automated immunostaining machine (Vision Biosystems, Newcastle, UK). For all methods, sections were deparaffinized, rehydrated and antigen-retrieved as detailed previously (Welsh *et al.*, 2006). For methods (i) and (ii) above, non-specific binding sites were blocked, sections were incubated with the primary antibody diluted accordingly (see Table 1), and immunostaining was detected using the secondary antibody and detection system specified in Table 1. Details of these methods have been published previously (Welsh *et al.*, 2009). For method (iii) above, a specific polymer high-contrast programme was used on a Bond-X automated immunostaining machine; briefly, slides were peroxidase blocked for 5 min, incubated for 2 h with the primary antibody diluted to the optimal concentration (detailed in Table 1) in the diluent supplied and then incubated with the post-primary reagent for 15 min. Control sections were incubated with diluent alone to confirm antibody specificity. Sections were then incubated with the polymer reagent for 15 min to increase sensitivity of detection prior to DAB detection for 10 min. Exact conditions were optimized for each antibody and all kits were purchased from Vision Biosystems. DAB-immunostained slides were counterstained with haematoxylin, dehydrated and mounted with Pertex (Histolab, Gothenburg, Sweden), and images were captured using a Provis microscope (Olympus UK Ltd, Southend-

Table 1 Immunohistochemistry antibody details

Antibody	Antibody source	Dilution	Detection system
3β-HSD/AR			
3β-HSD	Santa Cruz (Santa Cruz, USA)	1 : 4000	Tyramide 488
AR	Santa Cruz	1 : 50	Goat anti-rabbit Alexa 546
LHR/3β-HSD			
LHR	Santa Cruz	1 : 200	Tyramide 488
3β-HSD	Santa Cruz (Santa Cruz, USA)	1 : 4000	Tyramide 633
SF-1	Upstate	1 : 1500	Streptavidin-HRP, DAB
IGF-1	Abcam	1 : 3	Streptavidin-HRP, DAB
CYP17A1	Santa Cruz	1 : 2000	Bond-automated polymer system
CYP11A1a	Chemicon	1 : 1000	Bond-automated polymer system
3β-HSD	Santa Cruz	1 : 1000	Bond-automated polymer system
INSL3	Gift from Steven Hartung	1 : 300	Bond-automated polymer system

3β-HSD, 3β-hydroxysteroid dehydrogenase; AR, androgen receptor; LHR, luteinizing hormone receptor; SF-1, steroidogenic factor 1; IGF-1, insulin-like growth factor; CYP17A1, cytochrome p450 17; CYP11A1a, cytochrome p450 11a; INSL3, insulin-like factor 3; HRP, Horseradish Peroxidase; DAB, 3,3′-Diaminobenzidine.

on-Sea, UK) equipped with a Kodak DCS330 camera. Fluorescent immunostained sections were mounted in Mowiol mounting medium (Calbiochem, San Diego, CA, USA) and fluorescent images were captured using a Zeiss LSM 510 Meta Axiovert 100 M confocal microscope (Carl Zeiss Ltd., Welwyn, UK). To ensure reproducibility of results, representative testes from at least three animals at each age were used, and sections from PTM-ARKO and control littermates were processed in parallel on the same slide on at least two occasions. Appropriate negative controls were included to ensure that any staining observed was specific. All antibodies used showed only minor non-specific staining.

LC ultrastructure in adult control and PTM-ARKO mice

The testes of four control and four d100 PTM-ARKO mice were perfusion-fixed with 4% (vol/vol) glutaraldehyde in 0.1 M cacodylate buffer (pH 7.3) followed by a brief saline wash. The testes were then diced into small pieces, placed into the same fixative for 1 h, washed in cacodylate buffer overnight, post-fixed with 1% (wt/vol) osmium/1.25% (wt/vol) potassium ferrocyanide, dehydrated in ethanol and embedded in Araldite (CY 212). Thin sections were prepared from each testis, mounted on 200-mesh grids, stained with uranyl acetate and lead citrate, and examined on an electron microscope Tecnai – G2-20-FEI (FEI, Hillsboro, OR, USA).

RNA extraction and reverse transcription

RNA was isolated from frozen testes from PTM-ARKO or control mice using the RNeasy Mini extraction kit with RNase-free DNase on the column digestion kit (Qiagen, Crawley, UK) according to the manufacturer's instructions. For quantitative RT-PCR, 5 ng Luciferase mRNA (Promega Corp., Madison, WI, USA) was added to each testis sample before RNA extraction as an external standard (Tan *et al.*, 2005). RNA was quantified using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Random hexamer primed cDNA was prepared using the Applied Biosystems TaqMan reverse transcription kit (Applied Biosystems, Foster City, CA, USA) according to manufacturers' instructions.

Quantitative analysis of gene expression

Quantitative PCR was performed on d100 PTM-ARKO and control testes for the genes listed in Table 2, using an ABI Prism 7500 Sequence Detection System (Applied Biosystems) and the Roche Universal Probe library (Roche, Welwyn, UK), as described previously (Welsh *et al.*, 2009). The expression of each gene was related to an external positive control luciferase, as published previously (Baker & O'Shaughnessy, 2001a; De Gendt *et al.*, 2005; Tan *et al.*, 2005; Welsh *et al.*, 2009), and all genes were expressed per testis; as LC number is not signifi-

Table 2 Taqman primer details

Gene	Forward primer	Reverse primer
Steroidogenic acute regulatory protein (*StAR*)	ttgggcatactcaacaacca	acttcgtccccgttctcc
3β-hydroxysteroid dehydrogenase type 1 (*3β-HSD1*)	tgtgaccatttcctacattctga	ccagtgattgataaaccttatgtcc
3β-hydroxysteroid dehydrogenase type 6 (*3β-HSD6*)	accatccttccacagttctagc	acagtgaccctggagatggt
17β-hydroxysteroid dehydrogenase type 3 (*17β-HSD*)	aatatgtcacgatcggagctg	gaagggatccggttcagaat
Cytochrome p450 11a (*cyp11a1* or *p450scc*)	aagtatggccccatttacagg	tggggtccacgatgtaaact
Cytochrome p450 17 (*cyp17a1* or *17aOH*)	catcccacacaaggctaaca	cagtgcccagagattgatga
Cytochrome p450 21a1 (*cyp21a1)*	ccaacctggatgagatggtt	ggattcttcccaggttccag
Oestrogen sulphotransferase (EST)	tcccagaatagtaaaaactcacctg	gcgttccggcaaagatag
HMGCS1	cagggtctgatccccttg	cagagaactgtggtctccaggt
HMGCR1	tgcgtaagcgcagttcct	ttgtagcctcacagtccttgg
Scavenger receptor b1 (*Scarb1*)	atggtgccctccctcatc	acaggctgctcgggtctat
Steroidogenic factor 1 (*SF-1*)	tccagtacggcaaggaaga	ccactgtgctcagctccac
Insulin-like factor 3 (*Insl3*)	aagaagccccatcatgacct	tttatttagacttttgggacacagg
Insulin-like growth factor (*IGF-1*)	agcagccttccaactcaattat	gaagacgacatgatgtgtatctttatc
Insulin-like growth factor binding protein 3 (*IGFBP3*)	gcagcctaagcacctacctc	tcctcctcggactcactgat
Desert hedgehog (*Dhh*)	cacgtatcggtcaaagctgat	gtagttccctcagccccttc
GLI-Kruppel family member (*Gli1*)	ctgactgtgcccgagagtg	cgctgctgcaagaggact
Platelet-derived growth factor alpha (*PDGFα*)	tccaacctgaacccagacc	gccggctctatctcacctc
Kallikrein-1-related peptidase b21 (*Klk1b21*)	gcagcattacacccacgaa	attaggcaggggcttgatg
Kallikrein-1-related peptidase b24 (*Klk1b24*)	gtcctgttgaaccccaactg	tttgcccagccaaacatta
Kallikrein-1-related peptidase b26 (*Klk1b26*)	ctgtccctaggagggattga	tcacagttaaatcctccaacca
Kallikrein-1-related peptidase b27 (*Klk1b27*)	cccaactgggttctcacag	tttgcccagccaaacatta

cantly different in PTM-ARKO adult testes compared with controls, this should reflect gene expression per LC.

Statistical analysis

Data were analysed using GraphPad Prism version 5 (Graph Pad Software Inc., San Diego, CA, USA) using a two-tailed unpaired *t*-test or a one-way ANOVA followed by Bonferroni post hoc tests. Values are expressed as mean ± SEM. Normality was confirmed using D'Agostino and Pearson omnibus normality test.

Results

LC function is altered in PTM-ARKO adult testes

We previously identified compensatory adult LC failure in PTM-ARKO testes in which AR is ablated from around 40% of the PTM cells (Welsh *et al.*, 2009); to understand better the mechanisms underlying this, expression of genes involved in testosterone biosynthesis was examined at various ages. There was no effect on the expression pattern of any of the LC markers or steroidogenesis enzymes in PTM-ARKO testes at d1 (data not shown) suggesting that foetal LCs develop normally. Conversely, at d12, before any testicular histological abnormalities were obvious, there was a significant increase in *3β-HSD1* and *17β-HSD3* mRNA in PTM-ARKO testes compared with controls; expression of the other genes involved in steroidogenesis was not affected (Fig. 1B). At d100, there was a significant reduction in testicular expression of mRNAs encoding *HMGCS1* and *HMGCR1*, both of which are involved in de novo cholesterol synthesis, and in *3β-HSD1 and 3β-HSD6*, compared with age-matched controls. Conversely, there was a significant increase in the concentration of mRNA for cyp17a1 in PTM-ARKO d100 testes compared with controls (Fig. 1A). There was a reduction in *SF-1* mRNA expression in PTM-ARKO testes at d12, d50 and d100 compared with age-matched controls (Fig. 1C) and this reduction was confirmed at the protein level by immunohistochemistry (Fig. 1D). Interestingly, at d100, some LC nuclei stained positive for SF-1 in PTM-ARKO testes, whereas others were less positive or even negative for SF-1; both cell populations were located in the same interstitial spaces and were identified by a mouse pathologist as steroidogenic cells. This suggested that there might be two sub-populations of LCs in the adult PTM-ARKO testis; one which expressed all the normal LC markers examined, termed 'normal' throughout, and one in which normal LC markers were only weakly expressed/absent, termed 'abnormal' throughout the manuscript.

Evidence of two sub-populations of LCs in adult PTM-ARKO

At d100, two sub-populations of LCs could be identified in PTM-ARKO testes compared with controls (Fig. 2A, right panels); the 'normal' sub-population expressed CYP11A1 (data not shown), CYP17A1 (Fig. 2A) and 3β-HSD (Fig. 2A) as in control testes, whereas staining for these steroidogenic enzymes was weaker or even absent in the 'abnormal' population. These 'abnormal' LCs were interspersed amongst the 'normal' LCs expressing normal LC markers (Fig. 2A). This differential immunostaining pattern could be identified at d35 (Fig. 2A, left panels), but not at d21 (data not shown). Furthermore, immunostaining for 3β-HSD and LHR revealed that unlike in the control testes, some 3β-HSD-positive LCs did not express the mature adult LC marker LHR in PTM-ARKO d100 testes (Fig. 2B); these cells were predominantly the LCs which were less immunopositive for 3β-HSD (i.e. the 'abnormal' population). AR could be detected in LCs in control testes, as expected, and in both sub-populations of LCs in PTM-ARKO testes (Fig. 2C). LCs appeared more lipid-filled in PTM-ARKO testes, an observation confirmed by staining with Oil Red O (Fig. 2C). This staining suggested that there may be more lipid-filled interstitial cells in KO testes than in controls at d100 and that each lipid-filled cell stained more intensely with Oil Red O (Fig. 2D).

Confirmation of two sub-populations of LC in PTM-ARKO testes by electron microscopy

Typical adult mouse LC ultrastructures were identified in adult control mice (Christensen & Fawcett, 1966), such as considerable smooth endoplasmic reticulum in whorl form (WER) that encompasses centrally located cytoplasmic components, and continues with cylindrical bodies (CB; Fig. 3A,C,E). Interestingly, WER and CB and ordinary smooth endoplasmic reticulum were rarely observed in any PTM-ARKO LCs (Fig. 3B,D,F). Furthermore, the two different LC sub-populations described before could be identified by transmission electron microscopy. The first contained few lipid droplets, similar to control LCs, and a higher number of mitochondria (L1 in Fig. 3B,D,F), whereas the second LC sub-population had cytoplasm almost totally filled with lipid droplets (L2 in Fig. 3B,D). In comparison with control LCs (Fig. 3E, insert), mitochondria were obviously larger in both sub-populations of LCs in PTM-ARKO testes and were irregularly shaped with unevenly distributed tubular cristae (Fig. 3F, insert). In addition, the outer mitochondrial membrane was not evident in either PTM-ARKO LCs (Fig. 3E).

LC size and number

We have previously published that there was no significant change in LC size ($p = 0.34$) or number ($p = 0.35$) in PTM-ARKO mice at d100 compared with controls (Welsh *et al.*, 2009). However, in light of the identification of the two populations of LCs in adult PTM-ARKO testes, LC size and number were re-measured. This confirmed that there was no significant difference in LC size ($p = 0.74$) or overall number (Fig. 4) in d100 PTM-ARKO testes compared with controls. However, in the PTM-ARKO testes, total LC number comprised roughly equal numbers of 'normal' and 'abnormal' LCs (Fig. 4); these cells were distinguished by their histology and 3β-HSD expression. There was no significant difference in average LC size in PTM-ARKOs ($p = 0.27$) compared with controls. Furthermore, there was no significant difference in the average size of 'normal' LCs compared with

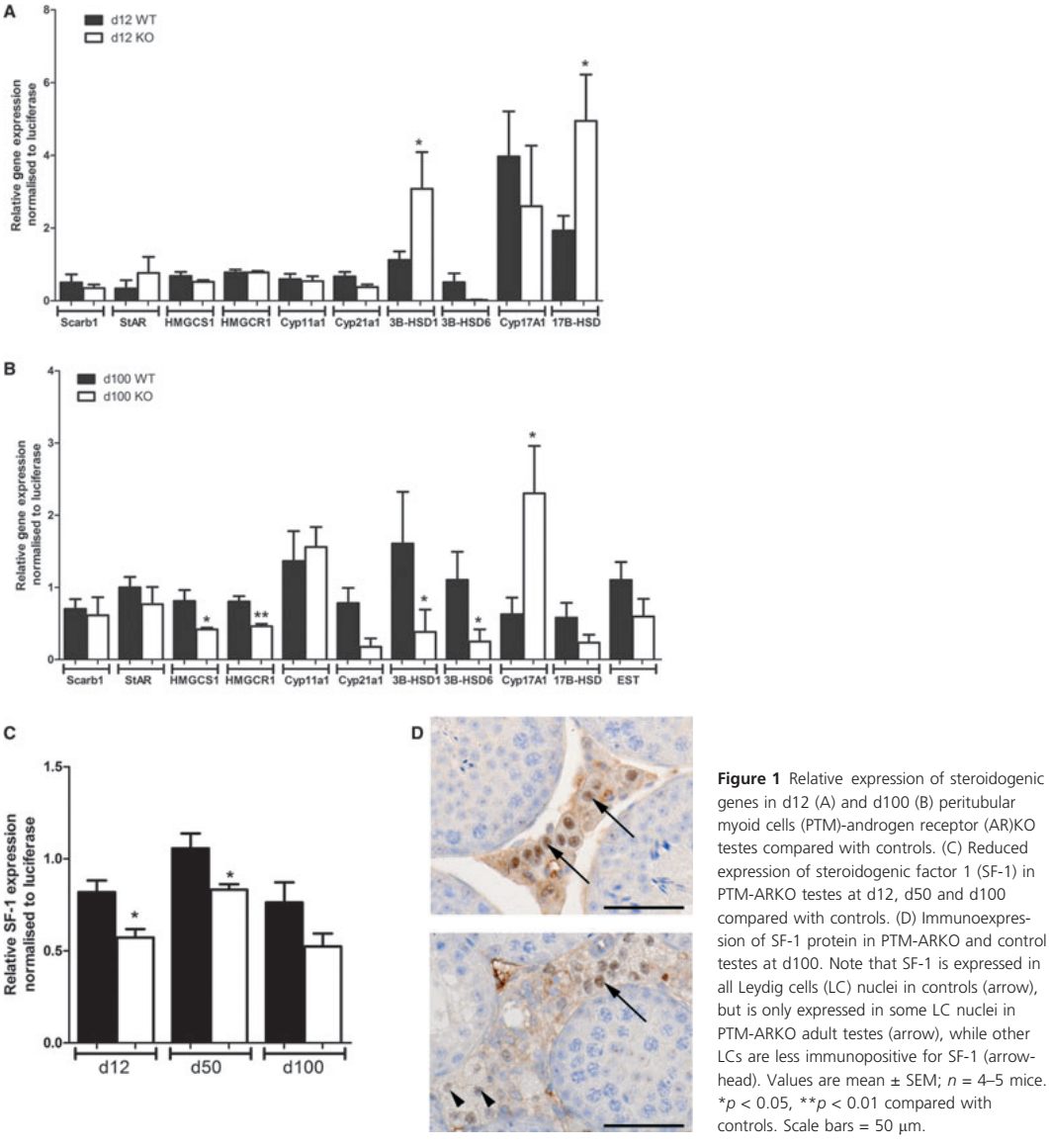


Figure 1 Relative expression of steroidogenic genes in d12 (A) and d100 (B) peritubular myoid cells (PTM)-androgen receptor (AR)KO testes compared with controls. (C) Reduced expression of steroidogenic factor 1 (SF-1) in PTM-ARKO testes at d12, d50 and d100 compared with controls. (D) Immunoexpression of SF-1 protein in PTM-ARKO and control testes at d100. Note that SF-1 is expressed in all Leydig cells (LC) nuclei in controls (arrow), but is only expressed in some LC nuclei in PTM-ARKO adult testes (arrow), while other LCs are less immunopositive for SF-1 (arrowhead). Values are mean ± SEM; $n = 4$–5 mice. *$p < 0.05$, **$p < 0.01$ compared with controls. Scale bars = 50 μm.

'abnormal' LCs within the PTM-ARKO testes ($p = 0.7$). Note that there was no significant difference in the number (Fig. 4) or size ($p = 0.8$) of 3β-HSD-positive LCs at d12.

Development of adult LCs is altered in PTM-ARKO testes

To gain insight into the origin of the two LC sub-populations, various markers of LC differentiation were examined. Expression of *Insl3* mRNA was significantly reduced in PTM-ARKO testes at d12, d21, d50 and d100 (Fig. 5A); this was confirmed by immunohistochemistry at d100 (Fig. 5B). Note that INSL3 protein was detected in all LCs in control d100 testes, but was only detected in some LCs in PTM-ARKO testes; the 'abnormal' LCs appeared less immunopositive or even negative for INSL3 in PTM-ARKO testes, whereas the 'normal' LCs expressed INSL3 similar to LCs in controls (Fig. 5B). IGF-1 signalling was also examined as it

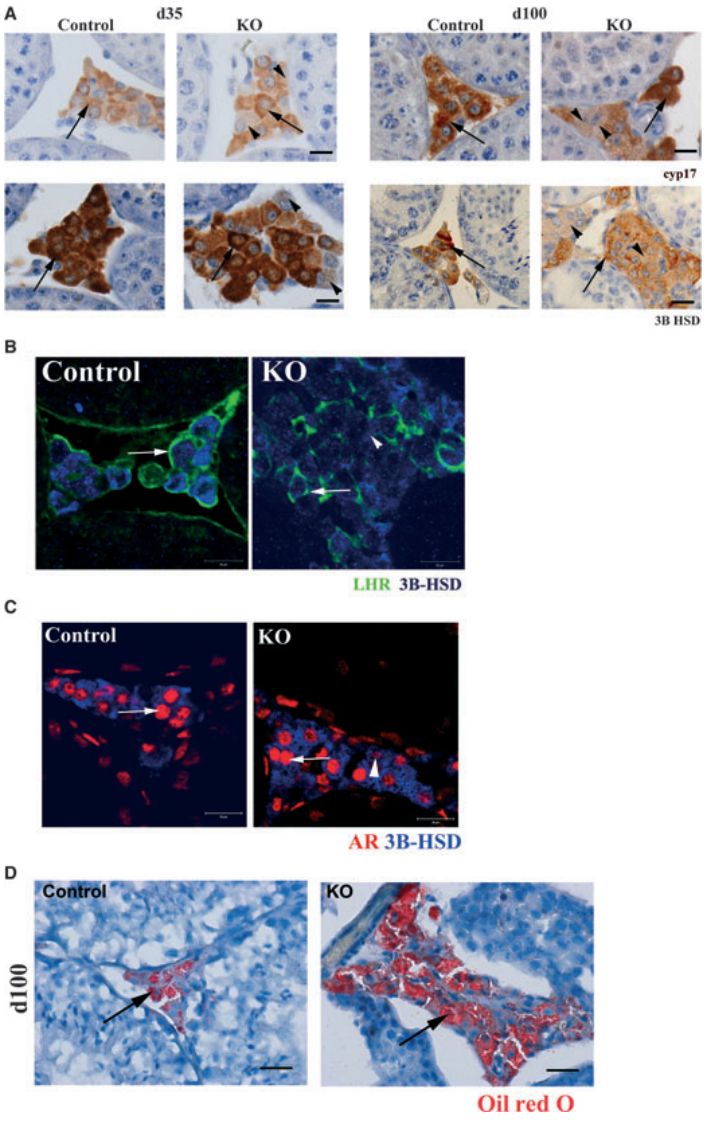


Figure 2 Immunoexpression of proteins involved in Leydig cell (LC) function suggests that there are two groups of LCs in peritubular myoid cells (PTM)-androgen receptor (AR)KO testes. (A) Immunoexpression of cytochrome p450 17 (CYP17A1) and 3β-hydroxysteroid dehydrogenase (3β-HSD) in d35 and d100 PTM-ARKO and control testes. Note that unlike in controls, at d100 some LCs (arrowhead) in PTM-ARKO testes appear less positive for CYP17A1, and 3β-HSD than other LCs (arrow). Scale bars = 20 μm. (B) Immunoexpression of luteinizing hormone receptor (LHR, green) and 3β-HSD (blue) in d100 PTM-ARKO and control testes. Note that all 3β-HSD-positive LCs were positive for LHR (arrow) in control testes, whereas only some 3β-HSD-positive cells in PTM-ARKO testes were positive for LHR (arrow), while others appeared not to express LHR (arrowhead). (C) Immunoexpression of androgen receptor (AR, red) and 3β-HSD (blue) in d100 PTM-ARKO and control testes. Note that AR is expressed in both 'normal' (arrow) and 'abnormal' (arrowhead) LCs in PRM-ARKO testes. (D) Staining of PTM-ARKO and control testes with Oil Red O revealed an increase in the amount of lipid (stained red, arrow) in LCs in PTM-ARKO d100 testes compared with controls. Scale bars = 50 μm.

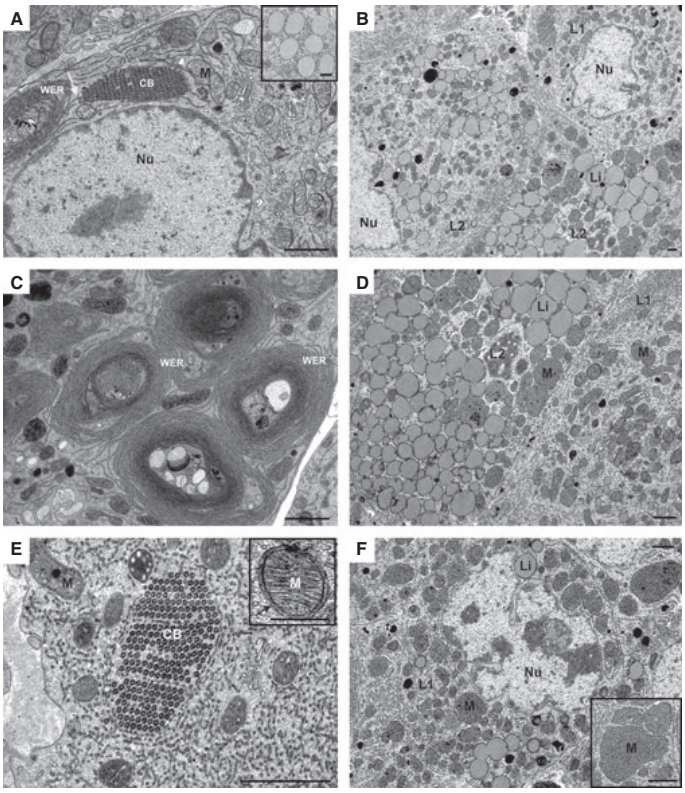


Figure 3 Leydig cell (LC) ultrastructure in d100 control (A, C and E) and PTM-ARKO (B, D and F) mice. Note typical mouse LC structures in the control such as smooth endoplasmic reticulum in whorl form (WER; A and C) that encompasses centrally located cytoplasmic components, and which continues (white arrow in A) with cylindrical bodies (CB; A and E). Both WER and CB were rarely observed in PTM-ARKO LCs. Two sub-populations of LCs could be identified in PTM-ARKO adult testes. The first (L1 in B, D and F) contained few lipid droplets (Li), as seen in controls (insert in A), and a higher amount of mitochondria (M); in the second population, the cytoplasm is almost totally filled with lipid droplets (L2 in B and D). In comparison with the wild type (see insert in E), in both LC populations found in the PTM-ARKO (insert in F), mitochondria were usually much larger, irregularly shaped and presented unevenly distributed tubular cristae. Also, the outer mitochondrial membrane (black arrow in the insert in E) was not evident in PTM-ARKO LC; Nu indicates nucleus (Nu). Bar = 1 μm (insert in E; bar = 0.5 μm).

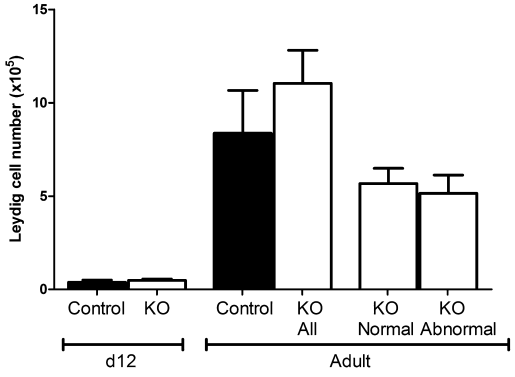


Figure 4 Leydig cell (LC) number in peritubular myoid cells (PTM)-androgen receptor (AR)KO and control testes at d12 and d100 (adult). 'Normal' LCs strongly immunostained for 3β-hydroxysteroid dehydrogenase (3β-HSD) whilst 'abnormal' LCs were only weakly positive for 3β-HSD. Values are mean ± SEM; n = 4–5 mice.

is reported to play a role in LC development (Khan *et al.*, 1992). *IGF-1* mRNA was significantly decreased at d12 in PTM-ARKO testes (Fig. 5C), whereas *IGFBP3* mRNA, an inhibitor of IGF-1, was significantly increased in PTM-ARKO testes at d12, d21, d50 and d100 compared with controls (Fig. 5D). It is known that kallikreins degrade and so reduce IGFBP3 expression (Schill & Miska, 1992; Matsui *et al.*, 2000, 2005; Matsui & Takahashi, 2001); *kallikreins 21* and *24* mRNA levels were both significantly reduced in PTM-ARKO testes at d12, d21 and d50, and *kallikrein 27* was significantly reduced at d12 and d21, compared with age-matched controls (Fig. 5E). Expression of IGF-1 was also examined by immunohistochemistry and confirmed that IGF-1 protein expression was reduced or absent in the 'abnormal' LCs in PTM-ARKO testes at both d35 and d100, while other 'normal' LCs within the same interstitial area demonstrated normal expression of IGF-1 protein (Fig. 5F). Expression of *Dhh* mRNA was significantly reduced in PTM-ARKO testes at d12 and d100 compared with controls (Fig. 6A). Furthermore, expression of GLI-Kruppel family member (*Gli1*) mRNA (Fig. 6B) and *PDGFα* (Fig. 6C) was also significantly reduced in PTM-ARKO testes at d50 and d100.

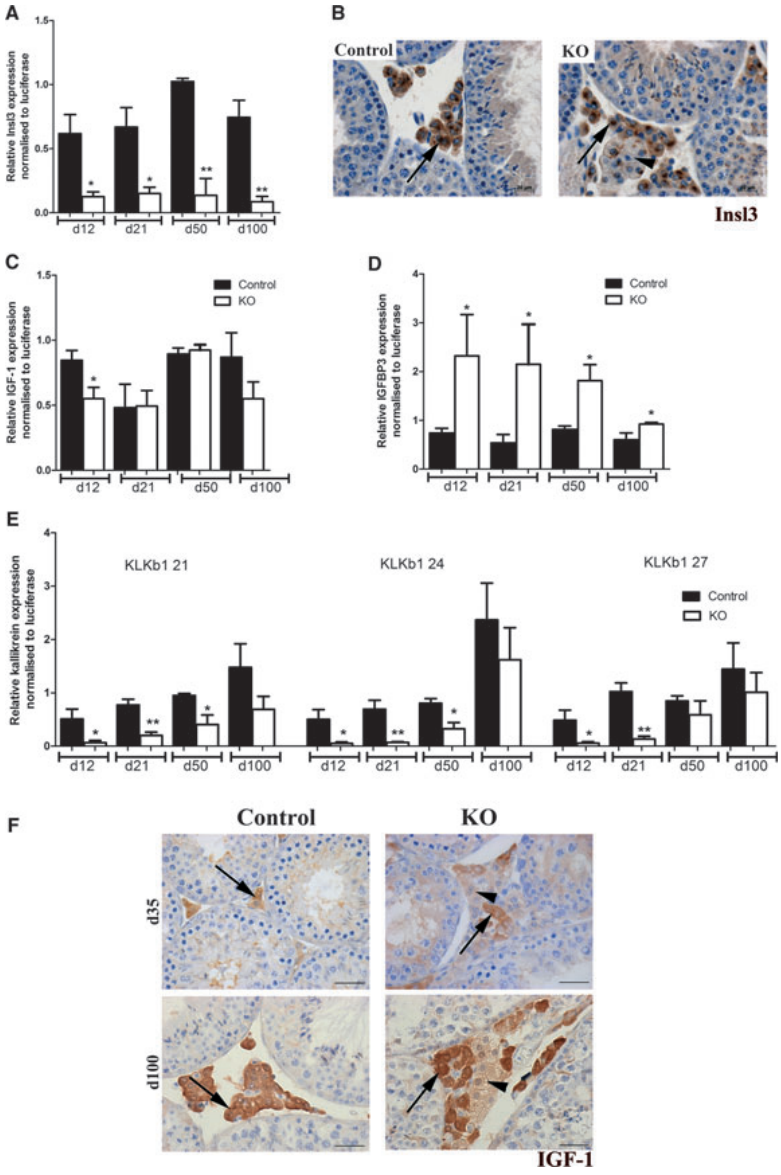


Figure 5 Expression of Leydig cell (LC) differentiation markers in peritubular myoid cells (PTM)-androgen receptor (AR)KO and control testes. (A) Relative expression of insulin-like factor 3 (*Insl3*) mRNA is significantly reduced in PTM-ARKO testes at d12, d21, d50 and d100 compared with controls. (B) INSL3 protein expression is differentially altered in PTM-ARKO adult testes; some LCs express INSL3 in KO testes (arrow), while other neighbouring LCs are negative for INSL3 (arrowhead). All LCs are INSL3 positive in control testes. (C) Expression of *IGF-1* mRNA in PTM-ARKO and control testes at d12–100. (D) Increased expression of *IGFBP3* mRNA in PTM-ARKO testes at d12–100, compared with controls. (E) Relative expression of *kallikreins 21, 24* and *27* in d12–100 control and PTM-ARKO testes. (F) Differentially altered expression of insulin-like growth factor (IGF-1) protein in PTM-ARKO testes compared with controls at d35 (top panels) and d100 (bottom panels). Note that some LCs in PTM-ARKO testes (arrow) express IGF-1 similar to that seen in controls (arrow), while others appear negative for IGF-1 (arrowhead). Values are mean ± SEM; $n = 4$–5 mice. $*p < 0.05$, $**p < 0.01$ compared with controls. Scale bars = 50 μm.

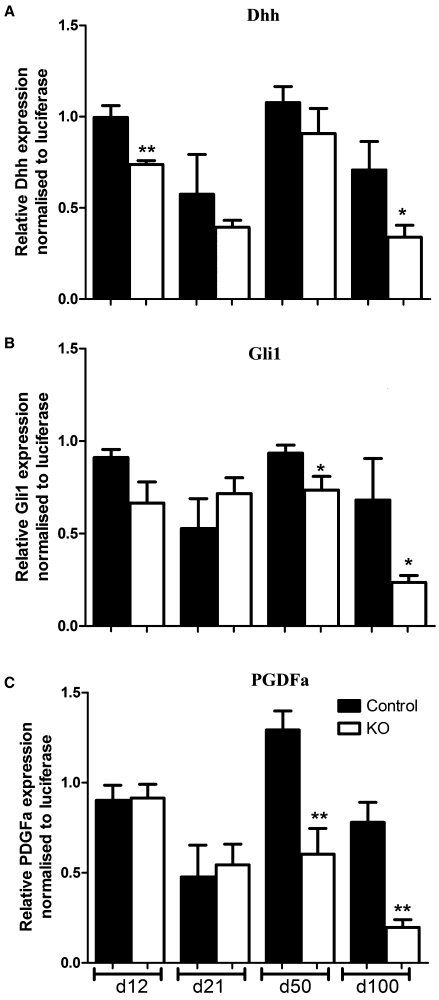


Figure 6 Relative expression of (A) *Dhh* (desert hedgehog), (B) *Gli1* (GLI-Kruppel family member) and (C) *PDGFα* (platelet-derived growth factor alpha) in peritubular myoid cells (PTM)-androgen receptor (AR)KO and control testes at d12, d21, d50 and d100. Values are mean ± SEM; $n = 4–5$ mice. $*p < 0.05$, $**p < 0.01$ compared with controls.

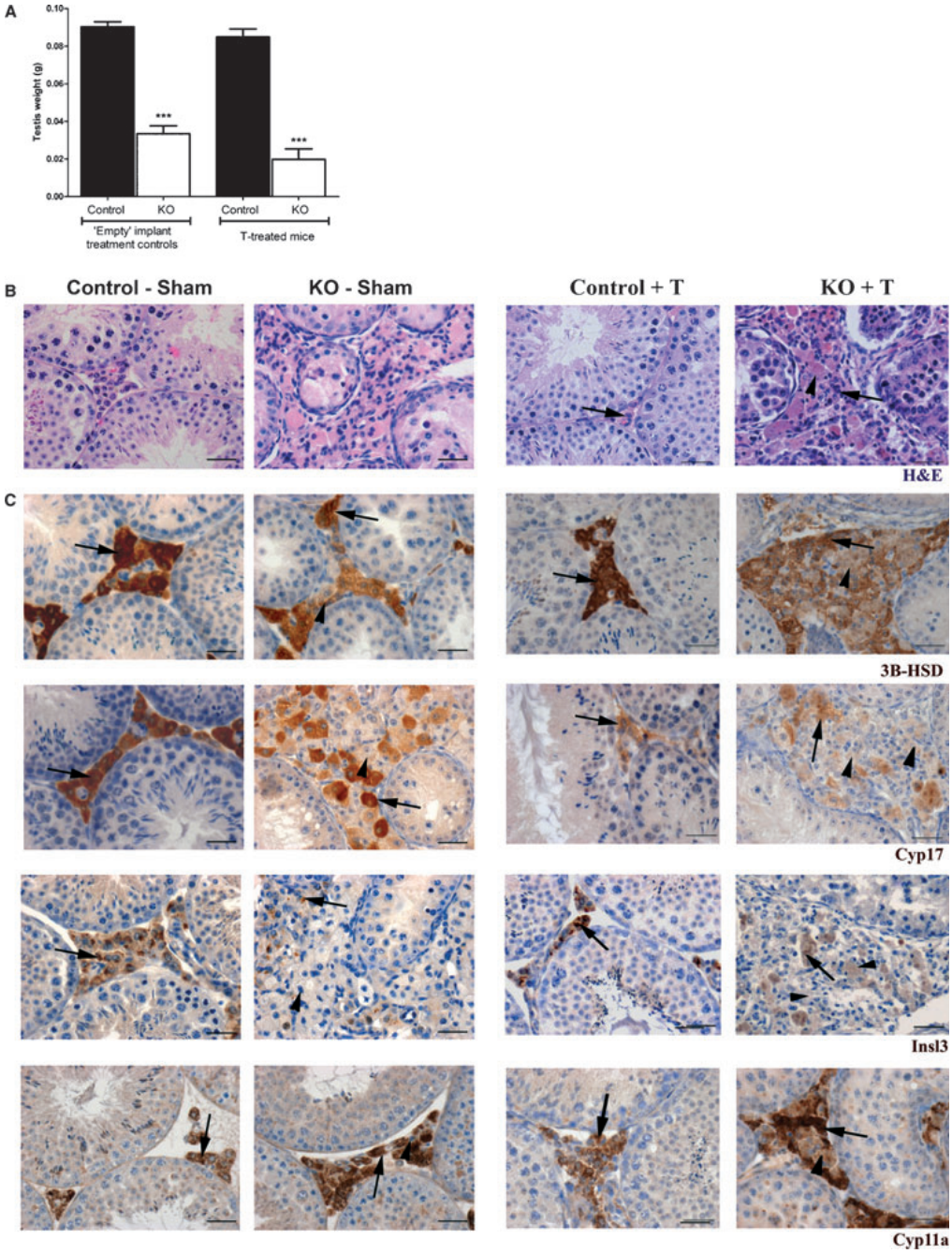
Abnormal LC response to exogenous testosterone treatment in vivo

Exposure to exogenous testosterone has been previously shown to inhibit endogenous testosterone production by suppressing LH secretion, leading to involution of LCs (Keeney *et al.*, 1988, 1990). In our study, exposure to exogenous testosterone for 13 weeks significantly reduced serum LH concentrations, but resulted in normal serum testosterone concentrations in both control and PTM-ARKO-treated males (data not shown). This treatment had no significant effect on testis weight in either PTM-ARKO or control mice (Fig. 7A) compared with age-matched 'empty' implant-treated PTM-ARKOs or controls, respectively. As expected, LCs involuted in testosterone-treated control testes (Fig. 7B, left panel), characterized by a reduction in LC size (Keeney *et al.*, 1988, 1990); conversely, only a proportion of LCs involuted in testosterone-treated PTM-ARKO testes in response to exogenous testosterone, whereas others remained large and lipid-filled (Fig. 7B, right panel). In testosterone-treated control testes, the small involuted LCs continued to express 3β-HSD, CYP17A1, INSL3 and CYP11A1 after testosterone treatment (Fig. 7C–F, respectively). Similarly, in testosterone-treated PTM-ARKO testes, the small involuted LCs normally expressed these LC markers (Fig. 7C–F, arrows), whereas the larger lipid-filled 'abnormal' LCs were less positive or even negative for 3β-HSD, CYP17A1, INSL3 and CYP11A1 (Fig. 7C–F, arrowheads).

Discussion

We demonstrated previously that serum testosterone concentrations are normal in adult PTM-ARKO males, but that supranormal LH concentrations are needed to achieve this (Welsh *et al.*, 2009), indicative of compensated LC failure. The aim of the current study was to investigate the basis of this impairment. The present studies have demonstrated that development of adult LCs is impaired in PTM-ARKO mice, probably as a consequence of reduced IGF, Dhh, PDGF and INSL3 signalling. Unexpectedly, not all LCs were similarly affected in PTM-

Figure 7 Response of peritubular myoid cells (PTM)-androgen receptor (AR)KO and control adult testes to exogenous testosterone. (A) Testis weight in adult mice treated with 'empty' or testosterone-filled implants for 13 weeks. PTM-ARKO testes were significantly smaller than control testes in both sham-implanted 'empty' and testosterone-treated mice. Testosterone treatment had no effect on testis weight in either KO or control mice, compared with age-matched 'empty' sham-implanted mice. (B) Haematoxylin and eosin staining of sham-operated and testosterone-treated PTM-ARKO and control testes. (C) Expression of 3β-hydroxysteroid dehydrogenase, cytochrome p450 17, insulin-like factor 3 and cytochrome p450 11a in sham-operated and testosterone-treated PTM-ARKO and control testes highlighting the presence of two populations of Leydig cells (LCs) in PTM-ARKO testes, which show differential expression of these LC markers. Arrow, normal LC; arrowhead, 'abnormal' LC. Values are mean ± SEM; $n = 4–5$ mice. $***p < 0.01$ compared with treatment controls. Scale bars = 50 μm.


A
Testis weight (g)
0.10
0.08
0.06
0.04
0.02
0.00
***
***
Control
KO
Control
KO
'Empty' implant treatment controls
T-treated mice
B
Control - Sham
KO - Sham
Control + T
KO + T
H&E
C
3B-HSD
Cyp17
Insl3
Cyp11a

ARKO adult testes, but instead there appeared to be two sub-populations of LCs: one grossly 'normal' sub-population which normally expressed adult LC makers such as INSL3, LHR and steroidogenesis enzymes, and another grossly 'abnormal' sub-population which appeared not to have developed fully into mature adult LCs and which only weakly expressed INSL3, LHR, and some of the steroidogenesis enzymes, and which had abnormal accumulation of cytoplasmic lipid droplets. It is likely that the more 'normal' LC population (approximately 50% that of control testes) has to work harder to compensate for the 'abnormal' LC population to maintain normal serum testosterone concentrations; this is likely to explain the compensatory LC failure in PTM-ARKO males.

Our previous studies showed that overall LC function is impaired in PTM-ARKO mice, based on the elevated LH, but normal serum testosterone concentrations (O'Shaughnessy *et al.*, 2002). The studies presented here showed a likely basis for this as expression of genes involved in steroidogenesis was significantly reduced in adult PTM-ARKO testes. These studies are based on QRT-PCR analysis of global gene expression in the testis, which takes no account of the differential gene expression between the different LC sub-populations, but this still gives a clear indication of an overall decrease in steroidogenic gene expression in adult PTM-ARKO testes. For example, expressions of *HMGCS1* and *HMGCR1*, which are involved in cholesterol synthesis de novo, were both significantly reduced. This would result in less starting product available for steroidogenesis and so could limit testosterone production per LC. Furthermore, in adult PTM-ARKO testes, expressions of *3β-HSD1* and *3β-HSD6* and *17β-HSD3*, the enzymes important for the conversion of DHEA into testosterone, were also reduced. This could further limit testosterone production; reduced expression of these enzymes is consistent with failure of adult LC differentiation (O'Shaughnessy *et al.*, 2002). Interestingly, relative expression of *3β-HSD1* and *3β-HSD6* and *17β-HSD3* was not significantly reduced in PTM-ARKO testes at d12, the age at which adult LC development first begins. In fact, relative expression of 3β-HSD1, which is expressed in both foetal and adult LCs, and 17β-HSD3, which is mainly expressed in adult LCs, were both increased in PTM-ARKO testes at d12 compared with controls. Increased expression of these genes is among the earliest changes associated with adult LC development suggesting that adult LC development begins normally in PTM-ARKO testes, but that problems arise at later stages. Furthermore, LC size and number were normal at d12 and LC gene expression was largely normal at d1 and d12. Together, these results suggest that there is a reduction in adult LC function at both d12, when the adult LCs are just developing, as well as in adulthood (d100). Our results do not indicate any gross dysfunction of foetal LC as determined by steroidoegenic enzyme expression at d12, and by the normal masculinisation of PTM-ARKO males, although there was a relative reduction in expression of foetal LC markers in adult PTM-ARKO testes compared with controls.

Steroidogenic enzyme expression, and thus adult LC differentiation, is critically dependent on SF-1 expression (reviewed in Hoivik *et al.*, 2010). *SF-1* expression was reduced in PTM-ARKO testes from d12 onwards, which suggests that PTM cell androgen signalling may affect normal LC development and function by regulating (directly or indirectly) SF-1 expression. This requires further investigations as there are several mechanisms controlling SF-1 expression (reviewed in Hoivik *et al.*, 2010) which PTM cell signalling may affect. Interestingly, immunohistochemical analysis revealed that SF-1 protein was expressed as expected in some adult PTM-ARKO LCs, whereas other LCs were less positive or even negative for SF-1. These 'abnormal' cells were confirmed as steroidogenic cells as they expressed some LC markers and steroidogenic enzymes, albeit weakly, but had an abnormal LC appearance at both light microscopic and electron microscopic levels. This provided the first evidence to us of two sub-populations of LCs in adult PTM-ARKO testes, one of which did not normally express several adult LC markers, indicating impaired function. These two LC sub-populations could be identified from d35 onwards. Adult LCs start developing just before puberty with immature LCs not identified until d35 (Wu *et al.*, 2010; Ge *et al.*, 1996). 3β-HSD is the first enzyme to be expressed in progenitor adult LCs, with CYP11A1 and CYP17A1 switching on slightly later (Mendis-Handagama & Ariyaratne, 2001; Zhang *et al.*, 2004). Indeed, in LHRKO mice, in which LC development is impaired, LCs express 3β-HSD, but not CYP11A1 or CYP17A1 (Zhang *et al.*, 2004). This is similar to PTM-ARKOs in which some 3β-HSD-positive LCs do not express CYP11A1 or CYP17A1. This suggests that the 'abnormal' PTM-ARKO LCs could be progenitor or immature LCs, which were arrested in their differentiation and thus failed to become mature adult LCs. This is the first evidence that AR signalling via the PTM cells is important for this aspect of normal adult LC development.

Further investigation of PTM-ARKO testes revealed that both the 'normal' and 'abnormal' LCs exhibited an altered ultrastructure (rare smooth WER and arranged in CBs, larger irregular mitochondria) compared with control adult mouse LCs. This suggests that all LCs in PTM-ARKO testes have impaired development and/or function, but that some LCs are more affected than others. Thus, there were clear ultrastructural differences between

the grossly 'normal' and 'abnormal' LCs with 'abnormal' LCs showing an obvious increase in lipid droplets and reduced LHR expression, a pattern reminiscent of progenitor or immature LCs rather than mature adult LCs, which normally express LHR and have few lipid droplets (Shan & Hardy, 1992; Shan *et al.*, 1993; Ge *et al.*, 1996). The abundant lipid droplets in 'abnormal' PTM-ARKO LCs could reflect a mechanism to overcome the reduced expression of *HMGCS1* and *HMGCR1*, which are required for de novo synthesis of cholesterol. This is similar to immature LCs, which rely on imported lipids for cholesterol synthesis rather than de novo synthesis, as in mature adult LCs (Shan *et al.*, 1993). Conversely, accumulation of lipid droplets could simply be a consequence of reduced steroidogenic output. The latter stages of adult LC development are dependent on LH signalling and LHR expression (Zhang *et al.*, 2001). Serum LH is elevated in adult PTM-ARKO mice, but a lack of LHR expression on some LCs presumably means that these cells are unable to respond normally to LH, which could explain their arrested differentiation; they are therefore unlikely to produce much testosterone. This was apparent in the reduced immunoexpression of 3β-HSD, CYP17A1 and CYP11A1 in the 'abnormal' PTM-ARKO LCs, whereas the adjacent 'normal' LHR-positive LCs normally immunoexpressed these proteins. LCs also enlarge as they differentiate from progenitor into immature LCs (Shan & Hardy, 1992; Shan *et al.*, 1993), but no difference was observed in the size of the abnormal LCs compared with either the normal PTM-ARKO LCs or LCs in controls. Furthermore, the adult size of these abnormal LCs and their expression of AR suggest that these cells are not aberrant foetal LCs. Taken together, our findings suggest that PTM-ARKO LCs undergo initial differentiation from progenitor to immature LCs, but do not complete their development into fully mature adult LCs.

In normal adult testes, only a few foetal LCs and immature adult LCs persist and a majority of LCs are mature adult LCs. In contrast, in PTM-ARKO testes, there were a similar number of LCs present as in controls, but the immature LCs do not fully differentiate into mature adult LCs in PTM-ARKO testes. Instead, there appears to be a continuum of LC development in PTM-ARKO testes, with some LCs (normal) maturing more than others (abnormal). This scenario is strikingly different from the simple reduction in adult LC number seen in both SCARKO and ARKO adult mice (De Gendt *et al.*, 2005). This suggests that androgen action via the SCs, and probably the LCs, is important for determining LC number, whereas androgen signalling via PTM cells is important for the normal differentiation of adult LCs. The appearance of abnormal LCs in PTM-ARKO testes at d35, but not before, is consistent with the LC abnormalities in PTM-ARKO testes resulting from impaired adult LC differentiation. However, it remains unclear why a proportion of LCs arrest in development more noticeably than others in PTM-ARKOs; this requires further investigation. We previously reported that AR is not ablated from all PTM cells in PTM-ARKO mice (Welsh *et al.*, 2009); however, we could not find any correlation between the location of the 'abnormal' immature LCs in the adult PTM-ARKO testes and the presence or absence of adjacent PTM AR expression. Furthermore, AR is also ablated from the blood vessel smooth muscle cells in the PTM-ARKO mouse (Welsh *et al.*, 2009), which might affect LC development. However, we did not identify a similar LC phenotype in SMARKO mice in which AR is only ablated from blood vessel smooth muscle cells (Welsh *et al.*, 2010). This suggests that the PTM-ARKO LC phenotype reported here is the consequence of AR ablation from PTM cells, either alone or in conjunction with AR ablation from the blood vessel smooth muscle cells.

We investigated the possible mechanisms underlying impaired LC development in PTM-ARKOs. First, we investigated INSL3, which switches on as adult LCs develop and is a marker of fully differentiated adult LCs (Pusch *et al.*, 1996; Ivell & Bathgate, 2002; Mendis-Handagama *et al.*, 2007). It has been suggested that INSL3 concentrations might serve as a marker for LC dysfunction (Foresta *et al.*, 2004). Consistent with this and the problems with LC development identified in PTM-ARKO mice, expression of *Insl3* mRNA was consistently decreased in PTM-ARKO mice at d12–100. This is similar to the reduction in *Insl3* expression observed in ARKO, but not in SCARKO, mice (De Gendt *et al.*, 2005) and could contribute to the impaired LC development in PTM-ARKO testes. Furthermore, immunohistochemistry revealed that INSL3 continued to be expressed in the 'normal' LCs in PTM-ARKOs, but was dramatically reduced or even absent from the 'abnormal' LCs; it remains unclear why this differential effect occurs and raises the question whether these cells do not develop normally because they do not express INSL3 or if they are programmed not to develop normally so do not switch on INSL3.

IGF-1 has also been reported to play a role in LC differentiation and induction of adult LC function, and IGF-1 knockout mice display abnormal functional and morphological differentiation of adult LCs (Wang *et al.*, 2003). *IGF-1* expression was reduced in PTM-ARKO testes and, at the protein level, this reduction was restricted to the 'abnormal' LCs, whereas the 'normal' PTM-ARKO LCs expressed IGF-1 similar to LCs in controls. We investigated the possible mechanisms underlying the reduction in IGF-1 expression and discovered that expression of

IGFBP3 mRNA, an inhibitor of IGF-1, was increased in PTM-AKRO mice. Kallikreins 21, 24 and 27 are expressed in LCs, are androgen-dependent and hydrolyse and degrade IGFBP3 (Schill & Miska, 1992; Matsui *et al.*, 2000, 2005; Matsui & Takahashi, 2001). Expression of these *kallikreins* was reduced in PTM-ARKO testes, which could result in increased IGFBP3 in LCs and thus a decrease in IGF1, which in turn leads to arrested LC differentiation and impaired function. Whilst disturbances to this signalling pathway offer a potential explanation for impaired LC differentiation in PTM-ARKO testes, further investigations are required to establish how PTM AR signalling affects LC IGF-1 and why this differentially affects only a proportion of LCs. Dhh and PDGF-A are also involved in adult LC development with adult LCs failing to develop in the absence of expression of ether gene (Clark *et al.*, 2000; Gnessi *et al.*, 2000). PDGF-A is expressed pre-pubertally in SCs and in adulthood in LCs (Mariani *et al.*, 2002) and was reduced in PTM-ARKO testes, as was *Dhh* expression. Dhh is made by SCs, which stimulate differentiation of LCs by upregulating SF-1 and cyp11a1 and its receptor, Ptch, is expressed on LCs (Clark *et al.*, 2000; Yao *et al.*, 2002). As expression of both Dhh and its effector, *Gli1* (Kroft *et al.*, 2001), was reduced in PTM-ARKO testes; this might impair LC differentiation and partly explain the reduced cyp11a1 and SF-1 expression in the 'abnormal' LCs. Disturbance of the Dhh signalling pathway in PTM-ARKO testes provides further evidence that AR signalling via the PTM cells can affect SC signalling which in turn affects other testicular cells; this further highlights the importance of paracrine regulation in the testis.

In normal mice, exposure to exogenous testosterone decreases LH secretion, via negative feedback, thereby causing LC involution and reduced testicular testosterone synthesis (Keeney *et al.*, 1988, 1990). We exposed PTM-ARKO mice to exogenous testosterone for 13 weeks to investigate whether the two sub-populations of LCs involuted as in control testes. Testosterone treatment inhibited LH production as expected, but maintained normal serum testosterone concentrations in controls and KOs. In control mice, this caused LC involution, as was also the case for 'normal' LCs in PTM-ARKO testes. In contrast, the 'abnormal' LCs in PTM-ARKOs did not involute. This suggests that the 'abnormal' LCs are not responding to altered LH stimulation, consistent with their low/absent expression of LHR. This confirms that these 'abnormal' LCs are not functioning normally. It may also provide an explanation for the compensated LC failure observed in PTM-ARKO mice as they have the same number of LCs as in control testes, yet only half of them appear to be functionally normal, based on immunoexpression data. These 'normal' LCs will therefore have to work harder to maintain normal serum testosterone concentrations in PTM-ARKO adults. Furthermore, this finding suggests that as these cells persisted even after dramatically reducing serum LH concentrations, their occurrence is not simply a consequence of the elevated LH in these mice. The idea of LC heterogeneity and differences in steroidogenic enzyme activity between the two populations has been proposed previously in both humans (Qureshi & Sharpe, 1993) and rats (Payne *et al.*, 1980). Furthermore, it was suggested that in rats population II, the more steroidogenically active LCs develop from population I, but the factors involved remained unknown (Payne *et al.*, 1980). The studies presented here offer new insight into this phenomenon and suggest that PTM cells are involved in the differential development of these two LC populations.

In conclusion, PTM AR signalling is important in regulating normal development, structure and function of adult LCs. LC development is impaired in PTM-ARKO testes, associated with altered SF-1, INSL3, IGF, PDGF and Dhh signalling, resulting in mixed populations of LC that may be arrested in different stages of development. As a consequence, the more 'normal' LC sub-population has to work harder to compensate for the 'abnormal' LC sub-population. It is not clear why these two populations of neighbouring LCs develop differentially in PTM-ARKO testes, but these findings uncover new androgen-dependent paracrine mechanisms underlying adult LC development. The PTM-ARKO provides a model in which to investigate these mechanisms.

Acknowledgements

The authors are grateful to Karel De Gendt and Guido Verhoeven for providing the AR^{flox} mice and Michael Kotlikoff for the smMHC-Cre mice. They thank the UK Medical Research Council for funding (WBS U.1276.00.002.0003.01) and David Brownstein, Mark Fisken, Nancy Nelson, Chris McKinnell and the members of the imaging facility and assay laboratory for technical assistance.

References

Baker PJ & O'Shaughnessy PJ. (2001a) Expression of prostaglandin D synthetase during development in the mouse testis. *Reproduction* 122, 553–559.

Baker PJ & O'Shaughnessy PJ. (2001b) Role of gonadotrophins in regulating numbers of Leydig and Sertoli cells during fetal and postnatal development in mice. *Reproduction* 122, 227–234.

Baker PJ, Johnston H, Abel M, Charlton HM & O'Shaughnessy PJ. (2003) Differentiation of adult-type Leydig cells occurs in gonadotrophin-deficient mice. *Reprod Biol Endocrinol* 1, 4. doi:10.1186/1477-7827-1-4.

Christensen AK & Fawcett DW. (1966) The fine structure of testicular interstitial cells in mice. *Am J Anat* 118, 551–571.

Clark AM, Garland KK *et al.* (2000) Desert hedgehog (Dhh) gene is required in the mouse testis for formation of adult-type Leydig cells and normal development of peritubular cells and seminiferous tubules. *Biol Reprod* 63, 1825–1838.

Corker CS & Davidson DW. (1978) A radioimmunoassay for testosterone in various biological fluids without chromatography. *J Steroid Biochem* 9, 373–374.

De Gendt K, Swinnen JV *et al.* (2004) A Sertoli cell-selective knockout of the androgen receptor causes spermatogenic arrest in meiosis. *Proc Natl Acad Sci USA* 101, 1327–1332.

De Gendt K, Atanassova N *et al.* (2005) Development and function of the adult generation of Leydig cells in mice with Sertoli cell-selective or total ablation of the androgen receptor. *Endocrinology* 146, 4117–4126.

Ferlin A, Pepe A *et al.* (2009) New roles for INSL3 in adults. *Ann N Y Acad Sci* 1160, 215–218.

Fisher JS, Macpherson S *et al.* (2003) Human 'testicular dysgenesis syndrome': a possible model using in-utero exposure of the rat to dibutyl phthalate. *Hum Reprod* 18, 1383–1394.

Foresta C, Bettella A *et al.* (2004) A novel circulating hormone of testis origin in humans. *J Clin Endocrinol Metab* 89, 5952–5958.

Ge RS, Shan LX *et al.* (1996) Pubertal development of Leydig cells. In: The Leydig Cell (eds AH Payne, M Hardy & LD Russell), pp. 160–172. Cache River Press, Vienna, IL.

Gnessi L, Basciani S *et al.* (2000) Leydig cell loss and spermatogenic arrest in platelet-derived growth factor (PDGF)-A-deficient mice. *J Cell Biol* 149, 1019–1026.

Habert R, Lejeune H *et al.* (2001) Origin, differentiation and regulation of fetal and adult Leydig cells. *Mol Cell Endocrinol* 179, 47–74.

Haider SG. (2004) Cell biology of Leydig cells in the testis. *Int Rev Cytol* 233, 181–241.

Handelsman DJ. (2008) Androgens. In: Endocrinology of Male Reproduction (ed. R. McLachlan), Available at: http://www.endotext.org/male/index.htm.

Hardy MP, Zirkin BR *et al.* (1989) Kinetic studies on the development of the adult population of Leydig cells in testes of the pubertal rat. *Endocrinology* 124, 762–770.

Hiroi H, Christenson LK *et al.* (2004) Regulation of transcription of the steroidogenic acute regulatory protein (StAR) gene: temporal and spatial changes in transcription factor binding and histone modification. *Mol Cell Endocrinol* 215, 119–126.

Hoivik EA, Lewis AE *et al.* (2010) Molecular aspects of steroidogenic factor 1 (SF-1). *Mol Cell Endocrinol* 315, 27–39.

Huhtaniemi I & Toppari J. (1995) Endocrine, paracrine and autocrine regulation of testicular steroidogenesis. *Adv Exp Med Biol* 377, 33–54.

Ikeda Y, Shen WH *et al.* (1994) Developmental expression of mouse steroidogenic factor-1, an essential regulator of the steroid hydroxylases. *Mol Endocrinol* 8, 654–662.

Ivell R & Bathgate RA. (2002) Reproductive biology of the relaxin-like factor (RLF/INSL3). *Biol Reprod* 67, 699–705.

Keeney DS, Mendis-Handagama SM *et al.* (1988) Effect of long term deprivation of luteinizing hormone on Leydig cell volume, Leydig cell number, and steroidogenic capacity of the rat testis. *Endocrinology* 123, 2906–2915.

Keeney DS, Sprando RL *et al.* (1990) Reversal of long-term LH deprivation on testosterone secretion and Leydig cell volume, number and proliferation in adult rats. *J Endocrinol* 127, 47–58.

Khan S, Teerds K *et al.* (1992) Growth factor requirements for DNA synthesis by Leydig cells from the immature rat. *Biol Reprod* 46, 335–341.

Kroft TL, Patterson J *et al.* (2001) GLI1 localization in the germinal epithelial cells alternates between cytoplasm and nucleus: upregulation in transgenic mice blocks spermatogenesis in pachytene. *Biol Reprod* 65, 1663–1671.

Mariani S, Basciani S *et al.* (2002) PDGF and the testis. *Trends Endocrinol Metab* 13, 11–17.

Matsui H & Takahashi T. (2001) Mouse testicular Leydig cells express Klk21, a tissue kallikrein that cleaves fibronectin and IGF-binding protein-3. *Endocrinology* 142, 4918–4929.

Matsui H, Moriyama A *et al.* (2000) Cloning and characterization of mouse klk27, a novel tissue kallikrein expressed in testicular Leydig cells and exhibiting chymotrypsin-like specificity. *Eur J Biochem* 267, 6858–6865.

Matsui H, Takano N *et al.* (2005) Characterization of mouse glandular kallikrein 24 expressed in testicular Leydig cells. *Int J Biochem Cell Biol* 37, 2333–2343.

McLachlan RI, O'Donnell L, Meacham SJ, Stanton PG, de Krester DM, Pratis K & Robertson DM. (2002) Identification of specific sites of hormonal regulation in spermatogenesis in rats, monkeys, and man. *Recent Prog Horm Res* 57, 149–179.

McNeilly JR, Saunders PT *et al.* (2000) Loss of oocytes in Dazl knockout mice results in maintained ovarian steroidogenic function but altered gonadotropin secretion in adult animals. *Endocrinology* 141, 4284–4294.

Mendis-Handagama SM & Ariyaratne HB. (2001) Differentiation of the adult Leydig cell population in the postnatal testis. *Biol Reprod* 65, 660–671.

Mendis-Handagama SM, Ariyaratne HB *et al.* (2007) Expression of insulin-like peptide 3 in the postnatal rat Leydig cell lineage: timing and effects of triiodothyronine-treatment. *Reproduction* 133, 479–485.

Miller WL. (1998) Early steps in androgen biosynthesis: from cholesterol to DHEA. *Baillieres Clin Endocrinol Metab* 12, 67–81.

Murphy L, Jeffcoate IA *et al.* (1994) Abnormal Leydig cell development at puberty in the androgen-resistant Tfm mouse. *Endocrinology* 135, 1372–1377.

Nef S, Shipman T *et al.* (2000) A molecular basis for estrogen-induced cryptorchidism. *Dev Biol* 224, 354–361.

O'Shaughnessy PJ, Johnston H *et al.* (2002) Failure of normal adult Leydig cell development in androgen-receptor-deficient mice. *J Cell Sci* 115(Pt 17), 3491–3496.

Payne AH, Downing JR *et al.* (1980) Luteinizing hormone receptors and testosterone synthesis in two distinct populations of Leydig cells. *Endocrinology* 106, 1424–1429.

Pusch W, Balvers M *et al.* (1996) Molecular cloning and expression of the relaxin-like factor from the mouse testis. *Endocrinology* 137, 3009–3013.

Quigley CA, De Bellis A *et al.* (1995) Androgen receptor defects: historical, clinical, and molecular perspectives. *Endocr Rev* 16, 271–321.

Qureshi SJ & Sharpe RM. (1993) Evaluation of possible determinants and consequences of Leydig cell heterogeneity in man. *Int J Androl* 16, 293–305.

Schill WB & Miska W. (1992) Possible effects of the kallikrein-kinin system on male reproductive functions. *Andrologia* 24, 69–75.

Scott HM, Mason JI *et al.* (2009) Steroidogenesis in the fetal testis and its susceptibility to disruption by exogenous compounds. *Endocr Rev* 30, 883–925.

Shan LX & Hardy MP. (1992) Developmental changes in levels of luteinizing hormone receptor and androgen receptor in rat Leydig cells. *Endocrinology* 131, 1107–1114.

Shan LX, Phillips DM *et al.* (1993) Differential regulation of steroidogenic enzymes during differentiation optimizes testosterone production by adult rat Leydig cells. *Endocrinology* 133, 2277–2283.

Sharpe RM. (1994) Regulation of spermatogenesis. In: The Physiology of Reproduction (eds E Knobil & JD Neill), pp. 1363–2434. Raven, New York.

Tan KA, De Gendt K *et al.* (2005) The role of androgens in Sertoli cell proliferation and functional maturation: studies in mice with total or Sertoli cell-selective ablation of the androgen receptor. *Endocrinology* 146, 2674–2683.

Vergouwen RP, Jacobs SG *et al.* (1991) Proliferative activity of gonocytes, Sertoli cells and interstitial cells during testicular development in mice. *J Reprod Fertil* 93, 233–243.

Wang GM, O'Shaughnessy PJ *et al.* (2003) Effects of insulin-like growth factor I on steroidogenic enzyme expression levels in mouse leydig cells. *Endocrinology* 144, 5058–5064.

Welsh M, Sharpe RM *et al.* (2010) Androgen action via testicular arteriole smooth muscle cells is important for Leydig cell function, vasomotion and testicular fluid dynamics. *PLoS ONE* 5, e13632.

Welsh M, Saunders PT *et al.* (2006) Androgen-dependent mechanisms of Wolffian duct development and their perturbation by flutamide. *Endocrinology* 147, 4820–4830.

Welsh M, Saunders PT *et al.* (2009) Androgen action via testicular peritubular myoid cells is essential for male fertility. *FASEB J* 23, 4218–4230.

Wu X, Arumugam R *et al.* (2010) Androgen profiles during pubertal Leydig cell development in mice. *Reproduction* 140, 113–121.

Xin HB, Deng KY *et al.* (2002) Smooth muscle expression of Cre recombinase and eGFP in transgenic mice. *Physiol Genomics* 10, 211–215.

Yao HH, Whoriskey W *et al.* (2002) Desert Hedgehog/Patched 1 signaling specifies fetal Leydig cell fate in testis organogenesis. *Genes Dev* 16, 1433–1440.

Zhang FP, Poutanen M *et al.* (2001) Normal prenatal but arrested postnatal sexual development of luteinizing hormone receptor knockout (LuRKO) mice. *Mol Endocrinol* 15, 172–183.

Zhang FP, Pakarainen T *et al.* (2004) Molecular characterization of postnatal development of testicular steroidogenesis in luteinizing hormone receptor knockout mice. *Endocrinology* 145, 1453–1463.

Chapter 8

Testis recovery from irradiation

8.1 Prelude

The aim of the project presented next in the manuscript *"Recovery of spermatogenesis after irradiation"* was to elucidate the transcriptional and cellular changes in the testis following low dose irradiation. Mice testes were irradiated with 1 Gy that in our previous study showed to eradicate the spermatocytes with the highest turnover (A1-B). The effects of the irradiation on the somatic cells were not investigated in our previous study, but spermatogenesis was recovered after 42 days (219).

In this study, we wanted to elucidate the effects of low dose irradiation on the somatic cells in the testis. We also wanted to compare the duration of the differentiation stages in adult spermatogenesis with that of the first cycle in postnatal (pn) mice. Adult spermatogenesis, the process after the first cycle, is a continuous process with all germ cell differentiation stages present in the testis at all times. Yet, the gap in germ cells created by the irradiation enabled us to estimate the duration of the individual differentiation steps in adult spermatogenesis for comparison with first pn cycle of spermatogenesis.

The project is in collaboration with Department of Growth and Reproduction, Rigshopitalet, Copenhagen, Denmark. The work presented is a combination of experimental work and data analysis. The basis of this study was mice testis samples collected in a time series after the irradiation. When a mouse was culled, one testis was snap-frozen for RNA purification and the other testis was fixed in paraffin for IHC. We have used both materials in this study and I participated in the gene expression microarray experiments and performed all data analysis.

The testis is comprised of various cell types, such as Leydig, PTM and Sertoli cells as well as germ cells in multiple stages of differentiation. This made the gene expression data from the testes following irradiation rather

complicated as the cellularity of the testes changed during the time series, because the gap in germ cells moved through spermatogenesis. Hence, the mRNA contribution from each cell type to the total mRNA pool varied according to the amount of gene expression from the cells and how much of the testis the cell type occupied, relative expression.

The testis tissue was apart from the gene expression data analysis also investigated by IHC for morphological changes. The results are presented in the manuscript. The manuscript is almost ready for publication, we need the final IHC validation of E2F1 as a marker for Leydig cell hyperproliferation.

8.2 Manuscript

Recovery of spermatogenesis after irradiation

KC Belling[1], M Tanaka[2,3], M Dalgaard[4], JE Nielsen[4], HB Nielsen[1], K Almstrup[4], H Leffers[4,5]

[1]Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, 2800 Lyngby, Denmark.
[2]Institute for Animal Experimentation, St. Marianna University Graduate School of Medicine, 2-16-1 sugao, Miyamae-ku, Kawasaki 216-8511, Japan.
[3]Department of Pharmacology, St. Marianna University School of Medicine, 2-16-1 sugao, Miyamae-ku, Kawasaki 216-8511, Japan.
[4]Department of Growth and Reproduction, Rigshospitalet, 2100 Copenhagen, Denmark.
[5]Corresponding author

Abstract

Spermatogenesis is the testicular process whereby spermatogonial stem cells (SSCs) divide and differentiate into mature sperm. Spermatogenesis takes place in the seminiferous tubules where Sertoli cells nurse the germ cells during development. The interstitium between the tubules contains the Leydig cells that produce testosterone essential for spermatogenesis. At birth the tubules only contain primordial germ cells and Sertoli cells. Spermatogenesis starts in mice at postnatal (pn) day 4-6 and after the first cycle is completed, it is a continuous process with all differentiation stages of germ cells present in the testis at all times.

Radiotherapy is used routinely to treat testis cancer patients. The radiosensitivity of the testis cells varies with the most sensitive cells being the highly dividing germ cells, whereas SSCs and the somatic cells are relative radio-resistant. Irradiation of the adult testis in low doses creates a gap of cells in the germinal epithelium, caused by the death of spermatogonial cells, which can be followed as it moves through later stages of germ cells post-irradiation (pi). The cellular changes in the testis tissue after irradiation are still not fully explained.

The aim of the present study was to elucidate the cellular changes in the testis following low dose irradiation and also to compare the duration and timing of adult spermatogenesis with the first wave in pn mice. For that purpose, gene expression array data was generated for a time series of testis RNA samples from pi day 3 to day 59. We identified a subset of 988 transcripts that changed during recovery and by Partitioning Around Medoids

(PAM) clustering we identified five unique associated with specific testis cell types. By comparison of the five clusters with studies of pn spermatogenesis it seemed that the duration of a spermatogenetic cycle is the same in adult mice as the length of the first wave of spermatogenesis in pn mice. By immunohistochemistry we did no find any changes in the Sertoli cells, but evidence suggests that the irradiation leads to hyperproliferation of Leydig cells at pi day 14-28.

Introduction

The testis is comprised of seminiferous tubules with interstitium in between. The interstitium contains the Leydig cells that in response to Luteinizing hormone (LH) produce testosterone that stimulates spermatogenesis (22). The seminiferous tubules consist of an outer layer of Peritubular myoid cells (PTM) cells that surround the basement membrane (220). Inside the tubules, Sertoli cells surround and nurse the germ cells during spermatogenesis (4).

Spermatogenesis is a process of various steps of divisions, meiosis and differentiations whereby spermatogonial stem cells (SSCs) develop into spermatozoa (221). Spermatogenesis initiates by division of A_{single} (A_s) (2N) to produce A_{paired} (A_{pr}), which can take one of two paths: either A_{pr} completes cytokinesis and becomes two new stem cells or A_{pr} continues to undergo further divisions during differentiation (29). After the first division of A_{pr} it is termed $A_{aligned}$ (A_{al}) that further differentiates to A1 spermatogonia that undergoes six divisions via A2, A3, A4, Intermediate (In) and B spermatogonia to primary spermatocytes (4N) (28). Primary spermatocytes undergo DNA recombination before the first meiotic division that produces secondary spermatocytes (2N), which further undergoes second meiotic division that produces haploid round spermatids (N) (30). Once the spermatids have differentiated to mature elongated spermatids they are released into the tubular lumen where after they are called spermatozoa (221). The maturation of spermatozoa continues while the cells are passively transported in the fluid to the rete testis and the epididymis through which spermatozoa leave the testis (29; 30).

The timing of the particular differentiation steps is relatively easy to study in rodents during the first wave of spermatogenesis, as the cell population in the seminiferous tubules at birth is simple with only early spermatogonia and Sertoli cells surrounded by a thin layer of PTM cells. Spermatogenesis in mice starts at postnatal (pn) day 4-6 when early spermatogonia undergo the first mitosis and start to differentiate. Primary spermatocytes start to be present in the testis at pn day 10, and secondary spermatocytes and early round spermatids at pn day 20 (222; 223). Spermatozoa are not found in the seminiferous tubules before pn day 35 (224).

After the first wave, spermatogenesis is a continuous process with all differentiation stages of germ cells present in the testis at all times (225), which makes the study of the adult spermatogenesis much more complicated. In rodents, irradiation of the adult testis in low doses creates a gap of cells

in the germinal epithelium (219), which disturbs the cell composition. The cellular and molecular changes caused by the movement of the gap along the different differentiation steps of germ cells, as the testis is repopulated from stem cells, is a way to study adult spermatogenesis.

Cells in the testis vary in radio-sensitivity and the degree of damage depends on the treatment, dose and fractionation. Sertoli cells in rodents seem affected by irradiation doses from 5 Gray (Gy) (226) and Leydig cells are damaged by the irradiation doses used in the clinic to treat testis cancer (16-20 Gy) (76; 227; 228), but testosterone production may be maintained. SSCs and As are the most radio-resistant germ cells, because they have a slow turnover (229). The cells that are most susceptible to irradiation are those that differentiate and proliferate the most, which are A1-A4 spermatogonia followed by A_{pr} and A_{al} (230; 231), but also In and B spermatogonia are susceptible to irradiation damage (232; 233). The further evolved germ cells, spermatocytes and spermatids, are not affected by irradiation in low doses and continue their maturation to spermatozoa (227).

The testis is a complex tissue that consists of many different cell types. Gene expression analysis of such complex tissue is difficult to interpret. The contribution from each cell type to the total RNA pool is determined by the concentration of the mRNA in specific cell types and the percentage of the total volume the cell type occupies, the cellularity. Thus, disappearance and reappearance of cells add to the complexity of the data like the huge differences in cell size (234).

The aim of the present study was to study the cellular effects of the irradiation effects and compare the duration and timing of adult spermatogenesis by the molecular changes with the first pn wave of spermatogenesis in mice. For this purpose, gene expression array data was generated for a time series of testis RNA samples from post-irradiation (pi) day 3 to 59. An enrichment score identified 988 transcripts that changed in expression in the time series and by an unbiased clustering analysis we identified five unique clusters, each with transcripts with similar expression patterns. We assigned the five clusters to specific testis cell types: spermatogonia, spermatocytes, early spermatids, late spermatids and somatic cells and gained further knowledge of the transcripts in each of the clusters by gene set enrichment analysis. The patterns of the five clusters were in addition compared with similar studies on pn induction of spermatogenesis. We further validated Leydig and Sertoli cell markers by immunohistochemistry (IHC) to see whether irradiation cause cellular changes in the somatic cells in the testis.

Materials and Methods

Mice testis preparation

Treatment of mice and preparation of testicular RNA and sections were performed as in Shah et al., 2009 (219). In brief, male C3H/He strain mice were obtained from Japan SLC (Shizuoka, Japan) and maintained under controlled conditions (22 ± 2°C, 55 ± 5% humidity, 12h light/dark cycle,

lights on 0600h) with laboratory chow (CE-2, Japan Crea, Tokyo, Japan) and water ad libitum.

Eleven-week old mice were anesthetized with pentobarbital and covered with lead sheeting except for the scrotum. The testes were locally exposed to X-ray irradiation with 1 Gy. Testes from 1 or 4 mice were sampled and weighted regularly during recovery on days 0, 3, 7, 10, 14, 17, 21, 24, 28, 31, 35, 38, 42, 45, 48, 52, 56, and 59. One testis was fixed in 4% paraformaldehyde in 0.1M phosphate buffer, pH 7.4, overnight at 4°C and subsequently dehydrated in graded series of ethanol and embedded in paraffin for IHC. The contralateral testis was snap-frozen in liquid nitrogen and used for preparation of total RNA.

The Japanese Pharmacological Society approved the animal study and the animals were treated according to generally accepted guidelines for animal experimentation at St. Marianna University Graduate School of Medicine and guiding principles for the care and use of laboratory animals.

Gene expression microarrays

Total RNA purification from whole testes was performed on samples from pi day 3 to day 59 in the time series with NucleoSpin RNA II (Macherey-Nagel, Dueren, Germany) according to the manufacturer s protocol. RNA quality was determined using Bioanalyzer nano kit (Agilent Technologies, Santa Clara, California, US). The samples were amplified (one round) using the MessageAmp II aRNA Amplification Kit (Applied Biosystems, Carlsbad, California, US) and the aRNA applied to Agilent whole mouse genome oligo microarrays 4×44K. Hybridization and scanning of one-color arrays were done as described by the manufacturer (Agilent Technologies, Santa Clara, California, US). The gProcessedSignal was loaded into the limma R/Bioconductor package, normalized between arrays using quantile normalization procedure, and probes were collapsed taken the median.

Enrichment score

Transcripts with expression patterns potentially explained by recovery from irradiation were selected for further analysis based on an enrichment score. The score was calculated for each transcript as the sum of the absolute differences in gene expression values between neighbouring time points in the time series divided by the sum of the differences from the mean of each gene expression value in the time series. The false discovery rate (FDR) of the enrichment score was calculated based on ten calculations on shuffled time series points. Transcripts included in the further analysis were chosen based on their cumulative minimum of the FDR for their enrichment score and the standard deviation of the most extreme gene expression value in each time series. Transcripts with extreme outliers were discarded for further analysis based on the standard deviation of the most extreme point in the time series.

Cluster analysis

Cluster analysis was performed on the subset of transcripts chosen based on their enrichment score described above. A distance matrix was calculated for each transcript as the correlation variances of the gene expressions during the time series centered around 1. The transcripts were clustered according to the distance matrix using Partitioning Around Medoids (PAM) clustering. Clustering was performed multiple times with different number of clusters. For each analysis, the gene expression patterns of the transcripts in each cluster were plotted to see how similar they were within a cluster and whether some clusters had similar patterns. The number of clusters that separated most clusters with unique patterns was chosen as the best separation of out data.

Assigning clusters to testis-specific cells

To determine which cells in the testis that express the transcripts in each cluster we included knowledge from our previous study of the same testes samples (219; 235). In our previous study, we used differential display and *in situ* hybridization (ISH) of chosen cell markers to determine their cell-specific expression patterns during the recovery from irradiation. The expression patterns of the five clusters identified by gene expression array analysis in this study were compared with the patterns of the four clusters identified in our previous study. Additionally, testis cell-specific transcripts from three previously published studies were used as markers (235; 236; 237). Identification of the markers in each cluster added to the clarification of which cells in the testis express the transcripts in each cluster.

Gene set enrichment analysis

Gene set enrichment analysis was performed of the transcripts in each cluster separately using DAVID (238) with the genes represented on the arrays as reference and set a cut off for the Bonferroni corrected p-value at 0.01.

Immunohistochemistry (IHC)

The following primary antibodies were used: Vimentin/HRP (Dako, Glostrup, Denmark; U7034), Transforming growth factor, beta receptor III (Tgfbr3) 1:75 (Santa Cruz Biotechnology, Santa Cruz, CA, USA; sc-6199), and 3beta-hydroxysteroid-dehydrogenase (Hsd3b) 1:6000 (R1484 a gift from Prof. J. Ian Mason, Edinburgh). Vimentin was used according to the manufacturer s protocol. In short the sections were deparaffinated, rehydrated and blocked for endogene peroxidase with H_2O_2, washed in tap water, placed 5min in TBS (0.5M Tris/HCl, 0.15M NaCl, pH 7.6) at 37°C, incubated with 1:10 Trypsine in TBS 15min at 37°C, exposed to the antibody for 1h at room temperature, development was performed with 3-amino-9-ethylcarbazole (AEC). Thorough washing with TBS was performed after each individual step and finally wash in water before a short staining with Meyers haematoxylin.

The two remaining antibodies, Hsd3b and Tgfbr3, were used as in a protocol based on a Zymed histostain kit (Invitrogen, Carlsbad, CA, USA). In short the sections were deparaffinated, rehydrated and blocked for endogen peroxidase as described above, followed by microwave treatment for 15min in TEG buffer (Tris 6.06g, EGTA 0.95g in 5l, pH 9.0). Cross reactivity of the antibodies was minimized by treatment with 0.5% milk powder diluted in TBS. Sections were exposed to the primary antibodies over night at 5°C and 1h at room temperature, then incubated with biotinylated goat anti-rabbit IgG or with biotinylated donkey anti-goat IgG 1:400 in TBS (The binding site Ltd., Birmingham, UK; AB360) for Tgfbr3 before exposure to a peroxidase-conjugated streptavidin complex. Finally, the sections were developed with AEC and counter stained with Meyers haematoxylin. Thorough washing with TBS was performed after each individual step. Slides without addition of primary antibodies were used as controls.

Results

IHC of somatic cell markers suggests Leydig cell hyperplasia

In order to describe histological changes in the testis after irradiation, we investigated the protein expression at pi day 7, 14, 21, 28, 31, 49 and 59 of two Leydig cell markers, the membrane proteoglycan and Transforming growth factor (TGF)-beta co-receptor Tgfbr3 and Hsd3b involved in steroidogenesis, and the type III intermediate filament, Vimentin, which is a Sertoli cell marker.

Vimentin showed a specific staining in Sertoli cells in the control material (pi day 0; Figure 8.1). The staining remained in Sertoli cells throughout recovery but the intensity and amount of staining seemed to decrease as the specific germ cells disappeared. Hence, already from pi day 7, where only spermatogonia are absent from the testis, the amount of staining was reduced, and it remained reduced until pi day 49 where the intensity and amount of staining was back to control levels. The classical extension of Sertoli cell cytoplasm towards the centre of tubules in addition was mainly observed at day 0 and after pi day 49. Although the staining was reduced as distinct germ cell populations were missing, there were no indications of cell death among the Sertoli cells.

In control material (pi day 0; Figure 8.1), Tgfbr3 was expressed in the cell membrane of the majority of the Leydig cells. Already at pi day 7 the Leydig cell membrane reaction however appeared disturbed, incomplete or became cytoplasmatic. In addition, the shape of the Leydig cells became more round and the close configuration in which they were arranged in the control material seemed to be lost. At pi day 21 some membrane reaction reappeared, but was lost again from day 28-31 where hyperplasia of the Leydig cells was pronounced. During pi day 49-59 some of the Leydig cell membrane reactions was restored, but only half of the Leydig cells had membrane reaction at pi day 59 while the rest were without – or showed a cytoplasmatic reaction.

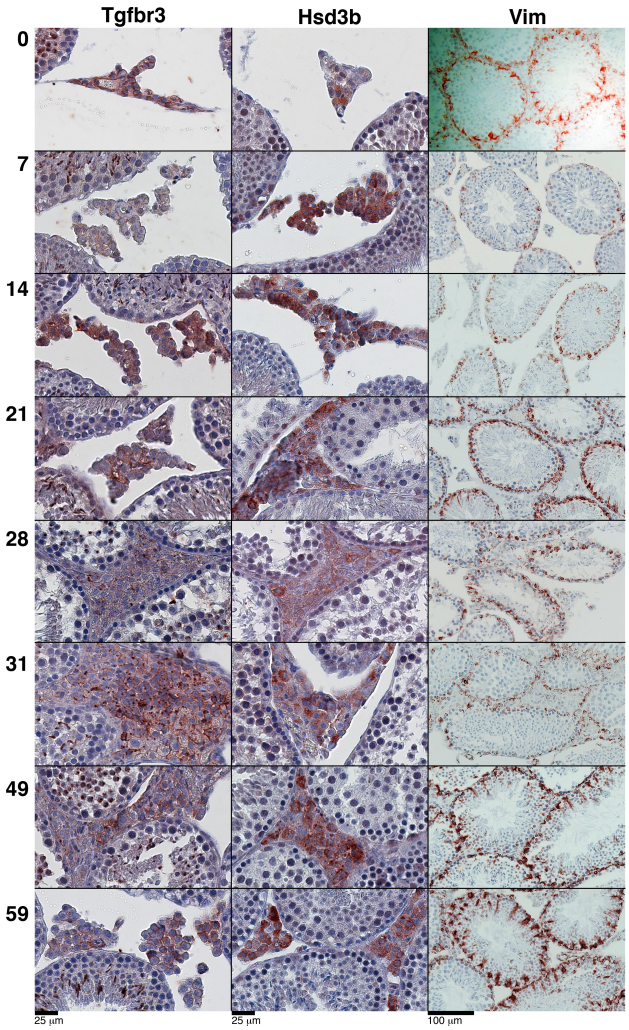


Figure 8.1. IHC using antibodies against the two Leydig cell markers, Tgfbr3 and Hsd3b, and of the Sertoli cell marker, Vimentin (Vim). A hyperproliferation of the Leydig cells are noticed at day 14-28 during recovery from irradiation, whereas the Sertoli cells do not seem affected.

Hsd3b was expressed in the cytoplasma of Leydig cells in the control (pi day 0; Figure 8.1). Only a few Leydig cells reacted strongly, whereas the majority reacted moderately to faint or were negative. Again, already from pi day 7, this picture changed with the vast majority of the Leydig cells now reacting strongly to the antibody and although the intensity of the reaction gradually declined, it was still elevated at pi day 59. Staining with Hsd3b

confirmed the increase in Leydig cells numbers from pi day 21-28, but the number of Leydig cells was almost back to control level at pi day 59.

It is difficult to evaluate the hyperplasia with microarrays since all the genes expressed in somatic cells appear upregulated on pi days 14 to 28 because of the absence of germ cells (see below). Nevertheless, we know that the transcription factor E2F transcription factor 1 (E2F1) is barely detectable in normal human Leydig cells, but strongly upregulated i hyperplastic human Leydig cells (results not shown). Thus, we investigated the expression of E2F1 and found that its expression became elevated from pi day 17 and remained upregulated until pi day 38. This seems to correlate nicely with the period where we observed an increase in Leydig cells.

Clusters of transcripts with similar expression pattern

We made a microarray study of gene expression during the recovery and an enrichment score was calculated for all transcripts represented on the arrays to identify transcripts with expression changes during the time series. A subset of 988 transcripts of interest was chosen for further analysis. The transcripts all had an enrichment score with cumulative minimum of FDR equal or less than 30% and a standard deviation of the most extreme point in the time series equal or less than 15. We performed numerous PAM clustering analysis of the distance matrix of gene expression changes of the subset with different number of clusters. We identified five clusters each with unique gene expression patterns during recovery from irradiation to be the number of clusters that could be identified from this dataset (Figure 8.2).

Figure 8.2. The mean gene expression values for each time point in the time series of the five clusters identified by PAM clustering.

Assigning transcripts in clusters to specific testis cells

To identify the cell types from which the transcripts in each cluster originates, we compared the patterns of the five clusters with the patterns of the four clusters identified in our previous study of the same irradiated testis samples (219). This identified that four of the clusters corresponded to transcripts originating from spermatogonia, spermatocytes, spermatids and somatic cells including Sertoli cells. The fifth identified cluster in this study was not found in the previous study, but from the gene expression pattern of transcripts in this cluster we estimated that it could originate from late elongated spermatids. Hence, the other spermatid cluster is properly mainly represented by transcripts originating from early round spermatids. Most transcripts seem to stabilize around day 40 pi where a complete round of spermatogenesis has taken place (30).

To add further evidence in associating specific cell types to the gene expression clusters, we identified cell-specific markers from three previous studies (235; 236; 237). In total, 39 cell-specific markers were identified in the five clusters representing spermatogonia, spermatocytes, spermatids, Leydig cells, PTM cells and Sertoli cells (Table 8.1 and Figure 8.3). The spermatogonia marker was found in the cluster we associated with spermatogonia. Spermatocyte markers were evenly distributed in all five clusters, however, from comparison of the profile to our earlier study we estimated that the cluster with a drop in expression at pi day 17 contains the transcripts expressed from spermatocytes. There were spermatid markers present in both clusters associated with early and late spermatids, which suggested that the transcripts in both clusters primarily are transcribed from spermatids. The somatic cell cluster contained all markers associated with somatic cells in the testis and hence is not limited to Sertoli cells.

Table 8.1: **Testis cell-specific markers. The 39 markers used to identify from which cells the genes in each cluster are expressed (235; 236; 237).**

Gene symbol	Cell	Cluster no.
Dazl	Spermatogonia	1
1500011H22Rik	Spermatocytes	2
1700029G01Rik	Spermatocytes	3
2410022L05Rik	Spermatocytes	2
Atl3	Spermatocytes	4
Akap12	Spermatocytes	5
Aldoa-ps1	Spermatocytes	4
Cypt3	Spermatocytes	3
D030056L22Rik	Spermatocytes	1
Gsg2	Spermatocytes	3
H3f3b	Spermatocytes	1
Hnrpa2b1	Spermatocytes	1
Lyar	Spermatocytes	2
Nt5c1b	Spermatocytes	4
Pgk2	Spermatocytes	4
Slc2a3	Spermatocytes	2

Spert	Spermatocytes	4
Stard10	Spermatocytes	4
Stmn1	Spermatocytes	1
Vkorc1	Spermatocytes	5
Acrv1	Spermatids	3
Actl7a	Spermatids	4
Akap4	Spermatids	4
Marcksl1	Spermatids	5
Pdpk1	Spermatids	4
Pdzk1	Spermatids	4
Prm1	Spermatids	2
Spag4l	Spermatids	3
Tctex1d1	Spermatids	4
Tnp2	Spermatids	3
Zpbp	Spermatids	3
Cyp17a1	Leydig cells	5
Hsd17b3	Leydig cells	5
Hsd3b1	Leydig cells	5
Acta2	PTM cells	5
Cldn11	Sertoli cells	5
Clu	Sertoli cells	5
Ctsl	Sertoli cells	5
Vim	Sertoli cells	5

Gene set enrichment analysis

Gene set enrichment analysis was performed on the transcripts from each cluster separately to gain knowledge of overrepresented functional categories. The full list of enriched categories in each cluster is presented in Table 8.2, whereof a few categories are highlighted next. The spermatogonia cluster genes were associated with acetylation and methylation and RNA binding and processing. The spermatocyte cluster was enriched with genes involved in cell cycle and spermatogenesis. The early spermatid cluster was associated with spermatogenesis, histone core and acrosomal vesicle. The late spermatid cluster was enriched in spermatogenesis and serine-type peptidase activity. Finally, the genes in the somatic cluster were enriched in steroidogenesis and lipid biosynthesis among other categories.

Discussion

Mice testes were irradiated with 1 Gy and the recovery of spermatogenesis were followed for 59 days with an interval of three to four days. In our earlier study of the same tissue, significant weight changes were identified during the recovery with a decrease the first 28 days to less than half of the pre-irradiation weight, followed by gradual restoration to normal pre-irradiation levels. The decrease in testis weight corresponds to loss of cells due to irradiation damage followed by re-population of the testis as it recovers from SSCs (219). The pachytene spermatocytes are physically the largest

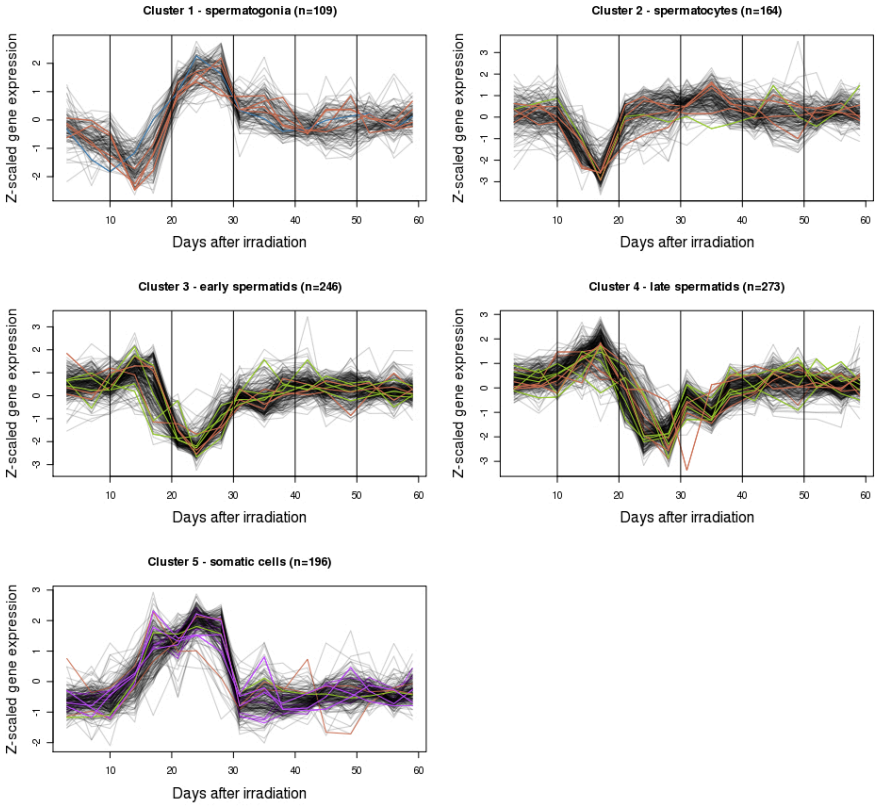


Figure 8.3. Gene expression profiles of the subset of 988 genes that changed in expression during recovery from irradiation divided into five clusters by PAM. The number of transcripts in each cluster is given in the title of each plot as well as which cell each cluster is associated with. The cell specific markers are coloured as following: Blue: Spermatogonia; Red: Spermatocytes; Green: Spermatids; Purple: Somatic cells.

germ cells in the testis (5) and the weight changes observed highly reflects the elimination and repopulation of these cells in the testis.

Table 8.2: **Gene set enrichment analysis.**

Cluster	Cell	Term	Bonferroni
1	Spermatogonia	Acetylation	7.161E-13
1	Spermatogonia	RNA-binding	4.630E-12
1	Spermatogonia	Ribonucleoprotein complex	1.286E-11
1	Spermatogonia	Methylation	1.492E-08
1	Spermatogonia	Ribonucleoprotein	1.531E-07
1	Spermatogonia	Nucleus	3.609E-06

1	Spermatogonia	mRNA processing	1.949E-05
1	Spermatogonia	Spliceosome	7.028E-05
1	Spermatogonia	mrna processing	7.086E-05
1	Spermatogonia	mRNA metabolic process	9.299E-05
1	Spermatogonia	mRNA splicing	9.679E-05
1	Spermatogonia	RNA splicing	1.313E-04
1	Spermatogonia	RNA recognition motif, RNP-1	1.704E-04
1	Spermatogonia	Nucleotide-binding, alpha-beta plait	1.854E-04
1	Spermatogonia	Viral nucleoprotein	1.583E-03
1	Spermatogonia	Isopeptide bond	2.014E-03
1	Spermatogonia	RNA processing	4.265E-03
1	Spermatogonia	Non-membrane-bounded organelle	4.933E-03
1	Spermatogonia	Parkinson's disease	8.467E-03
2	Spermatocytes	Sexual reproduction	1.099E-06
2	Spermatocytes	Gamete generation	7.443E-06
2	Spermatocytes	Spermatogenesis	2.053E-05
2	Spermatocytes	Male gamete generation	2.053E-05
2	Spermatocytes	Multicellular organism reproduction	1.212E-04
2	Spermatocytes	Cell cycle process	3.713E-03
3	Early spermatids	Acrosomal vesicle	1.926E-09
3	Early spermatids	Sexual reproduction	2.649E-09
3	Early spermatids	Spermatogenesis	6.407E-08
3	Early spermatids	Male gamete generation	5.283E-06
3	Early spermatids	Gamete generation	1.474E-05
3	Early spermatids	Secretory granule	1.848E-05
3	Early spermatids	Multicellular organism reproduction	1.987E-04
3	Early spermatids	Nucleosome core	2.520E-04
3	Early spermatids	Nucleosome	6.828E-04
3	Early spermatids	Systemic lupus erythematosus	1.143E-03
3	Early spermatids	Protein-DNA complex	2.077E-03
3	Early spermatids	Histone core	4.902E-03
3	Early spermatids	Single fertilization	5.533E-03
4	Late spermatids	Sexual reproduction	1.801E-05
4	Late spermatids	Spermatogenesis	5.010E-04
4	Late spermatids	Male gamete generation	5.010E-04
4	Late spermatids	Gamete generation	9.769E-04
4	Late spermatids	Multicellular organism reproduction	1.521E-03
4	Late spermatids	Serine-type peptidase activity	7.467E-03
4	Late spermatids	Serine hydrolase activity	7.745E-03
4	Late spermatids	Trypsin	8.970E-03
5	Somatic cells	Acetylation	2.062E-10
5	Somatic cells	Oxidoreductase	2.436E-09
5	Somatic cells	Oxidation reduction	1.977E-06
5	Somatic cells	Mitochondrion	1.069E-04
5	Somatic cells	Response to oxidative stress	4.123E-04
5	Somatic cells	Lipid biosynthetic process	4.182E-04
5	Somatic cells	Steroidogenesis	2.147E-03
5	Somatic cells	Pyruvate metabolism	3.780E-03
5	Somatic cells	Response to hydrogen peroxide	5.697E-03

Irradiation causes cellular changes in somatic cells in the testis

To investigate whether the irradiation affected the histology of the testis and Leydig and Sertoli cells we determined the protein expression of two

Leydig cell markers, Tgfbr3 and Hsd3b, and one Sertoli cell marker, Vimentin. Leydig cells are affected by irradiation although they are more resistant than germ cells. Previous studies report that Leydig cells are preserved after high-dose irradiation up to a total dose of 20 Gy, but that the endocrine function of the cells are affected also at lower doses (75; 239; 240; 241). In this study, the testes were irradiated with 1 Gy in a single dose, which is a low dose compared to clinical doses (76). Yet, from the validation of the two Leydig cell markers, Tgfbr3 and Hsd3b, it seems that the irradiation leads to a hyperproliferation of the Leydig cells around pi days 21-31 (Figure 8.1), which is supported by the increased expression of E2F1. Hyperproliferation of Leydig cells is also observed in many men with fertility problems and testis cancer (242). The proliferation might arise because of a decreased testosterone production, which will induce an increased LH production that may stimulate increased proliferation of Leydig cells. This is also the clinical observation upon irradiation of human testis harbouring carcinoma *in situ* (243) and in rats with Sertoli cell-only testis (244). However, it is not clear if a dose of 1 Gy leads to reduced testosterone production and thus increased LH, since much higher irradiation doses leads to very small changes in testosterone and LH levels (245). However, the presence of Follicle-stimulating hormone (FSH) receptors in Leydig cells could suggest that the hyperproliferation may be a response to decreased inhibin production caused by the absence of germ cells, which will lead to increased level of FSH that may induce hyperproliferation of Leydig cells (246).

The IHC staining of Vimentin of the Sertoli cells does not show a change during recovery from irradiation. Thus, from this study we could not detect any effect of this low dose. Earlier studies find Sertoli cell death follow irradiation, but these were all studies with use of higher irradiation doses (247; 248).

Irradiation creates a gap in spermatogenetic germ cells

The cells eradicated by the irradiation were in our previous study estimated to be A1 through B spermatogonia (219). This is in good agreement with findings in other studies that investigate the effects of irradiation on germ cells (227; 229; 230; 231; 232; 233). The length of the cellular gap created by the irradiation thus corresponds to the time it takes for Aal spermatogonia to differentiate to B spermatogonia. The length of the gap was in both this and our previous study estimated to last around 10 days (219), which is in accordance with the literature (30). As spermatogenesis proceed, the surviving germ cells will continue their differentiation and leave the seminiferous epithelium as spermatozoa, but at the same time the recovery will proceed with the spermatogonia being replaced from SSCs (A_s). This creates a gap in the differentiation stages of germ cells present in the testis that can be followed during the recovery.

Cell-specific gene clusters

Gene expression profiling showed that five clusters could be resolved each with transcripts with similar gene expression patterns during recovery from irradiation. We assigned the transcripts in each cluster to be transcribed primarily from five specific cell types in the testis: *(1)* spermatogonia, *(2)* spermatocytes, *(3)* early round spermatids, *(4)* late elongated spermatids, and *(5)* somatic cells (Figure 8.2 and 8.3). The expression of the transcripts in each cluster follows the gap reaching the cells assigned to the cluster. The transcriptional changes observed in this study are caused by a combination of events that change the total RNA pool extracted from the testis. The effects of irradiation on the cells influence the RNA pool in combination with gene expression changes as observed in healthy testis during spermatogenesis. Yet, the by far major reason for changes in the total RNA pool is caused by the gap moving along the spermatogenesis with the disappearance and reappearance of specific cell types. It is a complex data set and impossible to distinguish these changes from each other. So although we assign each cluster to a specific cell type that properly express most of the transcripts in the cluster, other cells in the testis will very likely also express some of the transcripts present.

Four of the five clusters identified corresponded to the four clusters previously identified with cell-specific markers by differential display and ISH (219). By the cluster analysis of the gene expression data, we were able to identify an additional fifth cluster. Transcripts of four of the clusters are associated with germ cells: spermatogonia, spermatocytes and early and late spermatids. Spermatogonia that have entered spermatogenesis and spermatocytes are highly dividing cells that express a lot of genes needed for the mitotic and meiotic divisions, respectively. A previous study of RNA from purified adult cell types reported 405 transcripts expressed by spermatogonia and 442 transcripts expressed by spermatocytes (249). We associated 109 transcripts with spermatogonia and 164 transcripts with spermatocytes from our analysis. The number of transcripts expressed by the two germ cell populations can be discussed, but for many of the genes we observed a mixed cellular expression e.g. a relatively low expression that initiates in spermatogonia but is followed by a massive increase in spermatocytes (for example Deleted in azoospermia-like (Dazl)).

We identified two spermatid clusters, an early and a late, corresponding to round and elongated spermatids. A massive wave of transcriptional activity in seen after meiosis in early spermatids before the chromatin is condensed and transcription is silenced during differentiation to late elongated spermatids (250; 251). The late elongated spermatid cluster identified in this study might contain transcripts expressed in early round spermatids, which then later accumulate and to a greater extent dominate the RNA pool from elongated spermatids. The further matured spermatozoa have a highly condensed transcriptionally inactive chromatin and therefore no cluster is identified representing these cells (252). We tried to separate more germ cells clusters from the data, but in trying to do so we came across uncertainties of whether it was actual clusters of just noise that made the clusters separate.

The somatic cell cluster mostly contains transcripts expressed from the somatic cells in the testis and we identified PTM, Leydig and Sertoli cell markers in this cluster. Looking at the gene expression pattern of the cluster during the time series it seems that the somatic cells undergo large expressional changes (Figure 8.2 and 8.3). Yet, the majority of the somatic cells are not affected by the low dose of irradiation and do to not change in their expression. Instead, the gene expression changes detected in the cluster are caused by the absence of other cell types in particular pachytene spermatocytes, which leads to the apparent upregulation of genes in the somatic cells. Thus, the gene expression change is caused by a change in cellularity, because Sertoli cells comprise a larger part of the testis when pachytene spermatocytes are missing due to the gap.

Gene set enrichment analysis

We performed a gene set enrichment analysis of the transcripts in each cluster separately to gain knowledge of the function of the proteins that they encode. All enriched categories are listed in Table 8.2 and a few categories are highlighted in this section. Yet as stated above, there is some overlap between clusters and transcripts expressed by several cell types are properly present in each cluster. Therefore, the result from the gene set enrichment analysis will not solely correspond to functions of the cell associated with the respective cluster. The spermatogonia cluster was enriched with genes involved in acetylation and methylation, and chromatin modifications are involved in transcriptional regulation during the numerous divisions and differentiation that spermatogonia undergo (253). The multiple divisions are properly also the reason why the genes in the cluster are enriched in RNA binding and processing.

The genes in the spermatocyte cluster are enriched in spermatogenesis and cell cycle. Spermatocytes undergo the two meiotic divisions whereby haploid spermatids are produced (221). The enrichment in the cell cycle might be caused by genes expressed from spermatogonia that for some reason follow the expression pattern of the spermatocyte cluster.

The early spermatid cluster is enriched in acrosomal vesicle, histone core and spermatogenesis. During the maturation of early round spermatids to late spermatids and further to spermatozoa the acrosome is formed (254). The chromatin in the germ cells are also condensed whereby transcription is silenced (255), which makes the enrichment of the acrosomal versicle and histone core very relevant for this cluster. The late spermatid cluster is enriched in spermatogenesis and serine-type peptidase activity. Especially one serine protease, acrosin, has been identified as one of the important enzymes in the acrosome (256).

The somatic cells in the testis are dominated by Sertoli cells, but also includes the PTM cells that surround the tubule and the cells in the interstitium, which include microvasculature, lymph vessels, nerve fibres, fibroblasts, connective tissue and the hormone-producing Leydig cells (5). Thus, the

transcripts expressed by these cells have very diverse functions. The genes in the somatic cell cluster are enriched in steroidogenesis and lipid biosynthesis. The Leydig cells produce both estrogens and androgens and the enrichment in both lipid biosynthesis and steroidogenesis is in agreement with Leydig cell function (257; 258).

Similar timing of fetal and adult spermatogenesis

One of the aims of this study was to determine whether the duration of the differentiation stages in the fetal and adult mice spermatogenesis is the same. The duration of each developmental stage in pn spermatogenesis has been determined in several histological studies (30; 222; 223; 259). Also, studies using transcriptome profiling of the first wave of spermatogenesis in mice have been published (235; 236; 260). All studies agree on the approximate length of each differentiation stage to 6 days for spermatogonia, 14 days as spermatocytes, 9 days as spermatids, and 6 days as spermatozoa before the germ cells leave the testis. We used these studies to estimate whether we found similar patterns in duration of adult spermatogenesis. We defined the length of a differential stage in our study from the day with the lowest relative expression to the day with expression peak.

Comparison of the expression profiles found in the study by Almstrup et al. (2004) of the first wave of spermatogenesis with the spermatogonia and spermatid clusters found in this study revealed roughly the same length of the gene expression profiles in the two studies. The duration of the pn pachytene and spermatid clusters were from day 8 to 22 and day 20 to 34, respectively (235), both corresponding to 14 days. In this study, the adult spermatogonia cluster was estimated to 15 days (pi day 24 to 37), the early spermatid cluster to 13 days (pi day 24 to 37) and the late spermatid cluster to 15 days (pi days 27 to 42), respectively.

A study by Ellis et al. of the pn spermatogenesis, analyzed gene expression microarray data generated from two cDNA libraries specific for the testis and germ cells. The analysis revealed 24 clusters based on the expression profiles of 1900 transcripts during pn spermatogenesis. These clusters were further divided into ten subgroups with similar patterns associated with different testicular cell types (260). The study identifies a pn spermatogonial cluster that has the same bell shape as the adult spermatogonial cluster identified in this study (cluster 3, Figure 8.3). The expression of the pn spermatogonia transcripts starts to increase at pn day 6, peaks around day 10-15 and is back at start levels at day 23, thus, 17 days in total. The pn spermatocyte clusters increased a little from day 5 with a steep increase pn day 10 until day 23, which gives an approximate length of 13-18 days. In this study, the adult spermatogonia and spermatid clusters had durations of 13-15 days that corresponds well to the pn length. The study by Ellis et al. also identified clusters with expression patterns corresponding to both early and late spermatids (260), which empower our identification of two separate spermatid clusters. The early spermatid cluster starts to increase in expression around

pn day 10 with a steep increase from day 15 to 23 where after the increase reduces. The late pn spermatid cluster increases steadily from day 15 until 8 weeks. Thus, the pn spermatid clusters identified have longer durations than the spermatid clusters identified in this study of adult spermatogenesis.

The previous gene expression profiling study by Shima et al. identified five clusters with significant patterns by cluster analysis of 9846 transcripts that correspond to A spermatogonia, B spermatogonia, pachytene spermatocytes, round spermatids, and somatic cells (236). The somatic cell cluster was highly expressed from birth to pn day 6-8. Type A and B spermatogonia clusters started with high expressions pn day 0-3 with repression at day 14-18 that gives a length of the clusters around 15 days, which corresponds to the length of the adult spermatogonia cluster identified in this study. The pachytene spermatocyte cluster identified in the study by Shima et al. (2004) started to increase around pn day 14-18 and the round spermatid cluster started to increase in expression from day 20, both of which are highly expressed throughout adulthood, which where that same observed in the study by Ellis et al. (2004).

The somatic cell cluster identified in this study is difficult to compare with the patterns of the somatic cell clusters identified in studies of pn spermatogenesis (235; 236; 260), because the testis at birth only contains Sertoli, PTM and Leydig cells that may still resemble fetal Leydig cells in addition to early spermatogonia. These cells along with the other somatic cells in the testis thus express all the transcripts detected in the beginning of pn spermatogenesis (222; 223). Thus, the somatic cell clusters observed in pn spermatogenesis studies start of by having a high expression that decreases until the germ cells dominate the testis. The observed time point for the lowest expression varies among the studies from pn day 14 to day 23 (235; 236; 260). In this study of the adult testis, somatic cells and a mixture of germ cells are present at all times during the times series with the gap moving along the germ cells. Thus, the somatic cell cluster only increases in expression from day 12 to 32 because of changes in cellularity of the testis, which is caused by the absence of pachytene spermatocytes.

Conclusion

Irradiation of the adult testis with 1 Gy lead to the death of A1 to B spermatogonia, which creates a gap in the germ cells in the testis that as spermatogenesis proceeds lead to the absence of specific germ cell differentiation types. We found that Sertoli cells did not seem affected by the irradiation, whereas the Leydig cells reacted to the irradiation with a hyperproliferation at pi day 14-28. We identified five clusters each containing transcripts with similar gene expression patterns during recovery from irradiation. The five clusters were associated with: spermatogonia, spermatocytes, early round spermatids, late elongated spermatids and somatic cells. We found by comparison with published studies of pn spermatogenesis that adult and pn spermatogenesis have similar duration and timing.

Acknowledgements

Ana Ricci Nielsen for skilful technical assistance.

Chapter 9

Testis CIS-specific miRNAs

9.1 Prelude

Testis cancer is one of the four phenotypes in TDS that all are assumed to evolve from fetal male maldevelopment. TGCTs accounts for approximately 95% of testis cancers (39) and develop from the common precursor cell, the CIS cell. CIS cells are believed to be gonocytes that were arrested during fetal development because of genetic predisposition and/or hormonal disturbances as the other three TDS phenotypes. CIS cells are pre-malignant and lie quiescent in the testis until puberty where they are stimulated for growth and differentiation by the increased testosterone level that trigger puberty and spermatogenesis.

In the study presented next, we focused on the miRNA signature in testis CIS cells. The exact mechanisms whereby CIS cells transform into malignant TGCTs still need to be elucidated. Therefore, it was of our interest to investigate whether miRNAs might be involved in the carcinogenesis. We also wanted to investigate whether CIS cells have an miRNA profile that resembles the one expressed in fetal gonocytes. The results will be presented in the manuscript *"MicroRNA expression profiling of carcinoma in situ (CIS) cells of the testis"* that was submitted before I handed in my thesis.

The project was a collaboration with Department of Growth and Reproduction, Rigshospitalet, University of Copenhagen where most of the experimental work was carried out. I performed the miRNA target prediction of the subset of interest, analyzed the gene expression microarray data from the microdissected cells, investigated miRNA target expression in this dataset as well as I performed the gene set enrichment analysis.

9.2 Manuscript

MicroRNA expression profiling of carcinoma *in situ* (CIS) cells of the testis

[1,2*]GW Novotny, [3*]KC Belling, [4]JB Bramsen, [1]JE Nielsen, [5]J Bork-Jensen, [1]K Almstrup, [1,6]SB Sonne, [4]J Kjems, [1]E Rajpert-De Meyts and [1,6]H Leffers

[1]Department of Growth and Reproduction, Rigshospitalet, Copenhagen, Denmark.
[2]Present address: Biomedical Institute, University of Copenhagen, The Panum Institute, Copenhagen, Denmark.
[3]Center for Biological Sequence Analysis, Institute for Systems Biology, Technical University of Denmark, Lyngby, Denmark.
[4]Department of Molecular Biology, Aarhus University, Aarhus C, Denmark.
[5]Steno Diabetes Center A/S, 2820 Gentofte, Denmark.
[6]Present address: Department of Biology, University of Copenhagen, The August Krogh Building, Copenhagen, Denmark.

*These authors contributed equally to the study.

Abstract

Testicular germ cell tumours (TGCTs) of young adult men, seminoma (SEM) and nonseminoma (NSEM), develop from a precursor cell, carcinoma in situ (CIS), which resembles foetal gonocytes and retains embryonic pluripotency. We used microarrays to analyse microRNA (miRNA) expression in 12 human testis samples with CIS cells and compared it to the miRNA expression profiles of normal adult testis, testis with Sertoli-cell-only (SCO) that lack germ cells, testis tumours (SEM and embryonal carcinoma, EC; an undifferentiated component of NSEM) and foetal male and female gonads. Principal component analysis revealed distinct miRNA expression profiles characteristic for each of the different tissue types. We identified several miRNAs that were unique to testis with CIS cells, foetal gonads, and testis tumours. These included miRNAs from the hsa-miR-371-373 and -302-367 clusters that have previously been reported in germ cell tumours and three miRNAs (hsa-miR-96, -141 and -200c) that were also expressed in human epididymis. We found several miRNAs that were upregulated in testis tumours: hsa-miR-9, -105 and the hsa-miR-182-183-96 cluster were highly expressed in SEM, while the hsa-miR-515-526 cluster was high in EC. We conclude that miRNA expression profile changes during testis development and that the miRNA profile of

adult testis with CIS cells shares characteristic similarities with the expression in foetal gonocytes.

Introduction

Testicular germ cell tumours (TGCTs) are the most common malignancies of adolescents and young men, with the majority of cases occurring from ages 18 to 45 (9). Testis cancers are treatable with a survival rate of 50-90% depending on tumour type and progression stage (261). However, testis cancer patients often have reduced fertility, and orchidectomy of the affected testicle and cisplatin treatment can result in a further reduction of their fertility. Survivors in addition suffer from a range of treatment effects, which significantly reduce later quality-of-life (262).

TGCTs arise from precursor cells named carcinoma in situ (CIS), which are of foetal origin (42). The CIS cells persist quiescently in the testis until puberty where hormone signals initiate spermatogenesis. These signals apparently result in the proliferation of the CIS cells, which progress to overt tumours that are classified as either seminoma (SEM) that generally do not differentiate or non-seminomas (NSEM), which have a striking ability to differentiate into virtually any type of tissue (263). The pluripotent nature of the NSEM tumours is likely inherited from the CIS cells, as gene expression studies have shown a significant overlap of gene expression between CIS cells and embryonic stem cells (ESCs) (264). Visually, CIS cells resemble foetal gonocytes (265), and recent investigations of microdissected tissues showed a close resemblance in gene expression between CIS cells and gonocytes (266). This indicates that CIS cells arise from foetal gonocytes, which failed to differentiate to pre-spermatogonia, and instead survived as quiescent gonocyte-like cells in the testis. The mechanisms behind the survival of gonocyte-like CIS cells in adult testis and how they develop into either SEM or NSEM tumours are unclear.

During the last years it has become widely appreciated that much of a cell's biology and function can be regulated at the translational level through a class of small RNAs, microRNAs (miRNAs). Most miRNAs specifically inhibit translation of mRNAs through binding to complementary sequences in the 3'-UTR (60; 267; 268), while some can enhance translation through binding to regulatory elements on the mRNA (62). In addition, binding of miRNAs usually also lead to reduction of the level of their target mRNAs (63). Deregulation of miRNAs that regulate tumour suppressors or oncogenes often play a role in cancer development (269; 270). In testis the oncogenic miRNAs hsa-miR-372, and -373 contribute to cancer development by disabling the p53 pathway through lowering LATS2 protein levels (68), while another oncogenic miRNA cluster, hsa-miR-17-92, (271; 272) may promote development of overt tumours through prevention of apoptosis by inhibiting E2F1 protein synthesis in CIS cells (273). Additionally, recent quantitative reverse transcriptase-PCR (qRT-PCR) studies of miRNA expression in germ cell tumours have shown differential expression of several miRNAs between

normal tissue and tumour samples (69; 70). Taken together these studies indicate that miRNA regulation may play a role in development of TGCTs, however no data on miRNA expression in the CIS cell is currently available. Achieving this information is not trivial as the CIS cell is a rare cell type, comprising at most 5-10% of the cells in the testis, and the techniques applicable for purifying CIS cells through microdissection are not suitable for detecting miRNAs.

The aims of this study were to identify the miRNA expression profile in CIS cells and investigate the role of individual miRNAs in tumour development and progression. For this purpose we applied microarray technology to detect miRNAs expressed in CIS cells by comparing normal human testis samples with samples containing varying amounts of tubules with CIS. We also performed miRNA analysis of a series of TGCT: SEM and embryonal carcinoma (EC), which is the pluripotent undifferentiated component of NSEM, and of male and female foetal gonads. Combining the miRNA expression data with data on mRNA expression in microdissected testicular tissues and CIS cells allowed us to perform a bioinformatics investigation to predict putative targets for the miRNAs expressed in CIS cells. Finally, we looked into the function of the predicted targets by miRNA enrichment analysis to gain further knowledge of the role of the miRNAs and their targets in CIS cell biology.

Materials and Methods

Tissue samples and RNA preparation

Use of adult testicular tissues was approved by the Regional Committee for Medical Research Ethics in Denmark. The samples were obtained from residual tissue of the orchidectomy specimens from patients diagnosed with malignant testicular cancer. Normal samples were obtained from areas with preserved complete spermatogenesis, where no tumour or CIS cells were present, or purchased commercially (Applied Biosystems/Ambion, CA, USA, and Biochain Institute, CA, USA). The tissue samples were either snap-frozen at -80°C or fixed overnight at 4°C in Stieve's fluid or paraformaldehyde, and subsequently embedded in paraffin. The samples were stained with hematoxylin-eosin for histological evaluation and with an antibody against placental alkaline phosphatase to confirm the presence of CIS, as previously described (274). Foetal gonads were collected in the United Kingdom according to Polkinghorne guidelines following ethical approval and informed consent of women who underwent elective abortions at 10-12 weeks of pregnancy as described previously (266). Three foetal gonadal samples were available for this study: two male (of which one contained testis and mesonephros tissue and the other only mesonephros) and one female (containing ovary and mesonephros tissue) (Table 9.1).

Table 9.1: **Tissue and RNAs used in this study.**

Biopsy ID	Histology	Estimated % CIS tubules[1]	Real % CIS tubules[2]	Gene expression data[3]	Comments
Control tissue					
SF5144 V-II	Leydig cell tumour (with normal rest tissue)	0			No CIS cells present
SF5297 V-II	Leyiomyosarcoma (with normal rest tissue)	0			No CIS cells present
Sertoli-cell-only (SCO)					
SF6237 H-II	Atrophy with SCO and hyalinization	0		Yes, Sertoli cells*	
SF5962	SCO			Yes, Sertoli cells*	
SF6310 H	SCO	0- ?%		Yes, Sertoli cells*	Extragonadal germ cell tumour, CIS cells found in testis.
Carcinoma in situ (CIS)					
SF6193 H	CIS only	20-70%	50%		miRNA array experiment 1
SF6220 V	CIS only	70-100%	99%	Yes, CIS cells*	miRNA array experiment 1
SF6246 H-II	CIS/SEM	40-50%	30%		miRNA array experiment 1
SF5472 H	CIS only	40-80%			
SF5954 V	CIS only	0-50%			
SF6204 V	CIS only	50-90%			
SF5967 V-II	CIS/SEM	10-50%			
SF5514-V-II	CIS/SEM	50-80%			
SF5814 V-I	CIS/EC	60-75%			
Seminoma (SEM)					
SF5967 V-I	CIS/SEM	10-50%		Yes, CIS cells*	
SF5703 V-I	CIS/SEM	40-60%		Yes, CIS cells*	
SF5514-V-I	CIS/SEM	50-80%			
Embryonal carcinoma (EC)					
SF5474 V-I	CIS/EC	95-99%			
SF5814 V-I	CIS/EC	60-75%			
Male Foetal Gonad					
764	week 10-11			Yes, gonocytes*	
752	week 12			Yes, gonocytes*	
Female Foetal Gonad					
473	week 12-13			Yes, oocytes*	

Purchased Normal Testis RNA					
Ambion					
Biochain Institute					

[1]Estimated from paraffin section; [2]Real % in frozen tissue; [3]From microdissected cells; *Sonne *et al.*, 2009 (266).

The foetal gonads were embedded in optimum cutting temperature compound (Sakura Fintek Europe) and snap-frozen at -80°C. Total RNA was prepared from minor amounts of the frozen tissue samples (adult testis) or from serial sections of the tissue (foetal gonads) using Trizol reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer s protocol.

miRNA array analysis and qRT-PCR expression profiling

The Agilent (Agilent Technologies, CA, USA) Human miRNA microarray v1.0 (4 samples: 3 CIS and 1 normal adult testis) and Agilent Human miRNA microarray v2.0 (24 samples: 9 CIS, 4 normal adult testis, 3 SCO, 2 EC, 3 SEM, and 3 foetal samples) were used for profiling the miRNA expression (Table 9.1). One hundred nanogram of total RNA from each sample was labelled and hybridized to the arrays following Agilent's instructions. Array images were scanned using Agilent scanner G2565BA with scanner software version A.7.0.1 with XDR function, and data obtained using Agilent Feature Extraction software version 9.5.3.1. The per-array lowess normalized signal (gProcessedSignal) from each array was normalized to each other by quantile normalization using the marray package in R/bioC version 2.7.2. Statistical analysis (Hierarchal clustering and significance analysis of microarrays (SAM)) of the normalized data was performed using TIGR MeV v4.0 (275), while principal component analysis (PCA) was performed using R/bioC version 2.7.2.

For the verification of the microarray results we performed qRT-PCR validation of 9 miRNAs using TaqMan miRNA assays (Applied Biosystems, CA, USA) as described by the supplier; hsa-miR-103 was used for normalisation.

In situ hybridization (ISH)

ISH probes for of pri-hsa-miR-96 and pri-hsa-miR-371 were prepared using two sets of specific primers for each. The first set (to amplify the cDNA) was for pri-hsa-miR-96: AGAGTGTGACTCCTGTTCTGT and TTGGCACTGCACATGATTGCT; and for pri-hsa-miR-371: GATCGCCGCCTTGCCGCA and TCGTGATGCCCTACTCAAACA. One µl from the first PCR sample was used for the 2nd amplification (for adding T3 or T7 promoters, respectively (underlined)): For pri-hsa-miR-96: AATTAACCCTCACTAAAGGGTATGGCACTGGTAG and TAATACGACTCACTATAGGGCAGAGCGGAGAGA and for pri-hsa-miR-371: AATTAACCCTCACTAAAGGGCACTCAAACTGT and TAATACGACTCACTATAGGGTGACGCTCAAAT. PCR amplification for the 1st set was: 95°C

for 5 min. 35 cycles of 95°C for 30 sec., 62°C for 1 min. and 72°C for 1 min. followed by 5 min. at 72°C. PCR for the 2nd set was: 95°C for 5 min. followed by 5 cycles of 95°C for 30 sec., 45°C for 1 min. and 72°C for 1 min., followed by 20 cycles of 95°C for 30 sec., 65°C for 1 min. and 72°C for 1 min.. After a final 5 min. extension step at 72°C the reaction was stopped. The PCR product was purified on an agarose gel and sequenced. Aliquots of 200ng were used for in vitro transcription labelling of the RNA probes using MEGAscript-T3 (sense) or MEGAscript-T7 (antisense) following the manufacturers protocol (Ambion, Applied Biosystems, CA, USA). ISH was performed as previously described (276; 277).

miRNA target prediction and enrichment analysis

miRNA targets were predicted for the 13 miRNAs from the first miRNA study (Table 9.2), as well as for hsa-miR-141, hsa-miR-200c and hsa-miR-183 using miRecords, an integrated database of 11 prediction algorithms.

Table 9.2: **Expression of miRNAs correlated to the percentage of CIS tubules.**

miRNA	Normal	CIS30	CIS50	CIS100	R*	Upreg. stem cell/cance	Reference
hsa-miR-367	3	85	236	395	0.987	iPSC, hESC, Cancer	(69; 278)
hsa-miR-371	0	253	969	1917	0.983	hESC, Cancer	(68; 278)
hsa-miR-373	8	510	1917	3520	0.982	hESC, Cancer	(68; 278)
hsa-miR-96	1	6	18	47	0.978	hESC, Cancer	(278; 279)
hsa-miR-302b*	-2	-1	4	12	0.975	iPSC, hESC, Cancer	(69; 278)
hsa-miR-183	1	4	17	37	0.972	Cancer	(69; 246)
hsa-miR-302b	-1	26	88	128	0.967	iPSC, hESC, Cancer	(69; 278)
hsa-miR-302d	-1	28	113	168	0.963	iPSC, hESC, Cancer	(69; 278)
hsa-miR-302a*	-1	30	94	130	0.96	iPSC, hESC, Cancer	(69; 278)
hsa-miR-302a	-1	45	152	203	0.953	iPSC, hESC, Cancer	(69; 278)
hsa-miR-372	10	801	3656	5095	0.95	hESC, Cancer	(68; 278)
hsa-miR-302c	-1	4	26	34	0.933	iPSC, hESC, Cancer	(69; 278)

iPSC: Induced pluripotent stem cell; hESC: human embryonic stem cell.
*Pearson's correlation coefficient

Our criterion for the target prediction was that each target was predicted by at least six of the 11 prediction algorithms. Gene set enrichment analysis of the miRNA targets were performed using Panther Classification System (http://www.pantherdb.org/tools/compareToRefListForm.jsp). As reference we used all the identifiers on the array.

Microarray analysis of mRNA expression

Microarray data from three samples of microdissected CIS cells and three samples of total RNA from normal testis, which were obtained in our previous study (266) were available for this study. The microarray images were read by the Agilent feature extraction software version 9.5.3.1 and analysed in the R/BioC limma package using the gMedianSignals. Normalization between arrays was done using a quantile normalization procedure and probes were collapsed by taking the median. Statistical analysis was done by t-test and the p-values were corrected for multiple testing by Benjamini Hochberg correction.

Results

miRNA microarray results

To identify the miRNAs expressed in CIS cells, we performed an initial array experiment applying a procedure previously successfully used for mRNA profiling of CIS (264). Using RNA from one normal testis biopsy and three testis biopsies containing variable amounts of CIS cells, we could determine CIS-enriched miRNAs by correlating miRNA expression with CIS cell content. To visualize and quantify the CIS cell content prior to purifying RNA from the tissue, we applied a fast NBT-BCIP staining protocol (280) on sections from the CIS samples. We estimated that samples 1-3 contained CIS cells in 30%, 50% and 99-100% of the tubules, respectively (Table 9.1 and data not shown). We then performed miRNA expression profiling on the Agilent Human miRNA microarray v1.0 platform (containing probes for 470 human miRNAs and 64 human viral miRNAs). The expression levels were then correlated to the CIS cell content and the miRNAs correlating most significantly to the CIS content and showing no or very low expression in normal testis are shown in Table 9.2 (for a complete list of correlations see Suppl. Table 9.3). The miRNAs that displayed increasing expression with increasing CIS content and low expression in normal testis were miRNAs previously reported in stem cells and cancer (Table 9.2).

Following the initial experiments, we expanded our investigation by applying Agilent Human miRNA microarray v2.0 (containing probes for 723 human miRNAs and 76 human viral miRNAs) to profile miRNA expression of 24 human tissue samples (Table 9.1) divided between normal testis (n=4), testis with CIS cells (n=9), non-cancerous biopsies of SCO testis which lack germ cells (n=3), TGCTs (SEM, n=3 and EC, n=2) and foetal gonads (n=3)

(Table 9.1). Unsupervised hierarchal clustering (UHC) resulted in overlapping but distinct clustering of the different testis tissue subtypes (Figure 9.1a), indicating the presence of unique miRNA signatures in each respective tissue group. PCA (Figure 9.1b) also revealed discrete groupings of the tumour types (EC and SEM) and the foetal gonads, while the samples consisting of normal testis, SCO, and CIS formed a large group. The grouping of CIS with non-cancerous adult testes was expected, because at least 90% of the cells in the CIS samples were normal testicular cells (mainly somatic cells). Several of the samples were paired, where the samples of the tumour and the adjacent testis tissue with CIS originated from the same patient (Figure 9.1a and 9.1b, compare patient numbers). Although originating from identical genetic background, the miRNA signatures between tissue containing CIS and the corresponding tumour were sufficiently different to allow sorting according to tissue type rather than by patient in the UHC and PCA plots.

High expression levels of the testis/gonadal specific miRNA hsa-miR-202 (281; 282) was seen in all testicular samples, while the tumours showed varying levels of expression (Figure 9.1c). Of the foetal samples, the foetal gonad male_2 had a very low level of hsa-miR-202, indicating that it mainly consisted of tissue from the mesonephros (an embryonic kidney-like organ that eventually develops into the epididymis and seminal vesicles). Immunohistochemistry with anti-müllarian hormone (AMH) confirmed that the sample was of male origin (data not shown). The embryonic hsa-miR-17-92 cluster (283; 284; 285) was highly expressed in the foetal gonad samples, indicating their embryonic origin (Figure 9.1c).

miRNAs differentiating between tissue subtypes

To find miRNAs differentially expressed between the tissue samples we performed a multifactorial SAM analysis among the seven distinct tissue groups (group 1, samples with normal spermatogenesis; group 2, biopsies with SCO; group 3, biopsies with CIS; group 4, SEMs; group 5, ECs; group 6, purchased normal testis RNA; group 7, foetal gonads) resulting in a list of 193 microRNAs (Figure 9.1d). The multifactorial SAM analysis revealed a group of testis-specific miRNAs (hsa-miR-506-510, -513abc and -514) (286) that were highly expressed in all adult testis samples (Figure 9.2a). Interestingly, the foetal gonad male_1 sample (gestational week 10-11) had a high expression of the miRNAs while the female foetal gonad (gestational week 13-14) showed very low expression, indicating that these miRNAs may have a very early role in testis development. In accordance with the testis-specificity of this cluster (286), the foetal gonad male_2 sample, consisting of mesonephros, showed no expression of these miRNAs and the SEM and EC show greatly reduced levels, suggesting either that the miRNAs are expressed in somatic cells, which is supported by their presence in SCO, or a loss of expression during tumour formation. The SAM analysis also revealed several miRNAs that were present at much higher levels in the commercial testis RNA samples than in normal testis samples from patients (Suppl. Figure 9.4a). Based on the miRNA

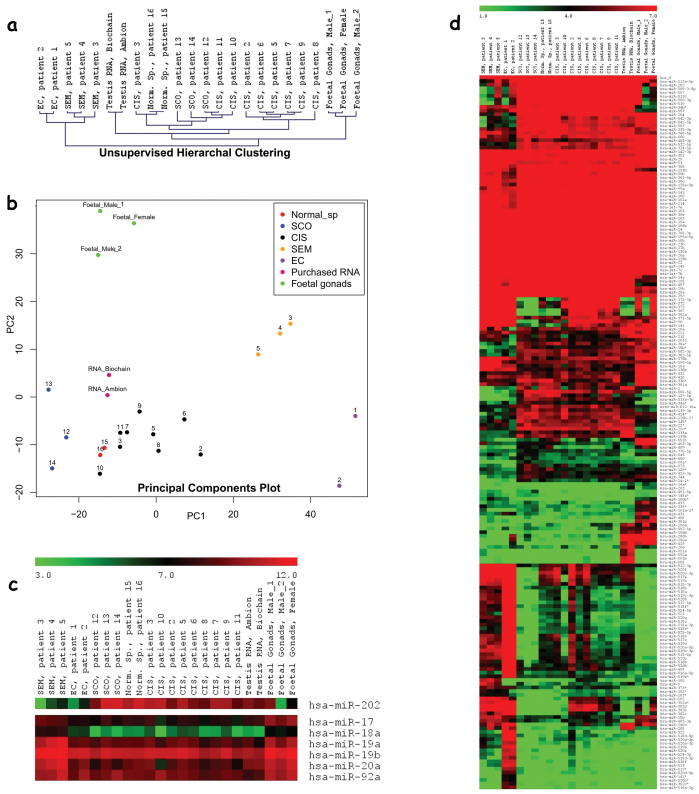


Figure 9.1. Visualization of the miRNA microarray data. (a) Unsupervised hierarchal clustering of all samples based on the expression of all probes: EC: Embryonal Carcinoma; SEM: Seminoma; CIS: Carcinoma *in situ*; Sp: Spermatogenesis (cells inside tubules without peritubular cells); SCO: Sertoli-cell-only. (b) Principal components analysis based on all expression data from all the samples. Numbers correspond to the Patient identifiers. (c) Log_2 fold heat map of the gonadal miRNA miR-202 and the embryonal miRNA cluster miR17-92. (d) Log_2 fold heat map of all differentially expressed miRNAs in the 7-way SAM analysis (delta 0.288; FDR 0.0%).

expression atlas (281), it was evident that these miRNAs corresponded to miRNAs expressed in the epididymis suggesting that the commercial testis RNA included RNA from both testis and epididymis (Suppl. Figure 9.4b).

To identify miRNAs important for the development of CIS we made a heat map of the expression of the miRNAs that showed a correlation higher than 0.8 with the percentage of CIS tubules (Table 9.2 and Suppl. Table 9.3) (Figure 9.2b) and performed a SAM analysis between the CIS samples and adult non-malignant testis samples (Normal spermatogenesis + SCO). We

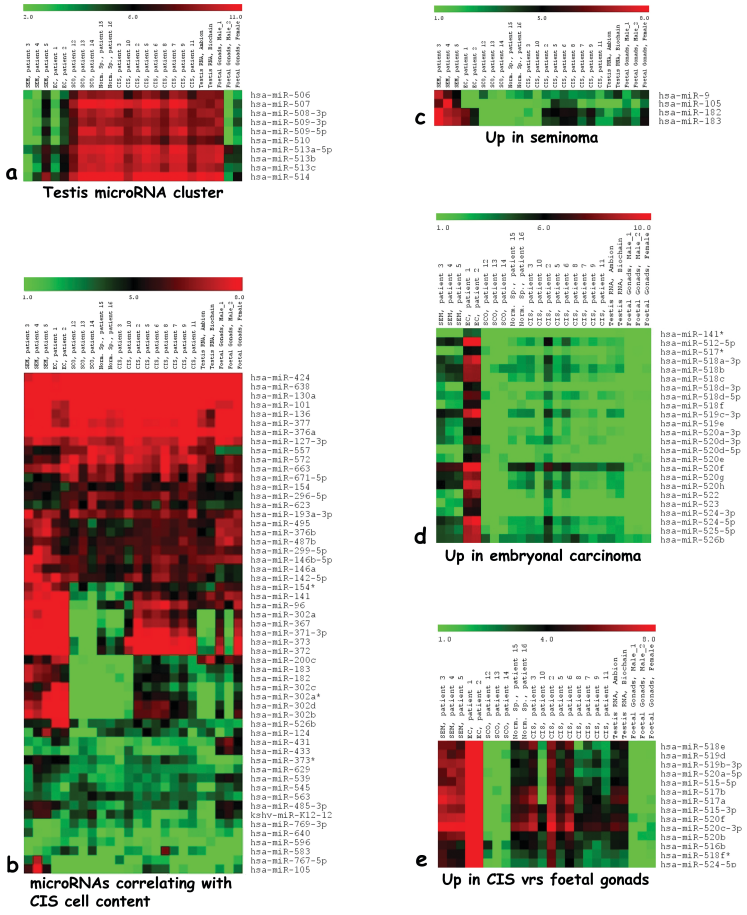


Figure 9.2. miRNAs differentially expressed between samples. (a) Cluster of testis-specific miRNAs. (b) Heat map of miRNAs correlating with CIS cell content. (c) miRNAs upregulated in seminomas (SEMs). (d) miRNAs upregulated in embryonal carcinomas (ECs). (e) miRNAs upregulated in CIS samples compared to foetal gonads, selected from a 2-way SAM analysis (CIS samples vs. foetal gonads, delta 0.588; FDR 0.0%).

observed a very consistent co-expression of almost all the miRNAs in samples with CIS cells and foetal gonads (i.e. a moderate to high expression in CIS samples and moderate to high expression in the foetal samples), where 50 of the 55 miRNA shown in Figure 9.2b also were expressed in the the foetal samples. Moreover, all the miRNAs we found in the SAM analysis were already in the list of miRNAs that correlated with the percentage of tubules with CIS (results not shown). The miRNAs enriched in CIS cells included

several miRNAs that were low expressed or absent from normal testis; these included the hsa-miR-371-373 and -302-367 clusters, which also have been described in germ cell tumours (68; 69; 70) and in ESC (287) and hsa-miR-182, -183 and -200c, which are expressed in the epididymis (Suppl. Figure 9.4).

We further investigated the differences between the CIS samples and foetal gonads in the multifactoral SAM analysis (Figure 9.2e and results not shown). However, the upregulated miRNAs were also expressed in the mesonephros sample (foetal gonad, male_2) indicating that these miRNAs did not originate from gonocytes. Similarly for the miRNAs downregulated in the foetal samples, their low expression compared to the CIS samples can be attributed to the presence of normal adult germ cells in the CIS samples, which is supported by the lack of expression in SCO. The most intriguing observation was a cluster of miRNAs (hsa-miR-515-526) that was not expressed in the foetal samples, expressed in samples with CIS and in SEM and highly expressed in EC (Figure 9.2d and e). Notably, they were not expressed in the SCO samples, but their presence in the normal testicular samples suggested a role for these miRNAs in germ cells and maybe during the transformation of CIS cells to overt tumours.

To gain insight into how CIS cells progress into overt tumours we performed SAM analysis between the CIS samples and SEM and EC tumours (Figure 9.2c and 9.2d). There were some differences in the expression in SEM and EC: the hsa-miR-9 and -105 were high in SEM, hsa-miR-182-183-96 cluster was high in SEM and EC (Figure 9.2b, c and d) while and the hsa-miR-515-526 cluster generally was higher expressed in EC than in SEM and CIS and very low or absent in foetal testis (Figure 9.2d).

Validation of array results

We selected eight miRNAs (hsa-miR-105, hsa-miR-141, hsa-miR-200c, hsa-miR-34c-5p, hsa-miR-367, hsa-miR-371-3p, hsa-miR-520c-3p and hsa-miR-96) for verification of the array data by qRT-PCR analysis. There was a strong correlation between the miRNA expression detected by microarrays and the qRT-PCR results (Figure 9.3a and b). We performed ISH experiments to assign expression of the selected miRNA precursors to the distinct cell types. We detected the pri-hsa-miR-182-183-96 cluster transcript in the nuclei of CIS cells (Figure 9.3c), while we were unable to detect transcripts of the hsa-miR-302-367 and hsa-miR-371-373 clusters, presumably because of scarce amounts of the precursor due to fast processing to the mature miRNA forms. For the hsa-miR-371-373 cluster, ISH detected cytoplasmic transcripts from the opposite strand (Figure 9.3d), which is in accordance with data from Ensembl (ensembl.org) that show several transcripts (DB443882, AI825624 and AW833903) from the opposite strand are expressed in the testis and testis cancers.

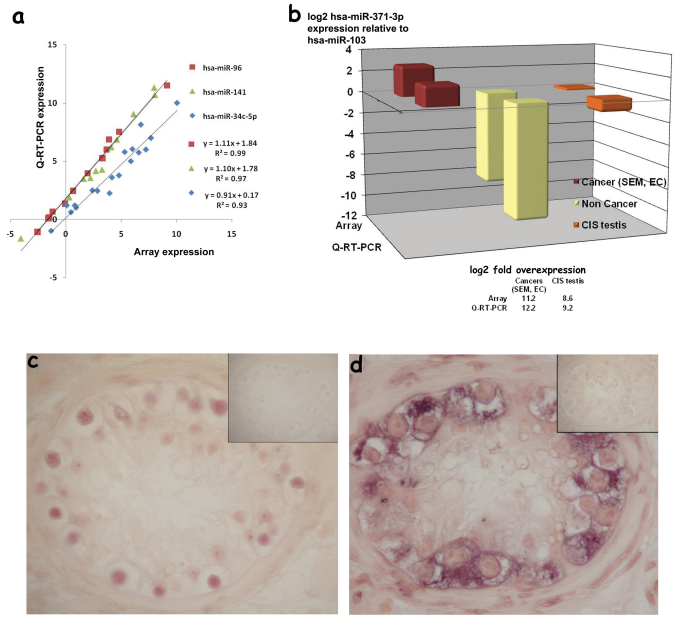


Figure 9.3. miRNA microarray validations. (a) qRT-PCR verification of hsa-miR-96, hsa-miR-141 and hsa-miR-34c-5p. Log_2 miRNA qRT-PCR expressions plotted against log_2 miRNA array expressions (both normalised to hsa-miR-103) from the same tissue samples. (b) qRT-PCR validation of fold-changes of hsa-miR-371-3p (normalised to hsa-miR-103). Log_2 expression values from microarrays and qRT-PCR from the same samples plotted next to each other for the cancers (red: n=5, STDEV 0.7 for array data, 2.12 for qRT-PCR data), non-cancers (yellow: n=4, array STDEV 0.84, 2.42 for qRT-PCR) and CIS (orange: n=3, array STDEV 1.37, 0.61 for qRT-PCR). Log_2 fold overexpression of hsa-miR-371-3p in cancers and testis CIS compared to non-cancers is shown below the plot. (c) *In situ* hybridization with pri-hsa-miR-96 transcript anti-sense probe on a testis CIS biopsy. pri-hsa-miR-96 transcript expression was detected in CIS cell nucleuses, while no staining of sense control RNA (inserted picture). (d) *In situ* hybridization with pri-hsa-miR-371-373 cluster sense probe, insert shows hybridisation with anti-sense probe. No expression of the primary transcript (anti-sense probe) could be detected, presumably due to fast processing of the transcript, while high expression from the opposite strand (sense probe).

Investigating biological function of miRNA expression

Since most mRNAs that are targeted by miRNAs are deadenylated and eventually degraded (63; 288; 289), potential targets can be discovered by mRNA profiling upon expression of a particular miRNA (70; 290; 291). Thus, we selected 16 miRNAs that were deregulated in CIS cells and testis tumours

(miRNAs in Table 9.3 as well as hsa-miR-141, -182, -200c and -323-3p; Figure 9.2b) for target prediction. We were able to predict targets for 13 out of 16 miRNAs using the database miRecords (http://mirecords.biolead.org/) with a total of 607 targets predicted. Using mRNA array data from specific microdissected testicular cell types (266) we then calculated the fold-change in expression of these mRNAs between CIS and normal testis. Investigating the full data-set of significantly changed genes (i.e. both up- and down-regulated genes, with no restrictions in fold-change levels) revealed 13,821 genes that were differentially expressed ($p<0.05$), with 7,757 genes upregulated in CIS and 6,064 genes down-regulated. Among the genes potentially targeted by the CIS-specific miRNAs, 412 were significantly changed, with 189 upregulated and 223 downregulated, which is significantly different from the expected numbers (231 up and 181 down) ($p<0.0001$, Chi2=17.581, df=1). To avoid unwanted noise in the calculation due to small fold-changes, the calculations were also performed after filtering out genes with fold-changes between -0.5 and +0.5. After this filtering, we found 5,237 up- and 3,905 down-regulated genes in the full data set of genes differentially expressed between CIS and normal testis, where 118 were up- and 150 downregulated amongst the miRNA targets, which is different from the expected values (154 up and 114 down) ($p<0.0001$, Chi2=19.244, df=1) indicating a significant enrichment of downregulated genes among the miRNA targets.

Assuming we could detect the influence of the miRNAs on the mRNA level of all their targets, we determined the expression changes (CIS cells vs. normal testis) of all miRNA target mRNAs (n=596) as a group compared to the rest of the annotated transcripts on the array (n=16,431). The expression of the miRNA-targeted mRNAs was significantly changed compared to the background (p=1.63e-8). To identify whether some of the miRNAs seemed to influence their targets more than others, we tested the miRNA target groups individually. The expression of two miRNA target groups, hsa-miR-141 (p=0.0037) and hsa-miR-200c (p=0.0003), was significantly changed compared to the background and hence could play a role in maintaining the CIS phenotype.

Discussion

In this study we have for the first time examined global miRNA expression in CIS, the precursor cell for testicular germ cell cancer. Since CIS cells only constitute a small percentage of the cells in the testis, we identified miRNAs that were expressed in CIS cells by identifying miRNAs whose expression correlated to the ratio of tubules with CIS cells. However, this may lead to an overestimation of the expression of miRNAs that are uniquely expressed in CIS cells as compared to miRNAs that also are expressed in other testis cell types (266). Nevertheless, this calculation likely identifies most miRNAs that mainly are expressed in CIS cells (Table 9.2 and Suppl. Table 9.3).

The results showed a strong correlation between miRNA expression in CIS cells and gonocytes, which is in accordance with the similar mRNA expression profiles (266). Especially, two miRNA clusters (hsa-miR-371-373 and

hsa-miR-302-367), which have been reported highly expressed exclusively in ESC (287) and in testicular (69) and extragonadal pediatric germ cell tumours (70; 292) were also expressed in CIS cells and in foetal gonads. This supports our hypothesis that CIS cells are arrested gonocytes that persist in the adult testis instead of differentiating to spermatogonia or entering apoptosis around birth (266). Nevertheless, we also identified a number of miRNAs that were differentially expressed between CIS and foetal gonads (Figure 9.2e) indicating that developmental arrest of gonocytes may be enforced by miRNAs. This is also further substantiated by the clear separation of CIS and foetal samples in the PCA plot (Figure 9.1c), but this may also be caused by the presence of normal adult cells in the CIS samples. The data confirmed a differential high expression of the hsa-miR-9, hsa-miR-105 and the hsa-miR-182-183-96 cluster in SEMs and the hsa-miR-515-526 cluster in ECs as previously suggested by Palmer et al., (2010). Although the highest expression appears to be in the tumours, we also observed a moderate expression of these miRNAs in testis with the CIS cells that later progressed to ECs or SEMs and in normal testis. The SAM analysis showed that the CIS samples that differed most from the normal spermatogenesis samples (SEM patients 5 and 6, and EC patient 2) (Figure 9.1c) were from patients who also harboured an invasive testis tumour, indicating that the miRNA expression in CIS cells is diverging towards the pattern seen in tumours. However, in CIS cells the expression of the clusters that are differentially expressed in SEM and EC does not seem to be predictive of the tumour type adjacent to the CIS cells.

The results suggested that hsa-mir-141 and hsa-mir-200c, which are not expressed in normal testis, but are expressed in the epididymis (Figure 9.2b), might be important for CIS cell biology and we used the enrichment tool at the Panther website (http://www.pantherdb.org/) to assign the mRNA targets of these miRNAs (84 targets for hsa-miR-141 and 157 targets for hsa-miR-200c) to Gene Ontology (GO) terms. Hsa-miR-141 targets were enriched in the biological processes chromatin architecture , organelle organization , and anterior/posterior axis specification . This enrichment indicated that hsa-miR-141 targets may be involved in many differentiation and developmental processes. The targets of hsa-miR-200c were enriched in the GO biological processes endoderm development, apoptosis, intracellular signalling cascade, and signal transduction and in GO molecular function categories associated with transcription factor activity, transcription cofactor activity, and DNA binding. Among the transcription cofactors targeted by hsa-miR-200c are Cited2, which is important in testis development (293).

The expression profiles confirmed the high expression of the embryonic and oncogenic hsa-miR-17-92 cluster in testis (Figure 9.1c) (273) and showed that the cluster already is expressed in foetal gonads of both sexes. The expression of the hsa-miR-17-92 cluster is regulated by MYC and several of the encoded miRNAs repress the expression of the transcription factor E2F1 (294). Thus, despite the abundant presence of the E2F1 transcript in meiotic germ cells and CIS cells (264), the hsa-miR-17-92 cluster prevents expression

of the E2F1 protein in CIS cells and in specific germ cell types during meiosis in normal testis (277). However, a survey of the E2F1 expression in 34 testis tumours revealed a moderate to strong E2F1 protein staining in 18-20 of the tumours (proteinatlas.org), which is surprising since testis tumours are derived from CIS cells. A potential explanation may come from a recent investigation of copy number variation in testicular seminomas reporting that the hsa-miR-17-92 cluster on chromosome 13q31.3 (overlap with GPC5) is deleted in 20-25% of the seminomas (295), which may indicate that regain of E2F1 expression in the absence of the hsa-miR-17-92 cluster may lead to the formation of some testis tumours.

The detection of miRNAs that in the testis are expressed exclusively in CIS cells and germ cells tumours may have clinical implications since the expression of developmentally regulated genes such as TFAP2C (AP2γ) and POU5F1 (OCT3/4) in CIS cells and germ cell tumours, but not in normal adult testis, has been exploited for screening of semen samples for pre-invasive testicular cancer (296; 297). Thus, the very restricted expression of the hsa-miR-371-373 and hsa-miR-302-367 clusters in CIS cells and germ cell tumours may open for a PCR-based screening of serum samples for testis cancer (298).

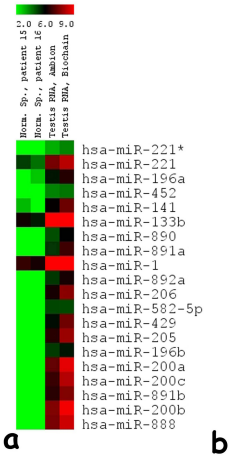
Funding

The work was supported by grants from the Villum Kann Rasmussen Foundation and the Danish Cancer Society.

Acknowledgements

The authors thank Sabina Soultanova and Brian Vendelbo Hansen for excellent technical assistance and Drs. Ludmila Ruban and Harry Moore for providing the foetal gonads.

Supplementary material



b

From Landgraf et al, 2007, Cell,
A mammalian microRNA expression
atlas based on small RNA library sequencing

miRNA	#Genes	Most matched sequence	Total copies	hsa_Epididymis	hsa_Testis
TOTAL			161006	1323	1373
hsa-miR-221	1	AGCUACAUUGUCUGCUGGGUUUC	225	5	0
hsa-miR-196a	2	UAGGUAGUUUCAUGUUGUUGGG	177	1	0
hsa-miR-452	1	AACUGUUUGCAGAGGAAACUGA	18	0	0
hsa-miR-141	1	UAACACUGUCUGGUAAAGAUGG	1307	12	0
hsa-miR-133b	1	UUUGGUCCCCUUCAACCAGCUA	4.5	0	0
hsa-miR-890	1	UACUUGGAAAGGCAUCAGUUG	10	9	0
hsa-miR-891a	1	UGCAACGAACCUGAGCCACUGA	24	22	0
hsa-miR-1	2	UGGAAUGUAAAGAAGUAUGUAU	25	0	1
hsa-miR-892a	1	CACUGUGUCCUUUCUGCGUAG	4	4	0
hsa-miR-206	1	UGGAAUGUAAGGAAGUGUGUGG	16	0	0
hsa-miR-582	1	3p UAACUGGUUGAACAACUGAACC	12	0	0
		5p UUACAGUUGUUCAACCAGUUACU	10	0	0
hsa-miR-429	1	UAAUACUGUCUGGUAAAACCGU	13	0	0
hsa-miR-205	1	UCCUUCAUUCCACCGGAGUCUG	143	27	0
hsa-miR-196b	1	UAGGUAGUUUCCUGUUGUUGGG	277	1	0
hsa-miR-200a	1	UAACACUGUCUGGUAACGAUGU	79	5	1
hsa-miR-200c	1	UAAUACUGCCGGGUAAUGAUGGA	955.5	47.5	0
hsa-miR-891b	1	UGCAACUUACCUGAGUCAUUGA	4	4	0
hsa-miR-200b	1	UAAUACUGCCUGGUAAUGAUGA	84.5	5.5	0
hsa-miR-888	1	UACUCAAAAAGCUGUCAGUCA	16	15	0
			3404.5	158	2
		Procent of miRNA content in tissue		11.94	0.15

Figure 9.4. miRNAs expressed in epididymis, but present in commercial testis RNAs . (a) Heat map of the expression of selected miRNAs in biopsies with normal spermatogenesis and in the testis RNA preparations from Ambion and the Biochain Institute. (b) Distribution of the miRNAs in testis and epididymis in the paper from Landgraf et al., 2007 (281).

Table 9.3: **miRNA with a correlation > 0.4 to percentage of tubules with CIS cells.**

Probe	Normal	CIS 30%	CIS 50%	CIS 100%	Corr.
hsa-miR-496	-7.390	-5.679	-4.520	-1.990	0.999
hsa-miR-368	230.548	316.190	358.706	509.543	0.999
hsa-miR-495	63.556	89.811	108.621	143.026	0.997
hsa-miR-545	-1.722	-1.070	-0.546	0.337	0.997
hsa-miR-154	78.172	98.748	106.318	131.546	0.995
hsa-miR-101	753.600	1053.293	1113.900	1489.670	0.991
hsa-miR-376a	1003.560	1242.125	1301.415	1766.765	0.991
hsa-miR-539	0.337	1.111	2.643	4.671	0.989
hsa-miR-367	-3.364	84.826	235.685	395.412	0.987
hsa-miR-146b	154.034	180.465	199.559	281.364	0.987
hsa-miR-371	-0.302	252.825	969.396	1916.780	0.983
hsa-miR-373	8.357	509.543	1916.780	3520.310	0.982
NC2_00092197	-9.231	-7.957	-6.500	-0.985	0.979
hsa-miR-96	0.872	6.167	17.857	47.396	0.978
hsa-miR-299-5p	128.014	136.649	163.513	190.599	0.978
hsa-miR-302b*	-2.186	-1.032	4.257	11.536	0.975
hsa-miR-629	11.016	19.728	26.261	32.625	0.973
hsa-miR-376b	54.427	104.550	140.945	177.093	0.973
hsa-miR-183	1.140	3.514	16.986	37.209	0.972
hsa-miR-302b	-0.876	25.859	87.560	128.014	0.967
kshv-miR-K12-12	-5.349	-2.824	-2.846	-0.496	0.967
hsa-miR-453	3.265	4.671	4.671	8.686	0.964
hsa-miR-302d	-1.390	27.663	113.026	167.807	0.963
hsa-miR-577	-7.957	-7.390	-7.039	-4.242	0.961
hsa-miR-675	-2.914	-2.846	-1.837	0.096	0.960
hsa-miR-302a*	-1.143	30.491	94.440	129.669	0.960
hsa-miR-377	341.863	701.075	647.419	1044.368	0.958
hsa-miR-302a	-0.546	45.381	151.762	202.846	0.953
hsa-miR-769-3p	10.054	10.054	11.804	15.440	0.953
hsa-miR-193a	84.826	143.026	128.014	210.216	0.951
hsa-miR-563	-0.692	-0.608	-0.200	1.270	0.950
hsa-miR-372	9.692	800.942	3656.003	5095.033	0.950
hsa-miR-485-3p	-0.070	6.847	18.264	23.104	0.948
hsa-miR-557	143.026	217.438	316.190	352.648	0.939
hsa-miR-127	271.091	464.608	523.205	608.662	0.937
hsa-miR-302c	-0.846	4.152	25.859	33.990	0.933
hsa-miR-640	0.260	1.891	1.270	3.658	0.931
hsa-miR-182	57.298	128.014	405.099	499.940	0.930
hsa-miR-136	199.559	395.412	509.543	577.025	0.930
hsa-miR-146a	135.379	123.890	178.256	245.404	0.928
hsa-miR-596	-1.698	-1.912	-0.760	0.402	0.927
hsa-miR-558	-4.990	-3.496	-2.437	-2.039	0.918
hsa-miR-154*	29.363	35.374	111.681	136.649	0.913
hsa-miR-767-5p	-2.824	-1.172	5.671	7.150	0.909
kshv-miR-K12-10b	6.952	9.161	8.068	20.351	0.906
hsa-miR-373*	9.161	20.351	15.062	29.663	0.902
hsa-miR-296	61.833	91.019	95.598	104.979	0.896
hsa-miR-620	-8.329	-7.039	-5.946	-5.806	0.890
hsa-miR-130a	1053.293	969.396	1137.485	1435.430	0.887
kshv-miR-K12-10a	8.686	12.939	9.416	27.663	0.881
hcmv-miR-US5-1	-2.591	-3.044	-2.209	-0.692	0.880
hsa-miR-424	5933.640	8427.245	8427.245	9331.450	0.879
hcmv-miR-US25-2-5p	-1.837	-2.356	-0.846	0.203	0.876
kshv-miR-K12-2	-3.496	-3.456	-3.496	-2.959	0.874
$NC2_00122731$	-11.690	-10.751	-11.690	-3.496	0.873
hsa-miR-626	-5.206	-5.124	-3.124	-2.879	0.868
hsa-miR-638	2865.870	6295.483	3257.248	14484.563	0.867
mr_1	7550.863	14484.563	19931.725	19931.725	0.863
hsa-miR-124a	13.300	10.939	23.104	26.817	0.861
hsa-miR-526b	12.650	9.416	13.708	25.237	0.857
hsa-miR-527	14.297	15.062	25.237	25.859	0.857
hsa-miR-142-5p	207.026	154.034	341.863	418.973	0.856

ebv-miR-BART17-5p	-3.124	-3.131	-3.249	-0.760	0.852
hsa-miR-623	47.396	66.402	50.912	91.019	0.851
hsa-miR-105	-2.535	-3.586	-2.490	2.261	0.848
hsa-miR-561	-4.650	-5.206	-3.230	-2.535	0.846
hsa-miR-572	418.973	665.715	352.648	1742.315	0.844
hsa-miR-487b	139.448	147.285	182.620	180.465	0.838
hsa-miR-663	523.205	446.551	292.624	2865.870	0.833
hur_2	333895.500	333895.500	366664.750	366664.750	0.824
hsa-miR-220	-4.242	-3.785	-4.326	-2.750	0.819
kshv-miR-K12-11	-3.414	-3.613	-3.445	-2.765	0.819
hsa-miR-671	217.438	333.025	161.101	766.003	0.815
hsa-miR-302c*	0.079	3.265	0.974	5.121	0.815
hsa-miR-583	19.728	17.541	4.794	118.175	0.812
hsa-miR-379	136.649	135.379	180.465	178.256	0.800
hsa-miR-518c*	42.898	54.427	34.432	102.981	0.799
hsa-miR-150	57.820	49.306	59.127	74.951	0.798
hsa-miR-411	83.289	74.951	84.826	98.748	0.790
hsa-miR-337	0.402	0.658	7.294	6.487	0.785
hsa-miR-214	1556.003	1515.303	1435.430	2103.935	0.784
hsa-miR-188	147.285	225.410	145.656	303.449	0.781
hsa-miR-381	91.019	182.620	391.509	333.025	0.775
hsa-miR-627	0.290	0.290	-0.496	3.587	0.774
hsa-miR-765	178.256	281.364	171.602	377.390	0.763
hsa-miR-134	73.330	113.026	102.981	115.042	0.761
hsa-miR-554	-6.900	-6.500	-4.761	-5.206	0.761
hcmv-miR-UL70-3p	118.175	358.706	123.890	525.686	0.756
hsv1-miR-H1	59.127	94.440	41.755	163.513	0.755
hsa-miR-493-5p	80.694	82.090	120.367	113.026	0.753
hsa-miR-155	77.504	56.121	98.748	111.681	0.750
hsa-miR-323	6.167	12.466	23.598	19.217	0.736
hsa-miR-92b	11.672	10.685	8.209	21.333	0.728
hsa-miR-758	21.333	30.984	49.306	41.755	0.724
hsa-miR-29a	3902.293	5300.118	5933.640	5569.658	0.722
hsa-miR-492	-1.857	-1.990	-2.238	-0.846	0.722
ebv-miR-BART7	5.879	5.879	3.514	12.410	0.720
hsa-miR-569	-3.551	-4.067	-3.551	-2.914	0.719
hsa-miR-636	7.018	5.302	5.520	11.166	0.719
hsa-miR-431	4.794	6.952	7.813	7.294	0.718
ebv-miR-BART10	0.203	-2.635	0.930	3.514	0.716
hur_4	19931.725	36489.400	36489.400	36489.400	0.714
hsa-miR-192	67.190	33.990	32.960	154.620	0.712
hsa-miR-566	10.939	10.485	4.152	28.151	0.709
hsa-miR-493-3p	22.188	38.394	40.871	38.982	0.707
hsa-miR-329	18.750	11.166	20.615	27.068	0.704
hsa-let-7c	3656.003	3520.310	3902.293	3902.293	0.703
hsa-miR-518f*	27.663	16.741	35.551	40.871	0.703
hsa-miR-487a	18.264	17.857	47.914	40.070	0.701
hsa-miR-215	49.306	22.570	24.360	102.130	0.680
hsa-miR-562	-2.846	-1.959	0.096	-0.876	0.680
dmr_3	-1.826	-2.298	-0.961	-1.032	0.677
hsa-miR-369-3p	25.859	11.241	22.188	41.496	0.672
hsa-miR-199a	2212.583	1916.780	1742.315	3093.128	0.667
hsa-miR-299-3p	39.199	80.694	91.019	80.694	0.662
hsa-miR-655	-2.070	-3.859	-2.535	-0.366	0.650
hsa-miR-18a*	6.385	25.578	24.798	23.598	0.649
hsa-miR-198	104.550	207.026	38.982	320.477	0.645
hsa-miR-33	311.810	275.693	320.477	341.863	0.644
hsa-miR-142-3p	989.400	525.686	1017.915	1359.818	0.636
hcmv-miR-UL148D	-0.264	-2.504	1.666	1.955	0.636
hsa-miR-770-5p	11.166	8.686	10.485	13.300	0.629
ebv-miR-BART3-5p	0.930	0.003	1.555	1.746	0.625
hsa-miR-382	43.480	59.127	73.330	62.595	0.620
hcmv-miR-US25-1	0.525	-1.260	-0.876	2.732	0.620
hsa-miR-622	33.150	50.912	20.351	69.633	0.620
hsa-miR-410	32.960	69.633	57.298	63.556	0.619
hsa-miR-409-3p	74.951	122.117	143.026	122.117	0.613
hsa-miR-520d*	1.270	1.824	6.952	4.574	0.608

hsa-miR-219	3.514	20.087	13.181	16.986	0.596
hsa-miR-29c	2001.235	4354.658	5569.658	4354.658	0.595
hsa-miR-448	-3.445	-2.686	-4.416	-1.460	0.591
hsa-miR-29b	2747.223	3257.248	4354.658	3656.003	0.589
hsa-miR-380-3p	1.018	2.960	12.778	7.467	0.582
hsa-miR-565	542.940	391.509	1668.123	1188.445	0.582
hsa-miR-575	1489.670	3902.293	252.825	5300.118	0.580
hsa-miR-432	40.871	67.190	78.914	65.609	0.573
hsa-miR-616	-3.928	-2.591	-3.063	-2.891	0.571
ebv-miR-BART5	-3.785	-0.760	-1.959	-1.433	0.569
hsa-miR-342	182.620	171.602	225.410	207.026	0.564
hsa-miR-580	-3.689	-2.186	-2.959	-2.504	0.563
hsa-miR-99a	2340.395	2103.935	2164.618	2543.055	0.555
hsa-miR-100	1435.430	1556.003	1638.658	1556.003	0.553
hsa-miR-9	14.731	7.294	14.731	18.750	0.552
NC2_00079215	-15.673	-12.132	-15.673	-12.132	0.549
hsa-miR-194	87.560	53.051	48.591	125.819	0.536
hsa-miR-433	4.348	5.348	3.658	5.879	0.524
$dmr_2$85	-4.520	-4.650	-4.909	-4.142	0.516
hsa-miR-644	-2.607	-4.794	-2.518	-1.640	0.511
hsa-miR-659	32.625	31.490	12.939	53.051	0.504
hsa-miR-184	12.466	1.555	9.692	17.348	0.502
hsa-miR-519e	0.096	0.142	5.302	2.441	0.497
hsa-miR-455	26.817	19.217	16.336	35.374	0.495
hsa-miR-25	483.805	542.940	940.178	688.623	0.494
hsa-miR-524	-2.238	-2.612	0.732	-0.905	0.491
ebv-miR-BART20-5p	-1.654	-3.763	-2.702	-0.608	0.490
kshv-miR-K12-3	129.669	163.513	136.649	156.084	0.488
hsa-miR-518f	-0.797	0.690	10.155	3.997	0.483
ebv-miR-BART12	2.643	14.666	22.570	13.181	0.483
ebv-miR-BART11-5p	-2.093	-2.665	-0.302	-1.286	0.478
hsa-miR-490	2.498	0.402	0.260	4.152	0.475
NC1_00000215	7.578	2.261	12.466	10.485	0.472
hsa-miR-639	-0.200	-1.857	2.261	1.018	0.470
hsa-miR-619	-2.518	-1.826	-1.070	-1.837	0.460
hsa-miR-523	-0.366	-0.200	5.062	1.824	0.456
hsa-miR-556	-3.613	-2.799	-3.301	-3.063	0.449
NC2_00106057	-2.298	-9.231	0.079	0.079	0.444
hsa-miR-601	44.216	108.621	21.333	108.621	0.436
hcmv-miR-US4	60.934	65.609	38.394	82.090	0.431
hsa-miR-148a	903.623	753.600	736.313	1003.560	0.430
hsa-miR-630	235.685	736.313	57.820	727.792	0.428
hsa-miR-197	44.908	57.298	68.175	55.510	0.428
ebv-miR-BART14-5p	-3.528	-2.879	-3.012	-3.124	0.427
ebv-miR-BART15	-1.136	-1.654	-2.591	-0.152	0.425
hsa-miR-498	25.578	33.150	56.121	37.469	0.423
hsa-miR-10b	3093.128	2543.055	2865.870	3257.248	0.414
hsa-miR-564	104.979	115.042	67.190	139.448	0.411

Chapter 10

TIMP-1 in breast cancer

10.1 Prelude

At time at diagnosis 70% of breast cancers are HR-positive and are offered adjuvant endocrine therapy for 5 years after primary surgery (99). Although positive IHC staining of HRs is predictive for endocrine therapy only 50-60% of HR-positive tumors respond to the treatment (125). Thus, a further classification of the HR-positive breast tumors are needed to determine benefit from endocrine therapy. Currently no predictive biomarkers for endocrine resistance are in clinical use.

TIMP-1 is a potential new prognostic and predictive marker in breast cancer. In a clinical study, patients with metastatic breast cancer and high levels of serum TIMP-1 had less benefit from endocrine therapy compared to patients with a lower level of serum TIMP-1 (299). TIMP-1 has several cancer promoting functions such as stimulation of growth and inhibition of apoptosis (300). Yet, the exact mechanisms behind the increased resistance to endocrine treatment in HR-positive cells are still unknown.

In the following manuscript *"TIMP-1 overexpression confers resistance of MCF-7 breast cancer cells to Fulvestrant"*, we investigated the biological role of TIMP-1 in antiestrogen resistance in breast cancer cells. The basis of the study was MCF-7 cell clones with different expression levels of TIMP-1. We investigated the response to antiestrogens by growth assays and also molecular changes by IHC in relation to TIMP-1 expression. Further, we generated gene expression array data from the cell clones to explore the transcriptional changes caused by changing TIMP-1 expression. Genes that correlated in expression with TIMP-1 were identified by linear regression analysis. The PR was identified both from experimental studies and from the gene expression data analysis to inverse correlate with TIMP-1 expression and seems to be an important factor in TIMP-1-mediated antiestrogen resistance.

The study was in collaboration with the Sino-Danish Breast Cancer Research Centre at Faculty of Life Sciences, University of Copenhagen. Some of the cell clones were also RNA-sequenced and I have analyzed the two gene expression datasets. The RNA-seq data analysis is ongoing and not included in this manuscript, but the analysis is described in chapter 5.

10.2 Manuscript

TIMP-1 overexpression confers resistance of MCF-7 breast cancer cells to Fulvestrant

C Bjerre[1]*, L Vinther[1]*, KC Belling[2]*, AS Rasmussen[1], R Yadav[2], J Wang[3], R Gupta[2], U Lademann[1], N Brünner[1] and J Stenvang[1,4]

[1]University of Copenhagen, Sino-Danish Breast Cancer Research Centre at Department of Veterinary Disease Biology, Faculty of Life Sciences (LIFE), Frederiksberg C, Denmark
[2]Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Lyngby, Denmark.
[3]BGI-Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen, China
[4]Corresponding author

*These authors contributed equally to the study.

Abstract

We hypothesized that the Tissue Inhibitor of Metalloproteinases-1 (TIMP-1) is involved in antiestrogen resistance in breast cancer cells. As a biomarker, TIMP-1 possesses both prognostic and predictive information and a high level of TIMP-1 is associated with a poor prognosis and reduced response to chemotherapy or endocrine therapy in breast cancer patients. To investigate the function of TIMP-1 in breast cancer cells we established a panel of 11 MCF-7 subclones with a wide range of cellular TIMP-1 expression. We found that cells with high expression of TIMP-1 are significantly more resistant to the antiestrogens Fulvestrant (ICI 182,780 or Faslodex), whereas TIMP-1 expression did not change the sensitivity to tamoxifen or 4-hydroxytamoxifen (4-OH-TAM). The ability of the estrogen receptor (ER) to bind Fulvestrant, 4-OH-TAM or 17β-estradiol (E2) was unchanged in the TIMP-1 high cells. The signaling status of pAkt, Her-2 and pERK was unchanged in the high TIMP-1 cells and gene expression analysis did not reveal general defects in ER signaling. However, an inverse expression of TIMP-1 and the E2 inducible progesterone receptor (PR) was found and PR inhibition by the antiprogestin, Mifepristone, largely recapitulated the effects of Fulvestrant.

Introduction

Endocrine therapy - either aromatase inhibitors or antiestrogens - plays a key role in the treatment of breast cancer, both in the adjuvant and metastatic

setting. Adjuvant endocrine therapy has been shown to reduce risk of recurrence and increase overall survival for breast cancer patients with estrogen receptor (ER) and/or progesterone receptor (PR) positive disease significantly. Despite of adjuvant therapy, recurrent metastatic disease will eventually occur in approximately one third of the patients treated with tamoxifen for 5 years after primary surgery (116).

It is apparent, that tumor biomarkers that could predict resistance to endocrine therapy in patients with ER and/or PR positive breast cancer would be of great value when making decisions for treatment of the individual patient, but currently no such biomarkers have reached a state where they can be recommended for routine clinical use.

Endocrine therapy aims to eliminate the tumor promoting effects of estrogen. Aromatase inhibitors or gonadal suppression is applied to minimize circulating levels of estrogen and antiestrogens are applied to block estrogen binding to ER. Currently clinically applied antiestrogens include the partial antagonist tamoxifen and the pure antagonist Fulvestrant (ICI 182, 780, Falsodex). During endocrine therapy resistance may develop due to various molecular alterations in the resistant cells with severe clinical implications (301). Some of the most prominent changes appear to be bidirectional crosstalk between the ER and growth factor receptors, for example EGFR/HER-2 and IGFR, leading to activation of the PI3-K/Akt/mTOR and MAPK pathways.

Tissue Inhibitor of Metalloproteinases-1 (TIMP-1) is a naturally occurring protein in the human organism regulating the activity of Matrix Metalloproteinases (MMPs). Besides this regulatory function, TIMP-1 has several MMP-independent functions such as modulation of angiogenesis (302), promotion of cell proliferation and inhibition of apoptosis (300; 302), all of which are highly important in the context of cancer and cancer progression. TIMP-1 possesses both prognostic and predictive information and a high level of TIMP-1 in tumor tissue is associated with a poor prognosis in breast cancer (303) as well as other cancers (304). Furthermore, high levels of TIMP-1 in tumor tissue are associated with a reduced response to anthracycline based chemotherapy in breast cancer patients (305) and reduced benefit from chemotherapy in patients with metastatic breast cancer (300). In the context of endocrine therapy, metastatic ER and/or PR positive breast cancer patients with high serum levels of TIMP-1 had significantly less benefit from tamoxifen and an aromatase inhibitor, which suggest an important role of TIMP-1 in endocrine resistance in breast cancer (299). Furthermore, TIMP-1 can activate the Akt survival pathway (306) and we therefore hypothesized that high levels of TIMP-1 could mediate antiestrogen resistance in human breast cancer cells.

We have established a panel of 11 stable ER-positive human MCF-7 breast cancer cell lines with high and low expression of TIMP-1 and investigated the sensitivity to tamoxifen, 4-hydroxytamoxifen (4-OH-TAM), Fulvestrant and Mifepristone. Furthermore, we have characterized the major survival signaling pathways in these cell lines and analyzed the gene expression in the

cell lines. We report an inverse correlation of TIMP-1 and PR and a TIMP-1 mediated resistance to Fulvestrant.

Materials and Methods

Transfection and generation of single cell clones

A MCF-7 S1 breast cancer cell line (kindly provided by Professor Marja Jäättela, the Danish Cancer Society, Copenhagen, Denmark) (307) was used for the study. Cells were propagated in RPMI 1640 Glutamax medium (Invitrogen, Denmark) supplemented with 10% heat inactivated fetal calf serum (FCS, Invitrogen, Denmark) and grown at 37°C in a humidified atmosphere with 5% CO_2. For the transfected cells, the media was supplemented with 100μg/ml hygromycin (Sigma-Aldrich, Denmark).

Full length human TIMP-1 cDNA cloned into the pGEM vector (kindly provided by Professor Dylan Edwards, University of East Anglia, United Kingdom) was released from the vector using restriction enzyme digestion with BamHI and HindIII and cloned into the mammalian expression vector pcDNA3.1/Hygro (-) (Invitrogen, Denmark).

MCF-7 S1 cells were stably transfected with the pcDNA3.1/Hygro/TIMP-1 vector or the empty control pcDNA3.1/Hygro vector. Transfections were performed using the FuGENE 6 transfection reagent (Roche, Denmark) according to the manufacturer's instructions.

Single cell cloning was performed by serial dilution of the transfected cells and the TIMP-1 level in each of the established cell lines were determined using an in-house enzyme-linked immunosorbent assay (ELISA) (308). For the experiments with tamoxifen and estrogen two clones with a high and low expression of TIMP-1 respectively were used. For the experiments with Fulvestrant and 4-OH-TAM four clones, two with a high and two with a low expression of TIMP-1 respectively were used.

TIMP-1 and total protein measurements

The total level of TIMP-1 was measured in conditioned cell media and cell lysates. Cells were plated in 10cm Petri dishes (106 cells/dish) and after 48 hours, cell medium was isolated and cells were harvested and lysed by incubation with 100μl lysis buffer (25mM Hepes, 1.5mM $MgCl_2$, 1mM EGTA) containing protease inhibitors (Aprotinin, Leupeptin, Pepstatin A and Pefa Block, 1μg/ml) and phosphatase inhibitors (sodiumfluoride and sodiumorthovanadate, 1mM) for 30 minutes. Samples were centrifuged at 20,000g for 5 minutes and stored at -20°C. TIMP-1 was measured using an in-house ELISA (308). All samples were run in duplicate. As internal control, duplicates of a control plasma pool were included on every plate. For normalization, TIMP-1 concentrations in cell lysates and corresponding media were related to total protein concentration in the lysate. Total amount of protein was determined by the BCA Protein Assay (Pierce, Thermo Fisher Scientific Inc., USA) according to the manufacturer s instructions.

Cell proliferation assay

Cell proliferation rates of each cell line were determined using the CyQuant Cell Proliferation Assay Kit (Molecular Probes Inc., USA/Invitrogen A/S, Denmark) according to the manufacturer's instructions.

Evaluation of cell growth

For the experiments with tamoxifen, 4-OH-TAM, Fulvestrant and Mifepristone cells were propagated in media as described earlier and supplemented with 1% penicillin (10.000 Units/ml) and streptomycin 10.000μg/ml (InVitrogen, Denmark) To avoid the estrogenic effects of phenol red and of components in the FCS, RPMI 1640 without phenol red (InVitrogen, Denmark) and charcoal stripped FCS (InVitrogen, Denmark) were used in the experiments with estrogen.

At day 0 cells were plated in 24 well plates, 10,000 cells/well using the propagation media. At day 2 and 5 experimental media consisting of propagation media supplemented with tamoxifen (T5648, Sigma Aldrich, Denmark), 4-OH-TAM (H7904, Sigma Aldrich, Denmark), Fulvestrant (ICI 182, 780) (I4409, Sigma Aldrich, Denmark), Mifepristone (M8046, Sigma Aldrich, Denmark) and estrogen (kindly provided by Ole William Petersen, University of Copenhagen, Copenhagen) were added in concentrations in the range from 1nM to 5μM for 4-OH-TAM, 1pM to 10μM for Fulvestrant, and 0.5μM to 10μ for Mifepristone. At day 7 cell growth was evaluated with a colorimetric assay using crystal violet (309).

The crystal violet assay is based on the cationic dye crystal violet, which binds to DNA and negatively charged proteins. In short, at day 7 the cells were washed with phosphate buffered saline (Invitrogen, Denmark) twice and then incubated with a 25% methanol solution containing 0.5% crystal violet for 10 minutes at room temperature. Unbound dye was removed by rinsing the plates in tap water and the plates were allowed to dry. The absorbed dye was resuspended in a 0.1M sodium citrate dehydrate in a 50% ethanol suspension for 30 minutes at room temperature. Optical density was measured for the individual wells at 570nm and absorbance correlated with cell number.

In all experiments a mean value for each treatment was estimated from triplicate or quadriplicates and expressed as a percentage with the cells being treated with vehicle only set to a 100%.

The results from the cell growth assays were analyzed using the Students T-test and the ANOVA procedure. We compared clones with a low and high expression of TIMP-1 respectively. The results were used for *(1)* a paired analysis which compared experiments run simultaneously and *(2)* a pooled analysis in which all results for each clone were pooled and compared. Results were considered significant when $p<0.05$.

Calculations were performed using SAS version 9.2 (SAS Institute, Cary, NC, USA), illustrations were made in Excel 2007 (Microsoft, Redmond, WA, USA).

Western Blot Analysis

Preparation of cell lysates: Day one, cells were seeded in T75 flasks (TPP, CM lab, Denmark), day two cells were treated with experimental medium and harvested at day four at 60-70% confluence. Cells were lysed in Proteo-JET Mammalian Cell Lysis Reagent (Fermentas, Thermo Fisher Scientific Inc., USA), containing both 1mM phosphatase and 1mM protease inhibitors (protease inhibitor cocktail set I and phosphatase inhibitor cocktail set II, Calbiochem/Merck4Biosciences, Germany), according to manufacturer's instructions. For the analysis of cells treated with endocrine therapy, cells where either treated with 10^{-6}M Fulvestrant, 10^{-8}M 4-OH-TAM or 96% EtOH. For the 17β-estradiol (E2) analysis, cells were plated in their regular media at day one, washed with PBS at day two, where after media was changed to media containing charcoal stripped media (Invitrogen, Denmark) containing either E2 10^{-9} M or 96% EtOH and harvested at day four. Protein concentration in each sample was determined by Pierce BCA Protein Assay (Thermo Fisher Scientific Inc, USA) according to manufacturer's instructions.

Western analysis: 10-50μg protein were resolved by 10% or 4-12% SDS-polyacrylamide gel electrophoresis (Invitrogen, Denmark). Proteins were transferred onto polyvinylidene difluoride (PVDF) membrane (GE Healthcare Life Sciences, Denmark). Blots were blocked using Tween-20 (0.1%) PBS blocking buffer (Invitrogen, Denmark). Primary monoclonal antibodies were used ON at 4°C at the following concentrations: TIMP-1 (VT7) 1:5000, ERα (RM-9101, Neomarkers, LabVision, United Kingdom) 1:5000, PR (RM-9102, Neomarkers, LabVision, United Kingdom) 1:500, β-actin (A5441, Sigma-Aldrich, Denmark) 1:1000.000, Hsp70 (MS-482-PO, Neomarkers, LabVision, United Kingdom) 1:750.000. Secondary rabbit and mouse IgG polyclonal antibodies (DAKO, Denmark) were used at 1:2.000-1:4.000 dilutions for 1 hour at RT. Finally, the blots were washed 3×10 minutes in washing buffer and incubated with ECLadvance and ECLplus (GE Healthcare Life Sciences, Denmark) and detected by using Biospectrum Imaging System, UVP, USA.

Gene expression microarray

Total RNA was purified from the 12 cell lines by the trizol protocol according to the manufacturer s instruction. The quantity of the total RNA was measured by using a Nanodrop ND1000 (Thermo Scientific Inc, USA). RNA quality was determined using Bioanalyzer nano kit (Agilent Technologies). The samples were amplified in one round using the MessageAmp II aRNA Amplification Kit (Applied Biosystems/Ambion). The aRNA was applied to Agilent Whole Human Genome Microarrays 4×44K v2. Hybridization and scanning of the one-color arrays were done as described by the manufacturer (Agilent Technologies).

Gene expression microarray analysis

Analysis of the data was performed in R where the gProcessedSignals were loaded into the limma R/Bioconductor package. The signals were normalized between arrays using the quantile normalization procedure and the probes were collapsed by taking the median. Principle component analysis (PCA) was performed using the *prcomp* function from the Stats package. Linear regression analysis was carried out to identify transcripts with an expression that correlated either negatively or positively with TIMP-1 expression changes in the MCF-7 subclones. This was done by use of the lm function also from the Stats package. A false discovery rate (FDR) was calculated based on ten random linear regression analysis on shuffled samples according to TIMP-1 expression levels. A subset of transcripts of particular interest was chosen based on the following criteria to the linear regression analysis: R-squared>0.8, absolute value of the slope>0.8 and FDR<0.1.

Estrogen-responsive genes

A metaanalysis on genes whose expression have previously been shown to be affected by E2 stimulation in MCF-7 cells were obtained from a published study by Ochsner et al., 2009 (310). The estrogen-responsive genes were divided into early (3-4h) and late (24h) E2-responsive genes. We investigated the expression changes of both early and late E2-responsive genes in the TIMP-1 expression array dataset. Heatmaps were created by the *gplot* function in R.

ERα-regulated genes

ERα-regulated genes were obtained from the Transfac Pro 2011.2 database (http://www.biobase-international.com/). Transfac contains data on transcription factors, their experimentally proven binding sites and regulated genes. We created a heatmap using the *gplot* libraries in R to visualize the gene expression of the ERα target genes with varying TIMP-1 expression.

Results

Following transfection and single cell cloning, 11 single cell clones expressing different levels of TIMP-1 protein were selected including one empty control vector-transfected clone. The parental MCF-7 S1 cell line was also included in the experiments. Table 10.1 shows an outline of the 12 included cell lines and individual TIMP-1 protein levels in culture medium and cell lysate as measured by ELISA. We determined the doubling times of all the included cell lines (Table 10.1) and found comparable doubling times of the 12 cell lines ranging from 1.38 to 2.37 days. Importantly, the TIMP-1 levels did not influence the growth rates. PCA showed that the clones separated according to the TIMP-1 expression levels with high TIMP-1 clones from the low TIMP-1 expression clones (Supplementary figure 10.9).

Table 10.1: **Characteristics of the 12 included cell lines. The naming of the generated cell lines is given and the doubling times and TIMP-1 protein levels in either cell lysate or culture medium is shown.**

Cell line	Transfection vector	Doubling times (days)	Lysate TIMP-1 conc. (ng/mg)	Medium TIMP-1 conc. (ng/mg)
V1	Empty control	1.48	0	6
T1	TIMP-1	1.38	0	89
T2	TIMP-1	1.39	0.4	11
MCF-7 S1	None	1.77	1.5	11
T3	TIMP-1	1.52	54.7	47
T4	TIMP-1	1.97	93.7	76
T5	TIMP-1	1.64	371.3	313
T6	TIMP-1	1.78	581.4	385
T7	TIMP-1	1.55	384.7	445
T8	TIMP-1	1.52	364.0	266
T9	TIMP-1	2.37	428.8	373
T10	TIMP-1	2.35	932.9	607

Since the level of TIMP-1 has been associated to response to antiestrogens in clinical investigation, it was investigated if increased levels of TIMP-1 in MCF-7 cells changed antiestrogen mediated growth inhibition. For this purpose, two representative clones with high (T5 and T6) or low (V1 and T2) levels of TIMP-1 were selected. The empty vector control (V1) was selected as one of the low expressing TIMP-1 clones. Initially, the growth response to E2 was analyzed in the V1 and T5 clones grown in E2 depleted medium (DCC) and revealed similar growth induction upon E2 stimulation indicating that high levels of TIMP-1 in MCF-7 cells does not influence E2 growth response (data not shown). Addition of the selective estrogen receptor modulator (SERM) 4-OH-TAM the prodrug of Tamoxifen - to the MCF-7 cell with low TIMP-1 levels caused dose-dependent growth-inhibition as previously reported (311). The growth inhibitory effect of 4-OH-TAM in the high TIMP-1 cells was not significantly different from the low TIMP-1 cells (Figure 10.1).

Fulvestrant (ICI 182, 780) is another type of antiestrogen belonging to the class of pure antiestrogens, which prevents ER dimerization and accelerates ER degradation (312) Exposure of the low TIMP-1 clones (V1 and T2) to Fulvestrant lead to dose-dependent growth inhibition (IC50 around 100pM) as normally observed in MCF-7 cells (313). However, the high TIMP-1 clones (T5 and T6) were significantly more resistant to the effect of Fulvestrant (IC50 around 1μM) (Figure 10.2). These data shows that high TIMP-1 levels in MCF-7 cells led to reduced sensitivity to Fulvestrant whereas the growth response to E2 and 4-OH-TAM is not influenced by the TIMP-1 levels.

To investigate if the antiestrogens worked by the relevant mechanisms of action we treated cells with vehicle (C), 4-OH-TAM (T) or Fulvestrant (F)

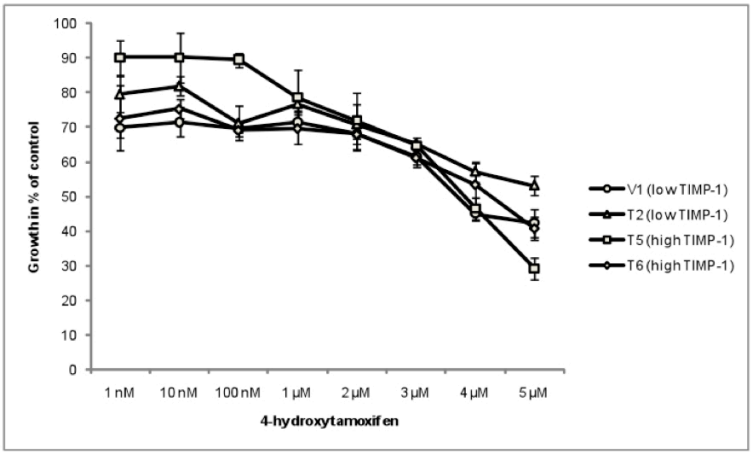


Figure 10.1. Growth response to 4-OH-TAM of high and low TIMP-1 MCF-7 cells. Growth response measured after 5 days of treatment with 4-OH-TAM with concentrations ranging from 1nM to 5μM. The growth response is calculated as a mean of triplicates or quadriplicates +/- SD and expressed as a percentage of the control. Some growth inhibition was observed for all the cell lines, but the inhibitory effect was independent of TIMP-1 status and there were no statistically significant differences in the growth responses. Three independent experiments were performed and the figure is representative of the results.

for 48 hours and harvested cells for western blotting. Firstly, we confirmed the different TIMP-1 expression among the selected cells (V1 and T5) and also found that TIMP-1 is not regulated by either 4-OH-TAM or Fulvestrant (Figure 10.3 upper panel). No difference in the ER levels were observed between high and low TIMP-1 cells and both 4-OH-TAM and Fulvestrant regulated the ER levels by stabilization or degradation, respectively (Figure 10.3 mid panel). This is in accordance with the mode of action for these antiestrogens (312) and suggest that differences in ER levels or changed mode of action of the antiestrogens cannot explain the differences observed in the growth assays. We also analyzed the functionality of the ER by looking at the level of a classical ER regulated protein, namely the progesterone receptor (PR). In the low TIMP-1 cells we observed that both antiestrogens reduced the expression of PR as would be expected upon inhibition of ER. Collectively, these data imply that the ER is normally regulated and fully functional in the low TIMP-1 cells. However, in the high TIMP-1 cells no PR was detected, which suggest an alteration in the ER functionality in these cells (Figure 10.3 lower panel). Equivalent data was obtained from the T2 and T6 cells (data not shown).

The lack of PR in the high TIMP-1 cells may be related to the resistance to Fulvestrant in these cells and therefore the western blot analysis of ER and

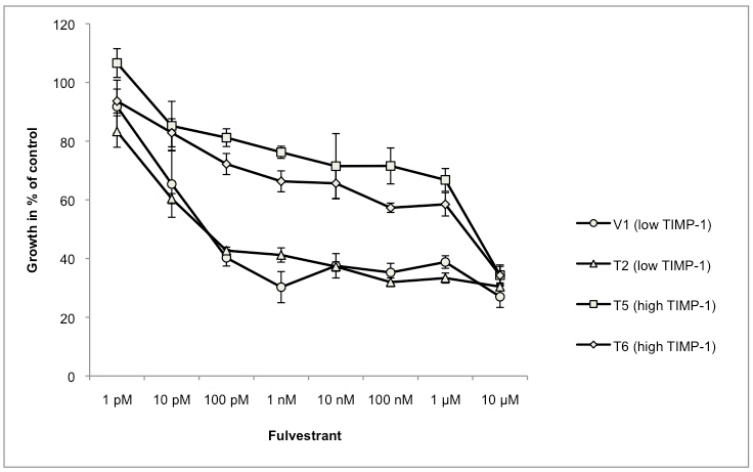


Figure 10.2. Growth response to Fulvestrant of high and low TIMP-1 MCF-7 cells. Growth response measured after 5 days of treatment with Fulvestrant at concentrations from 1pM to 10μM. The growth response is calculated as a mean of triplicates or quadruplicates +/- SD and expressed as a percentage of the control. For the cells with a high expression of TIMP-1 some growth inhibition were observed, whereas the growth inhibition for the clones with a low expression of TIMP-1 were more pronounced. The difference in response to Fulvestrant was observed in a wide concentration range from 10pM to 1μM and the difference was statistically significant ($p<0.001$). Three independent experiments were performed and the figure is representative of the results.

PR was expanded to all the clones described in Table 10.1. These ELISA data was confirmed (Figure 10.4 top panel). No change in the levels of ER in relation to the TIMP-1 levels was found. However, all the clones with low level of TIMP-1 expressed the PR, whereas clones with high levels of TIMP-1 have lost the PR (Figure 10.4 lower panel). Thus, it appears that high levels of TIMP-1 in MCF-7 cells abrogate expression of PR and that this inverse correlation is a general phenomenon.

Due to the TIMP-1 related changes in sensitivity to Fulvestrant and expression of the PR the response to E2 was investigated in the V1, T2, T5 and T6 clones. The cells were cultured in E2 deprived DCC-medium and treated for 48 hours with a concentration (1nM) of E2 that stimulated growth under these conditions. The levels of TIMP-1 were unchanged (Figure 10.5 top panel), which was also verified by ELISA analysis (data not shown). Independently of the TIMP-1 level, addition of E2 downregulated the levels of ER, thereby indicating normal binding of E2 to the ER. PR is a classical E2 responsive protein, which is reduced after addition of antiestrogens (Figure 10.3) and induced upon addition of E2. I the low TIMP-1 clones (V1 and T2) E2 induced PR whereas no induction was observed in the high TIMP-1

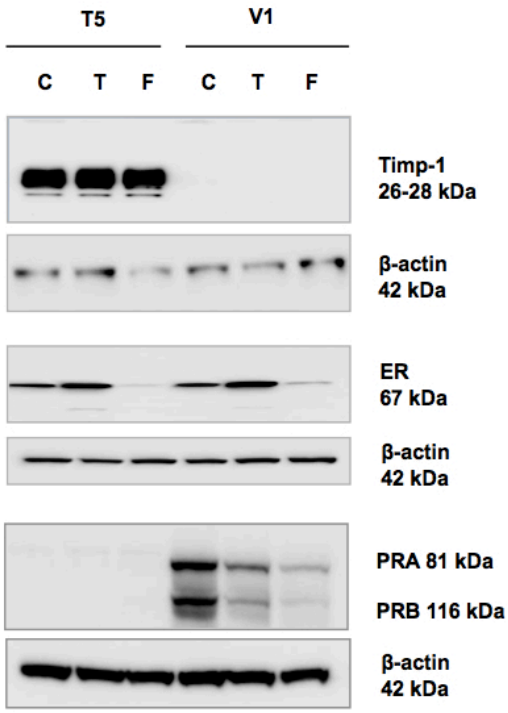


Figure 10.3. Results after 48h treatment with vehicle (C), 4-OH-TAM (T) and Fulvestrant (F). TIMP-1 protein expression levels are unaffected by 4-OH-TAM and Fulvestrant. The clones are as expected ER-positive, which is stabilized upon treatment with 4-OH-TAM. Treatment with Fulvestrant lead to degradation of the ER in both clones. Expression of PR (both isoforms, A and B) is observed only in the low TIMP-1 cells and decreases upon treatment with both 4-OH-TAM and Fulvestrant. The experiment was repeated for the clones T2 (low TIMP-1) and T6 (high TIMP-1) with similar results. All results were confirmed in a second, independent experiment.

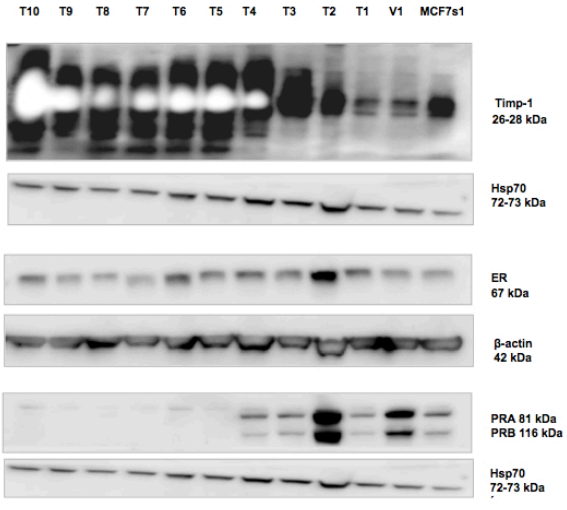


Figure 10.4. Protein concentrations of TIMP-1 in the MCF-7 cell clones and the parental cell line. TIMP-1, ER and PR protein expression levels for the 11 clones and the parental cell line MCF-7 S1. All clones are ER-positive, whereas PR expression is only seen in the low TIMP-1 clones. The results were confirmed in a second, independent experiment.

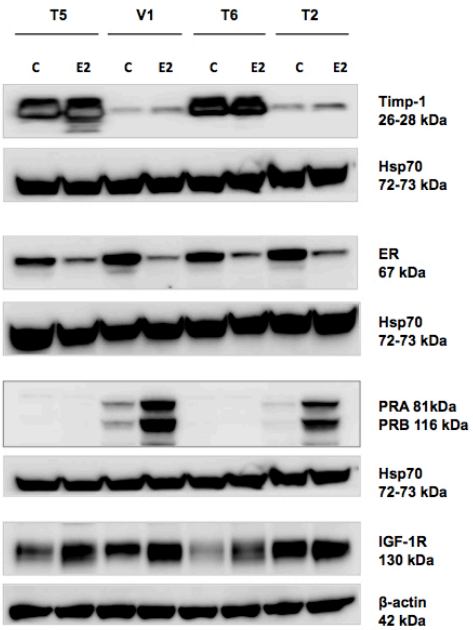


Figure 10.5. Western blot on estrogen response in high and low TIMP-1 MCF-7 cells. Results for cell lines with low TIMP-1 expression (V1, T2) and high TIMP-1 expression (T5, T6) treated with vehicle (C) or 17β-estradiol (E2) for 48h in phenolred free media with charcoal stripped fetal calf serum. As expected ER is downregulated independent of the TIMP-1 expression levels. For PR in the low TIMP-1 clones an induction is seen upon treatment with estrogen, whereas PR in the high TIMP-1 clones is not detected and not inducible with estrogen. IGF-1R is induced by estrogen independently of TIMP-1 expression. The results were confirmed in a second, independent experiment.

clones (T5 and T6) (Figure 10.5). These data and data in Figure 10.3 and 10.4 shows that high levels of TIMP-1 in MCF-7 cells leads to deprivation of PR under normal growth conditions and failure of E2 induction of PR in E2 depleted growth medium. PR is just one example of E2 regulated proteins and to investigate if high TIMP-1 levels leads to de-regulation of another E2 regulated gene the expression levels of the Insulin-like Growth Factor 1 Receptor (IGF-1R) were analyzed. Irrespective of the TIMP-1 levels E2 induced IGF-1R in the MCF-7 cells indicating a functional ER. This suggest that the dysfunctional E2 regulation of PR is a special case among ER regulated protein.

Activation of various signaling pathways have been reported to cause endocrine resistance in breast cancer (313) and we therefore analyzed the expression levels of pAkt, Akt, pERK1/2, ERK1/2 but the expression levels of these proteins were equal among the clones and unrelated to the levels of TIMP-1 (Supplementary figure 10.10 and 10.11). Thus, the observed resistance to Fulvestrant in the cells with high TIMP-1 is not associated to the classical signaling pathways reported in endocrine resistance.

Global gene expression microarray analysis was conducted for the 11 clones and the parental cell line. From preliminary analysis, we saw that the parental cell line MCF-7 S1 and one of the subclones were outliers compared to the other ten cell clones and these were excluded in the further analysis. We compared TIMP-1 ELISA protein measurements of the ten cell clones

with the gene expression measurements obtained with microarrays (Supplementary figure 10.12). The data was used to obtain a detailed understanding of the TIMP-1 mediated molecular alterations. Due to the observed alterations in response to Fulvestrant and the changed expression levels of the E2 inducible PR the focus was on E2-responsive genes. These genes were selected from a metaanalysis of genes that response to E2 in MCF-7 cells (310). This metaanalysis comprise 13,000 genes collected from ten published databases and both early (3-4h) and late (24h) E2-responsive genes that changed with a fold change of 2 in minimum one study was included. By linear regression analyses E2 regulated genes that correlated positively or negatively to the levels of TIMP-1 mRNA were identified to focus on genes that may be involved in the biological effects of increased TIMP-1 levels. The TIMP-1 mRNA and protein levels correlated very well apart from a single clone and the transcript levels were pictured in the heatmaps according to the TIMP-1 transcript levels (Figure 10.6).

This analysis identified a subset of 508 genes that correlated either positively or negatively with TIMP-1 expression and these genes may be related to the protective effect to Fulvestrant. This subset of transcripts were subjected to pathway analysis of TIMP-1 and ER-signaling pathways (Reactome), but this did not lead to any clear suggestions of the pathways that TIMP-1 may be regulating (data not shown). A gene enrichment approach based on linear regression discovered 37 genes that changed in correlation with TIMP-1 expression. Encouragingly, the TIMP-1 mRNA levels were also tightly linked to the levels of the PR being the transcript with the strongest negative association to TIMP-1. These genes may be secondary events of E2 and therefore we also analyzed transcript levels from genes that are directly targeted by ERα (Figure 10.6).

We investigated 45 ERα-targeted genes to see which direct targets were affected on gene expression level by TIMP-1 expression. PR is a direct target of ERα and are highly correlated with TIMP-1. We further identified three more ERα-targeted genes that correlated with TIMP-1: Transforming growth factor beta 3 (TGFB3) correlated negatively with TIMP-1, whereas CKB and RAS11B correlated positively (Figure 10.7). From these gene expression analysis and pathway searches it appeared that TIMP-1 does not generally affect the function of ERα, because we do not see a general change in all E2-responsive genes.

PR appeared to be very tightly associated to the TIMP-1 levels both at the transcript and the protein levels and Fulvestrant has also been reported to act as an antiprogestin (314; 315; 316). Therefore, the growth response to inhibition of PR was investigated by addition of the antiprogestin Mifepristone. The high TIMP-1 cells (T5 and T6) with no detectable PR protein only vaguely responded to Mifepristone with a growth reduction to 60% of untreated cells upon addition of 10μM Mifepristone (Figure 10.8). In contrast, the two low TIMP-1 clones (V1 and T2) displayed much more pronounced dose-dependent inhibition and both were reduced to 20% of untreated controls by addition of 10μM Mifepristone. These data suggest that inhibition

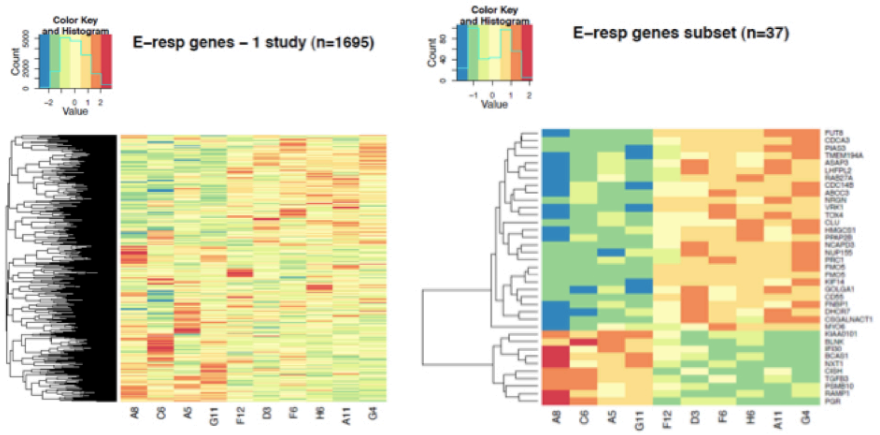


Figure 10.6. Gene expression of E2-reponsive genes in MCF-7 cells from a metaanalysis by Ochsner et al,. 2009 (310). The metaanalysis is a collection of 13,000 genes that change in response to E2 stimulation collected from ten published databases. Both early and late E2-responsive genes that changed with a fold change of 2 in minimum one study is included. Left: Hierarchical clustering of the gene expression of 1,695 E2-responsive genes of the ten MCF-7 cell clones with different expression of TIMP-1. Right: Hierarchical clustering of 37 of the 1,695 genes that changed in correlation with TIMP-1 expression identified with linear regression (R2>0.8, absolute slope>0.8 and FDR<0.1). The cell clones are listed from low TIMP-1 (left) to high TIMP-1 (high).

of PR leads to growth inhibition of high and low TIMP-1 cells in a manner comparable to Fulvestrant.

Discussion

We hypothesized that increased levels of TIMP-1 protein in MCF-7 cells could lead to increased resistance to antiestrogens. We found that elevating TIMP-1 caused a significant resistance in the growth inhibitory response to Fulvestrant, whereas the level of TIMP-1 did not influence the response to tamoxifen or 4-OH-TAM. This may point to a changed functionality of the ER. However, the expression levels of ER were independent of the TIMP-1 levels, the ER was regulated as previously described by both tamoxifen and Fulvestrant and the ER regulated gene IGF-1R was also induced by E2 in both high and low TIMP-1 cells. Together, these data indicates the ER binds both antiestrogens and E2 as expected and alterations in ER is not a likely reason for the observed phenotype. This was confirmed by gene expression analysis.

Interestingly, it has been shown that the endocrine sensitivity of antiestrogen resistant breast cancer cells can be restored by inhibition of the Akt sig-

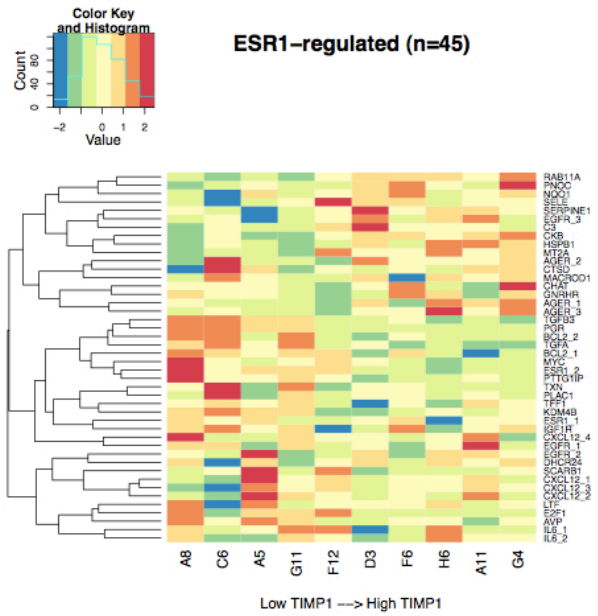


Figure 10.7. Hierarchical clustering of the expression of 45 ERα-regulated transcripts detected from the microarray study. The four genes showing most coordinated expression are PGR (Progesterone Receptor), TGFB3 (Transforming Growth factor, beta 3), CKB (Creatine kinase, brain) and RAB11A (member RAS oncogene family). PGR and TGFB3 are negatively correlated with TIMP1 concentration where as CKB and RAS11B are positively correlated.

nalling pathway (315) or the mammalian target of rapamycin (mTOR) which is downstream of Akt (314). Similar observations have been reported when blocking growth receptor signalling using for example the dual EGFR/HER-2 tyrosine kinase inhibitor Lapatinib (301; 315). Since TIMP-1 activates the Akt survival pathway (306) we investigated these signaling pathways but no changes in relation to the TIMP-1 levels were displayed. A very strong inverse correlation between the TIMP-1 levels and the PR levels was found and in the cells with high levels of TIMP-1 E2 was unable to induce PR expression. In contrast to tamoxifen, Fulvestrant has been reported to act as an antiprogestin (316; 317; 318) besides it antiestrogenic effects. Thus, it is likely that the Fulvestrant resistant phenotype in the high TIMP-1 cells is due to the loss of PR. To elaborate on this hypothesis we analyzed the growth response to the antiprogestin Mifepristone and indeed we found that TIMP-1 over-expressing cells without PR are more resistant to this antiprogestin. These are the very first data to link TIMP-1 to antiestrogens resistance in vitro but in ER and/or PR positive metastatic breast cancer high levels of serum

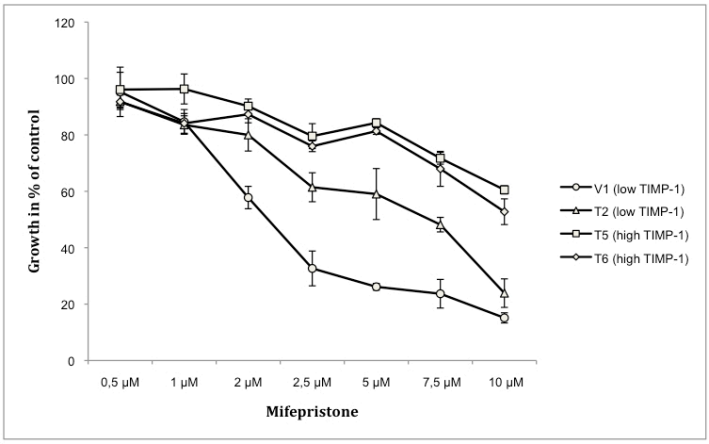


Figure 10.8. Growth effect of Mifepristone mediated PgR inhibition. Growth response measured after 5 days of treatment with Mifepristone at concentrations from 0.5μM to 10μM. The growth response is calculated as a mean of triplicates +/- SD. The TIMP-1 low PR positive clones demonstrate a marked reduction in growth whereas the growth inhibition of TIMP-1 high clones which are PR negative are less pronounced. The difference in response to Mifepristone was observed at concentrations ranging from 0.5μM to 10μM and the overall difference between TIMP-1 low and TIMP-1 high clones was statistically significant ($p < 0.005$) at these concentrations.Three independent experiments were performed and the figure is representative of the results.

TIMP-1 was recently associated to significantly less benefit from tamoxifen and the aromatase inhibitor letrozole. Fulvestrant is currently applied for the treatment of ER-positive metastatic breast cancer, typically after failure of prior endocrine therapy such as aromatase inhibitors and tamoxifen. Recent data (319) however suggest that Fulvestrant is at least as effective as the aromataseinhibitor anastrozole and even increases progression free survival in first line treatment of ER-positive metastatic breast cancer, which might suggest a new role for Fulvestrant in the endocrine therapy hierarchy. To our knowledge, no predictive markers for response to Fulvestrant exist. Our finding in vitro may have clinical implications indicating a potential role for TIMP-1 as a predictive marker in endocrine therapy. This hypothesis needs further investigation and validation in a clinical setting.

Supplementary material

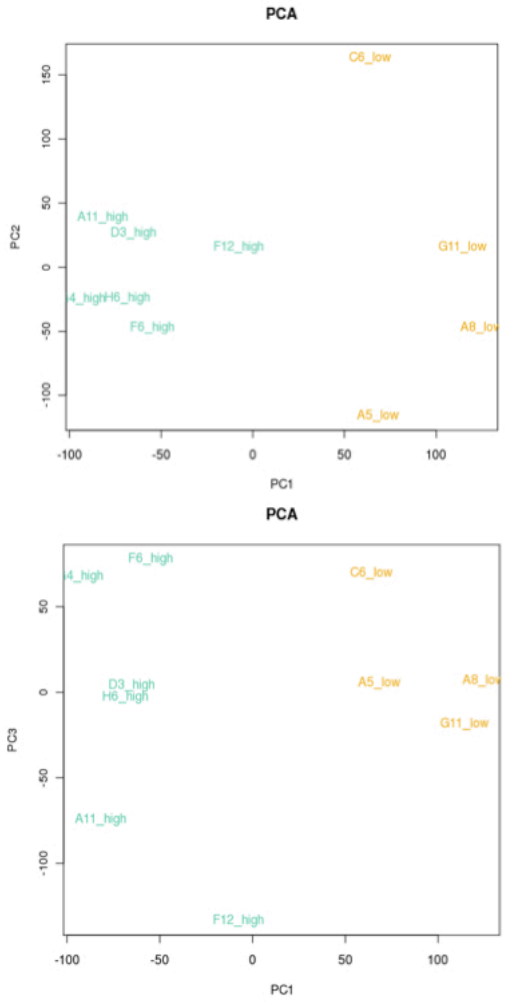


Figure 10.9. PCA plots of the gene expression data from the ten MCF-7 cell clones. The high TIMP-1 cell clones (green) cluster together away from the low TIMP-1 clones (orange). Top: principal component (PC) 1 vs. PC2, bottom: PC1 vs. PC3.

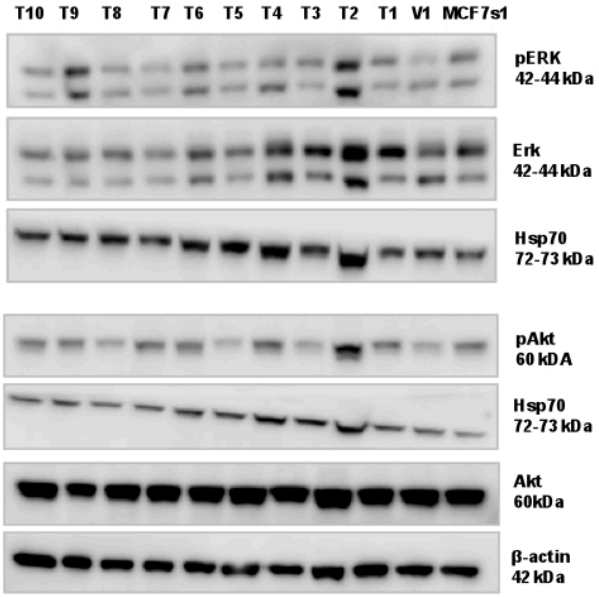


Figure 10.10. Western blot on pAkt, Akt, pERK1/2 and ERK1/2 that have previously been associated with increased endocrine resistance. The expression levels in the cell clones were equal and unrelated to TIMP-1 protein levels.

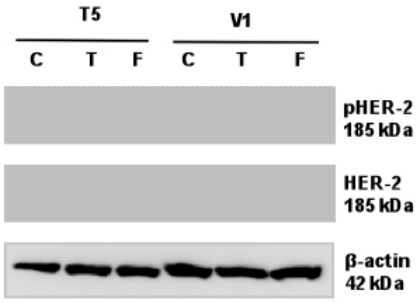


Figure 10.11. Western blot on p-Her2/Her2 that also have been associated with endocrine resistance. No expression was seen in high (T5) and low TIMP-1 (V1) expression cell clones. The expression was not affected by treatment with vehicle (C), 4-OH-TAM (T) and Fulvestrant treatment (F).

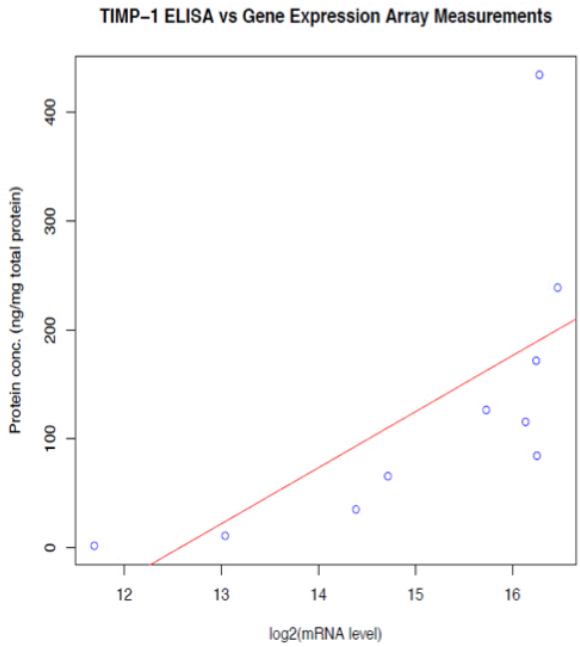


Figure 10.12. TIMP-1 protein concentrations measured by ELISA plotted against gene expression levels of the same ten cells clones. The measurements correlated quite well, except for one of the cell clones, Pearson's correlation coefficient=0.5191.

Part III

Conclusions

Chapter 11

Concluding remarks

This thesis presents projects that have the aim of elucidating the biology of primarily the testis and the breast. The thesis contains five projects that describes research performed in the fields of experimental biology, gene expression profiling and integrative systems biology.

Chapter 6 presents an integrative systems biological study on the regulatory role of OCT4 in ESC pluripotency and differentiation. We identified a highly interconnected PPI network of OCT4 targets that were differentially expressed during EB differentiation. From Gene Ontology analysis we confirmed that the network are involved in embryonal differentiation and system development.

The understanding of stem cells biology and the process of their differentiation is important in order to understand proper development and when development fail. Today stem cell therapy is a fast developing field to treat diseases and heal injuries by introducing new cells into damaged tissue (320). OCT4 may have a future role in inducing pluripotency in cells to reprogram them for use in stem cell therapy (321). Elucidating stem cell biology is also important in order to understand the biology of cancer stem cells that give rise to tumors and drives tumorigenesis (49), an example of such cancer stem cell is the CIS cell, the precursor cell for TGCTs.

TDS is a common syndrome of four disease phenotypes that are established during fetal life because of a disturbance in proper fetal development cause by genetic predisposition or hormonal disturbances. Three projects presented in this thesis studied different aspects of TDS. Chapter 7 describes an experimental study that focused on cell-specific AR signaling in the testis. We used PTM cell-specific AR knock out mice and studied the importance of PTM AR function for proper development and function of the testis. We

found that PTM AR signaling during embryonic development is essential for normal Leydig cell development and function in adult life, thus, for normal testosterone and sperm production in males.

Chapter 8 outline a project on adult spermatogenesis and the cellular and transcriptional changes that are caused by low dose irradiation. By a clustering analysis of gene expression time series data, we identified five clusters with unique gene expression patterns following irradiation that represented five specific testis cell types. We compared the duration of the differentiation stages with the same stages in the first wave of spermatogenesis in pn mice, and found that they have the same timing through spermatogenesis. Also, we found hyperproliferation of the Leydig cells caused by the low dose irradiation. Similar cellular changes likely impact sperm production in testis cancer patients after radiotherapy as they are allocated irradiation in the dose range 16-20 Gy, where we studied the lower dosage of 1 Gy in mice (76).

Chapter 9 presents a project where we identified a miRNA signature in testis CIS cells. We saw that the miRNA profile resembles that of fetal gonocytes. This support that CIS cells are undifferentiated gonocytes that are arrested during embryonal development and lie quiescent in the testis until puberty where CIS cells are triggered to further malignant develop (40). We further saw resemblances of the CIS miRNA profile with the miRNAs expressed in TGCTs, which indicate that these miRNAs are involved in testis carcinogenesis.

The incidence of TDS is increasing with one of the highest incidence rates in Denmark (1; 2). Thus, an increased understanding of the etiology of the syndrome will help to improve the treatment for these patients and potentially also limit the incidence.

HR-positive breast cancer patients account for approximately 75% of breast cancers at diagnosis (105). These patients are offered endocrine therapy if they have a high risk of disease recurrence after primary surgery (99). Unfortunately, only 50-60% of these patients benefit from the treatment (116). Thus, markers to predict resistance to endocrine therapy are important in order to avoid allocating inefficient therapy. One such potential marker is TIMP-1. In the project presented in chapter 10, we studied endocrine resistance caused by TIMP-1 in HR-positive breast cancer cells, and we identified that PR might be an important player in that resistance. This knowledge might contribute to a better treatment for cancer patients with an alternative therapy for the high TIMP-1 HR-positive breast cancer patients, which might be targeted therapy against TIMP-1 or PR.

Taken together, this thesis presents five studies where a combination of experimental work, gene expression profiling and integrative systems biology together elucidate the complex biological of cells both in health and disease. The work has contributed to the understanding of the biology of fetal development, the testis and the breast and will hopefully in the future contribute to a better understanding of the diseases and improve their treatment.

Bibliography

[1] Richiardi, L. *et al.* Testicular cancer incidence in eight northern european countries: secular and recent trends. *Cancer Epidemiol Biomarkers Prev* **13**, 2157–66 (2004). 3, 6, 9, 10, 130

[2] Carlsen, E., Giwercman, A., Keiding, N. & Skakkebaek, N. E. Evidence for decreasing quality of semen during past 50 years. *BMJ* **305**, 609–13 (1992). 3, 8, 130

[3] Kurian, A. W., Fish, K., Shema, S. J. & Clarke, C. A. Lifetime risks of specific breast cancer subtypes among women in four racial/ethnic groups. *Breast Cancer Res* **12**, R99 (2010). 3, 16

[4] Nel-Themaat, L., Gonzalez, G., Akiyama, H. & Behringer, R. R. Illuminating testis morphogenesis in the mouse. *J Androl* **31**, 5–10 (2010). 5, 70

[5] Holstein, A.-F. F., Schulze, W. & Davidoff, M. Understanding spermatogenesis is a prerequisite for treatment. *Reprod Biol Endocrinol* **1**, 107 (2003). 5, 8, 79, 83

[6] Skakkebaek, N. E., Rajpert-De Meyts, E. & Main, K. M. Testicular dysgenesis syndrome: an increasingly common developmental disorder with environmental aspects. *Hum Reprod* **16**, 972–8 (2001). 5, 6

[7] Petersen, P. M., Skakkebaek, N. E. & Giwercman, A. Gonadal function in men with testicular cancer: biological and clinical aspects. *APMIS* **106**, 24–34; discussion 34–6 (1998). 5

[8] Jacobsen, R. *et al.* Risk of testicular cancer in men with abnormal semen characteristics: cohort study. *BMJ* **321**, 789–92 (2000). 5

[9] Winter, C. & Albers, P. Testicular germ cell tumors: pathogenesis, diagnosis and treatment. *Nat Rev Endocrinol* **7**, 43–53 (2011). 5, 89

[10] Chemes, H. *et al.* Early manifestations of testicular dysgenesis in children: pathological phenotypes, karyotype correlations and precursor stages of tumour development. *APMIS* **111**, 12–23; discussion 23–4 (2003). 6

[11] Delbès, G., Levacher, C. & Habert, R. Estrogen effects on fetal and neonatal testicular development. *Reproduction* **132**, 527–38 (2006). 6

[12] Patrão, M. T. C. C., Silva, E. J. R. & Avellar, M. C. W. Androgens and the male reproductive tract: an overview of classical roles and current perspectives. *Arq Bras Endocrinol Metabol* **53**, 934–45 (2009). 6, 7, 8, 49

[13] Andersen, H. R. *et al.* Impaired reproductive development in sons of women occupationally exposed to pesticides during pregnancy. *Environ Health Perspect* **116**, 566–72 (2008). 6

[14] Damgaard, I. N. *et al.* Persistent pesticides in human breast milk and cryptorchidism. *Environ Health Perspect* **114**, 1133–8 (2006). 6

[15] Kristensen, D. M. *et al.* Many putative endocrine disruptors inhibit prostaglandin synthesis. *Environ Health Perspect* **119**, 534–41 (2011). 6

[16] Jensen, T. K. *et al.* Association of in utero exposure to maternal smoking with reduced semen quality and testis size in adulthood: a cross-sectional study of 1,770 young men from the general population in five european countries. *Am J Epidemiol* **159**, 49–58 (2004). 6

[17] Damgaard, I. N. *et al.* Cryptorchidism and maternal alcohol consumption during pregnancy. *Environ Health Perspect* **115**, 272–7 (2007). 6

[18] Ewen, K. A. & Koopman, P. Mouse germ cell development: from specification to sex determination. *Mol Cell Endocrinol* **323**, 76–93 (2010). 7

[19] Kashimada, K. & Koopman, P. Sry: the master switch in mammalian sex determination. *Development* **137**, 3921–30 (2010). 7

[20] Barbaro, M., Wedell, A. & Nordenström, A. Disorders of sex development. *Semin Fetal Neonatal Med* **16**, 119–27 (2011). 7, 49

[21] Wu, X., Wan, S. & Lee, M. M. Key factors in the regulation of fetal and postnatal leydig cell development. *J Cell Physiol* **213**, 429–33 (2007). 7

[22] Dohle, G. R., Smit, M. & Weber, R. F. A. Androgens and male fertility. *World J Urol* **21**, 341–5 (2003). 7, 70

[23] Hannema, S. E. & Hughes, I. A. Regulation of wolffian duct development. *Horm Res* **67**, 142–51 (2007). 7

[24] Grinspon, R. P. & Rey, R. A. Anti-müllerian hormone and sertoli cell function in paediatric male hypogonadism. *Horm Res Paediatr* **73**, 81–92 (2010). 7

[25] Biason-Lauber, A. Control of sex development. *Best Pract Res Clin Endocrinol Metab* **24**, 163–86 (2010). 7

[26] Hiort, O. Androgens and puberty. *Best Pract Res Clin Endocrinol Metab* **16**, 31–41 (2002). 7, 11

[27] Orth, J. M. The role of follicle-stimulating hormone in controlling sertoli cell proliferation in testes of fetal rats. *Endocrinology* **115**, 1248–55 (1984). 8

[28] Phillips, B. T., Gassei, K. & Orwig, K. E. Spermatogonial stem cell regulation and spermatogenesis. *Philos Trans R Soc Lond B Biol Sci* **365**, 1663–78 (2010). 8, 70

[29] de Rooij, D. G. Stem cells in the testis. *Int J Exp Pathol* **79**, 67–80 (1998). 8, 70

[30] Adler, I. D. Comparison of the duration of spermatogenesis between male rodents and humans. *Mutat Res* **352**, 169–72 (1996). 8, 70, 77, 81, 84

[31] seminary, S. J. http://ldysinger.stjohnsem.edu/thm_599d_beg/ 02_biology/02_spermatogenesis.jpg (2011). 9

[32] Mruk, D. D. & Cheng, C. Y. Sertoli-sertoli and sertoli-germ cell interactions and their significance in germ cell movement in the seminiferous epithelium during spermatogenesis. *Endocr Rev* **25**, 747–806 (2004). 8

[33] Setchell, B. P. Blood-testis barrier, junctional and transport proteins and spermatogenesis. *Adv Exp Med Biol* **636**, 212–33 (2008). 8

[34] Bonde, J. P. *et al.* Year of birth and sperm count in 10 danish occupational studies. *Scand J Work Environ Health* **24**, 407–13 (1998). 8

[35] Forti, G. & Krausz, C. Clinical review 100: Evaluation and treatment of the infertile couple. *J Clin Endocrinol Metab* **83**, 4177–88 (1998). 8

[36] Parkin, D. M., Bray, F. I. & Devesa, S. S. Cancer burden in the year 2000. the global picture. *Eur J Cancer* **37 Suppl 8**, S4–66 (2001). 9

[37] M, H., Prener, A. & Skakkebaek, N. E. Testicular cancer, cryptorchidism, inguinal hernia, testicular atrophy, and genital malformations: case-control studies in denmark. *Cancer Causes Control* **7**, 264–74 (1996). 9

[38] Toppari, J. *et al.* Male reproductive health and environmental xenoestrogens. *Environ Health Perspect* **104 Suppl 4**, 741–803 (1996). 9

[39] Dilworth, J. P., Farrow, G. M. & Oesterling, J. E. Non-germ cell tumors of testis. *Urology* **37**, 399–417 (1991). 10, 87

[40] Rajpert-de Meyts, E. & Hoei-Hansen, C. E. From gonocytes to testicular cancer: the role of impaired gonadal development. *Ann N Y Acad Sci* **1120**, 168–80 (2007). 10, 130

[41] Ulbright, T. M. Germ cell neoplasms of the testis. *Am J Surg Pathol* **17**, 1075–91 (1993). 10

[42] Skakkebaek, N. E., Berthelsen, J. G., Giwercman, A. & Müller, J. Carcinoma-in-situ of the testis: possible origin from gonocytes and precursor of all types of germ cell tumours except spermatocytoma. *Int J Androl* **10**, 19–28 (1987). 10, 89

[43] Rajpert-De Meyts, E., J, N., Br-Nielsen, K., Müller, J. & Skakkebaek, N. E. Developmental arrest of germ cells in the pathogenesis of germ cell neoplasia. *APMIS* **106**, 198–204; discussion 204–6 (1998). 10

[44] Dieckmann, K. P. & Skakkebaek, N. E. Carcinoma in situ of the testis: review of biological and clinical features. *Int J Cancer* **83**, 815–22 (1999). 10, 11, 12, 13

[45] J, N. *et al.* Expression of immunohistochemical markers for testicular carcinoma in situ by normal human fetal germ cells. *Lab Invest* **72**, 223–31 (1995). 10

[46] Rajpert-De Meyts, E. & Skakkebaek, N. E. Expression of the c-kit protein product in carcinoma-in-situ and invasive testicular germ cell tumours. *Int J Androl* **17**, 85–92 (1994). 10

[47] Kristensen, D. M. *et al.* Presumed pluripotency markers utf-1 and rex-1 are expressed in human adult testes and germ cell neoplasms. *Hum Reprod* **23**, 775–82 (2008). 10

[48] Rajpert-De Meyts, E. Developmental model for the pathogenesis of testicular carcinoma in situ: genetic and environmental aspects. *Hum Reprod Update* **12**, 303–23 (2006). 10

[49] Wicha, M. S., Liu, S. & Dontu, G. Cancer stem cells: an old idea–a paradigm shift. *Cancer Res* **66**, 1883–90; discussion 1895–6 (2006). 10, 129

[50] Kristensen, D. M. *et al.* Origin of pluripotent germ cell tumours: the role of microenvironment during embryonic development. *Mol Cell Endocrinol* **288**, 111–8 (2008). 11

[51] Albertson, D. G. Gene amplification in cancer. *Trends Genet* **22**, 447–55 (2006). 11

[52] Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000). 11

[53] Maslov, A. Y. & Vijg, J. Genome instability, cancer and aging. *Biochim Biophys Acta* **1790**, 963–9 (2009). 11

[54] MacDonald, D. J. Germline mutations in cancer susceptibility genes: An overview for nurses. *Seminars in Oncology Nursing* **27**, 21 – 33 (2011). URL `http://www.sciencedirect.com/science/article/pii/S0749208110000860`. 11

[55] Garber, J. E. & Offit, K. Hereditary cancer predisposition syndromes. *J Clin Oncol* **23**, 276–92 (2005). 11

[56] Evan, G. I. & Vousden, K. H. Proliferation, cell cycle and apoptosis in cancer. *Nature* **411**, 342–8 (2001). 11

[57] Vogelstein, B. & Kinzler, K. W. Cancer genes and the pathways they control. *Nat Med* **10**, 789–99 (2004). 11

[58] Fearon, E. R. & Dang, C. V. Cancer genetics: tumor suppressor meets oncogene. *Curr Biol* **9**, R62–5 (1999). 11

[59] Albertson, D. G., Collins, C., McCormick, F. & Gray, J. W. Chromosome aberrations in solid tumors. *Nat Genet* **34**, 369–76 (2003). 12

[60] Bartel, D. P. Micrornas: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–97 (2004). 12, 89

[61] Bartel, D. P. Micrornas: target recognition and regulatory functions. *Cell* **136**, 215–33 (2009). 12

[62] , U. A., Nielsen, F. C. & Lund, A. H. Microrna-10a binds the 5'utr of ribosomal protein mrnas and enhances their translation. *Mol Cell* **30**, 460–71 (2008). 12, 89

[63] Eulalio, A. *et al.* Deadenylation is a widespread effect of mirna regulation. *RNA* **15**, 21–32 (2009). 12, 89, 99

[64] Zhang, B., Pan, X., Cobb, G. P. & Anderson, T. A. microrna as oncogenes and tumor suppressors. *Dev Biol* **302**, 1–12 (2007). 12

[65] Brait, M. & Sidransky, D. Cancer epigenetics: above and beyond. *Toxicol Mech Methods* **21**, 275–88 (2011). 12

[66] Bernstein, E. *et al.* Dicer is essential for mouse development. *Nat Genet* **35**, 215–7 (2003). 12

[67] Maatouk, D. M., Loveland, K. L., McManus, M. T., Moore, K. & Harfe, B. D. Dicer1 is required for differentiation of the mouse male germline. *Biol Reprod* **79**, 696–703 (2008). 12

[68] Voorhoeve, P. M. *et al.* A genetic screen implicates mirna-372 and mirna-373 as oncogenes in testicular germ cell tumors. *Cell* **124**, 1169–81 (2006). 12, 89, 93, 98

[69] Gillis, A. J. M. *et al.* High-throughput microrna oe analysis in human germ cell tumours. *J Pathol* **213**, 319–28 (2007). 12, 90, 93, 98, 101

[70] Palmer, R. D. *et al.* Malignant germ cell tumors display common microrna profiles resulting in global changes in expression of messenger rna targets. *Cancer Res* **70**, 2911–23 (2010). 12, 90, 98, 99, 101

[71] Horwich, A., Shipley, J. & Huddart, R. Testicular germ-cell cancer. *Lancet* **367**, 754–65 (2006). 12

[72] Masters, J. R. W. & Köberle, B. Curing metastatic cancer: lessons from testicular germ-cell tumours. *Nat Rev Cancer* **3**, 517–25 (2003). 12, 13

[73] Spermon, J. R. *et al.* Comparison of surveillance and retroperitoneal lymph node dissection in stage i nonseminomatous germ cell tumors. *Urology* **59**, 923–9 (2002). 12

[74] Dieckmann, K. P., Besserer, A. & Loy, V. Low-dose radiation therapy for testicular intraepithelial neoplasia. *J Cancer Res Clin Oncol* **119**, 355–9 (1993). 13

[75] Giwercman, A. *et al.* Localized irradiation of testes with carcinoma in situ: effects on leydig cell function and eradication of malignant germ cells in 20 patients. *J Clin Endocrinol Metab* **73**, 596–603 (1991). 13, 81

[76] Mortensen, M. S., Gundgaard, M. G. & Daugaard, G. Treatment options for carcinoma in situ testis. *Int J Androl* **34**, e32–6 (2011). 13, 71, 81, 130

[77] Looijenga, L. H. & Oosterhuis, J. W. Pathogenesis of testicular germ cell tumours. *Rev Reprod* **4**, 90–100 (1999). 13

[78] Dieckmann, K. P. & Loy, V. Paternity in a patient with testicular seminoma and contralateral testicular intraepithelial neoplasia. *Int J Androl* **16**, 143–6 (1993). 13

[79] Wanderås, E. H., Fosså, S. D. & Tretli, S. Risk of a second germ cell cancer after treatment of a primary germ cell cancer in 2201 norwegian male patients. *Eur J Cancer* **33**, 244–52 (1997). 13

[80] Weissbach, L. Organ preserving surgery of malignant germ cell tumors. *J Urol* **153**, 90–3 (1995). 13

[81] Stordal, B. & Davey, M. Understanding cisplatin resistance using cellular models. *IUBMB Life* **59**, 696–9 (2007). 13

[82] Lampe, H., Horwich, A., Norman, A., Nicholls, J. & Dearnaley, D. P. Fertility after chemotherapy for testicular germ cell cancers. *J Clin Oncol* **15**, 239–45 (1997). 13

[83] Shaw, J. Diagnosis and treatment of testicular cancer. *Am Fam Physician* **77**, 469–74 (2008). 13

[84] Fosså, S. D. & Kravdal, O. Fertility in norwegian testicular cancer patients. *Br J Cancer* **82**, 737–41 (2000). 13

[85] Hovey, R. C., McFadden, T. B. & Akers, R. M. Regulation of mammary gland growth and morphogenesis by the mammary fat pad: a species comparison. *J Mammary Gland Biol Neoplasia* **4**, 53–68 (1999). 15

[86] Gudjonsson, T., Adriance, M. C., Sternlicht, M. D., Petersen, O. W. & Bissell, M. J. Myoepithelial cells: their origin and function in breast morphogenesis and neoplasia. *J Mammary Gland Biol Neoplasia* **10**, 261–72 (2005). 15

[87] Ramsay, D. T., Kent, J. C., Hartmann, R. A. & Hartmann, P. E. Anatomy of the lactating human breast redefined with ultrasound imaging. *J Anat* **206**, 525–34 (2005). 15

[88] Wikimedia. http://upload.wikimedia.org/wikipedia/commons/0/0f/breast_anatomy_normal_scheme.png. 15

[89] Sainsbury, J. R., Anderson, T. J. & Morgan, D. A. Abc of breast diseases: breast cancer. *BMJ* **321**, 745–50 (2000). 16

[90] GLOBESCAN. Globocan 2008 database available from url:http://globocan.iarc.fr/ (2011). 16

[91] Palacios, J., Robles-Frías, M. J., Castilla, M. A., López-García, M. A. & Benítez, J. The molecular pathology of hereditary breast cancer. *Pathobiology* **75**, 85–94 (2008). 16

[92] Song, M., Lee, K.-M. M. & Kang, D. Breast cancer prevention based on gene-environment interaction. *Mol Carcinog* **50**, 280–90 (2011). 16, 17

[93] Russo, J., Moral, R., Balogh, G. A., Mailo, D. & Russo, I. H. The protective role of pregnancy in breast cancer. *Breast Cancer Res* **7**, 131–42 (2005). 17

[94] Li, C. I., Littman, A. J. & White, E. Relationship between age maximum height is attained, age at menarche, and age at first full-term birth and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* **16**, 2144–9 (2007). 17

[95] Rose, D. P. & Vona-Davis, L. Interaction between menopausal status and obesity in affecting breast cancer risk. *Maturitas* **66**, 33 – 38 (2010). URL `http://www.sciencedirect.com/science/article/pii/S0378512210000460`. 17

[96] Fernandez, S. V. & Russo, J. Estrogen and xenoestrogens in breast cancer. *Toxicol Pathol* **38**, 110–22 (2010). 17

[97] cancer.dk. Danish cancer society. url:http://www.cancer.dk/ (2011). 17

[98] Philpotts, L. E. Comprehensive breast imaging 2010. *Semin Roentgenol* **46**, 7–17 (2011). 17

[99] DBCG. Danish breast cancer group. recommendations available afrom url:http://www.dbcg.dk/ (2011). 17, 18, 19, 20, 107, 130

[100] Clarke, C., Sandle, J. & Lakhani, S. R. Myoepithelial cells: pathology, cell separation and markers of myoepithelial differentiation. *J Mammary Gland Biol Neoplasia* **10**, 273–80 (2005). 17

[101] Elston, C. W. & Ellis, I. O. Pathological prognostic factors in breast cancer. i. the value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* **41**, 154–61 (2002). 18

[102] Mainiero, M. B. Regional lymph node staging in breast cancer: the increasing role of imaging and ultrasound-guided axillary lymph node fine needle aspiration. *Radiol Clin North Am* **48**, 989–97 (2010). 18

[103] Knoop, A. S. *et al.* retrospective analysis of topoisomerase iia amplifications and deletions as predictive markers in primary breast cancer patients randomly assigned to cyclophosphamide, methotrexate, and fluorouracil or cyclophosphamide, epirubicin, and fluorouracil: Danish breast cancer cooperative group. *J Clin Oncol* **23**, 7483–90 (2005). 18, 19

[104] Bardou, V.-J. . J., Arpino, G., Elledge, R. M., Osborne, C. K. & Clark, G. M. Progesterone receptor status significantly improves outcome prediction over estrogen receptor status alone for adjuvant endocrine therapy in two large breast cancer databases. *Journal of Clinical Oncology* **21**, 1973 (2003). 18

[105] Borgquist, S. *et al.* Oestrogen receptors alpha and beta show different associations to clinicopathological parameters and their co-expression might predict a better response to endocrine treatment in breast cancer. *J Clin Pathol* **61**, 197–203 (2008). 18, 130

[106] Platet, N., Cathiard, A. M., Gleizes, M. & Garcia, M. Estrogens and their receptors in breast cancer progression: a dual role in cancer proliferation and invasion. *Crit Rev Oncol Hematol* **51**, 55–67 (2004). 18

[107] Gutierrez, C. & Schiff, R. Her2: biology, detection, and clinical implications. *Arch Pathol Lab Med* **135**, 55–62 (2011). 18, 20

[108] Mass, R. D. *et al.* Evaluation of clinical outcomes according to her2 detection by fluorescence in situ hybridization in women with metastatic breast cancer treated with trastuzumab. *Clin Breast Cancer* **6**, 240–6 (2005). 18

[109] Champoux, J. J. Dna topoisomerases: structure, function, and mechanism. *Annu Rev Biochem* **70**, 369–413 (2001). 18

[110] Järvinen, T. A. H. & Liu, E. T. Topoisomerase iialpha gene (top2a) amplification and deletion in cancer–more common than anticipated. *Cytopathology* **14**, 309–13 (2003). 18

[111] Pritchard, K. I. *et al.* Her-2 and topoisomerase ii as predictors of response to chemotherapy. *J Clin Oncol* **26**, 736–44 (2008). 18

[112] Khan, S. A. & Eladoumikdachi, F. Optimal surgical treatment of breast cancer: implications for local control and survival. *J Surg Oncol* **101**, 677–86 (2010). 18

[113] (EBCTCG), E. B. C. T. C. G. Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials. *The Lancet* **366**, 2087 – 2106 (2006). URL `http://www.sciencedirect.com/science/article/pii/S0140673605678877`. 19

[114] Mouridsen, H. T. & Andersen, J. [adjuvant treatment of breast cancer. endocrine therapy]. *Ugeskr Laeger* **169**, 3072–6 (2007). 19

[115] Goldhirsch, A. *et al.* Meeting highlights: updated international expert consensus on the primary therapy of early breast cancer. *J Clin Oncol* **21**, 3357–65 (2003). 19

[116] (EBCTCG), E. B. C. T. C. G. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet* **365**, 1687–717 (2005). 19, 20, 21, 110, 130

[117] Orlando, L. *et al.* Molecularly targeted endocrine therapies for breast cancer. *Cancer Treat Rev* **36 Suppl 3**, S67–71 (2010). 20

[118] Puhalla, S., Brufsky, A. & Davidson, N. Adjuvant endocrine therapy for premenopausal women with breast cancer. *Breast* **18 Suppl 3**, S122–30 (2009). 20

[119] Clynes, R. A., Towers, T. L., Presta, L. G. & Ravetch, J. V. Inhibitory fc receptors modulate in vivo cytoxicity against tumor targets. *Nat Med* **6**, 443–6 (2000). 20

[120] Miyoshi, Y. *et al.* Predictive factors for anthracycline-based chemotherapy for human breast cancer. *Breast Cancer* **17**, 103–9 (2010). 20

[121] Gewirtz, D. A. A critical evaluation of the mechanisms of action proposed for the antitumor effects of the anthracycline antibiotics adriamycin and daunorubicin. *Biochem Pharmacol* **57**, 727–41 (1999). 20

[122] McGrogan, B. T., Gilmartin, B., Carney, D. N. & McCann, A. Taxanes, microtubules and chemoresistant breast cancer. *Biochim Biophys Acta* **1785**, 96–132 (2008). 20

[123] Tagliabue, E., Balsari, A., Campiglio, M. & Pupa, S. M. Her2 as a target for breast cancer therapy. *Expert Opin Biol Ther* **10**, 711–24 (2010). 21

[124] Nikolényi, A. *et al.* Tumour topoisomerase ii alpha protein expression and outcome after adjuvant dose-dense anthracycline-based chemotherapy. *Pathol Oncol Res* (2011). 21

[125] Goldhirsch, A., Colleoni, M. & Gelber, R. D. Endocrine therapy of breast cancer. *Ann Oncol* **13 Suppl 4**, 61–8 (2002). 21, 107

[126] Valabrega, G., Montemurro, F. & Aglietta, M. Trastuzumab: mechanism of action, resistance and future perspectives in her2-overexpressing breast cancer. *Ann Oncol* **18**, 977–84 (2007). 21

[127] Moreno-Aspitia, A. & Perez, E. A. Anthracycline- and/or taxane-resistant breast cancer: results of a literature review to determine the clinical challenges and current treatment trends. *Clin Ther* **31**, 1619–40 (2009). 21

[128] Arpino, G., Wiechmann, L., Osborne, C. K. & Schiff, R. Crosstalk between the estrogen receptor and the her tyrosine kinase receptor family: molecular mechanism and clinical implications for endocrine therapy resistance. *Endocr Rev* **29**, 217–33 (2008). 21

[129] Rivera, E. & Gomez, H. Chemotherapy resistance in metastatic breast cancer: the evolving role of ixabepilone. *Breast Cancer Res* **12 Suppl 2**, S2 (2010). 21

[130] Vidal, M. A unifying view of 21st century systems biology. *FEBS Lett* **583**, 3891–4 (2009). 23

[131] Ahn, A. C., Tewari, M., Poon, C.-S. S. & Phillips, R. S. The limits of reductionism in medicine: could systems biology offer an alternative? *PLoS Med* **3**, e208 (2006). 23

[132] Sauer, U., Heinemann, M. & Zamboni, N. Science (2007). 23

[133] Snoep, J. L., Bruggeman, F., Olivier, B. G. & Westerhoff, H. V. Towards building the silicon cell: a modular approach. *Biosystems* **83**, 207–16 (2006). 23, 24

[134] Ehrenberg, M., Elf, J. & Hohmann, S. Systems biology: Nobel symposium 146. (2009). 23

[135] Kitano, H. Computational systems biology. *Nature* **420**, 206–10 (2002). 24

[136] Schneider, M. V. & Orchard, S. Omics technologies, data and bioinformatics principles. *Methods Mol Biol* **719**, 3–30 (2011). 24, 35

[137] Kiechle, F. L., Zhang, X. & Holland-Staley, C. A. The -omics era and its impact. *Arch Pathol Lab Med* **128**, 1337–45 (2004). 24

[138] Xie, Z., Hu, S., Qian, J., Blackshaw, S. & Zhu, H. Systematic characterization of protein-dna interactions. *Cell Mol Life Sci* **68**, 1657–68 (2011). 25, 26, 43

[139] Finchtalk. Finchtalk http://finchtalk.geospiza.com/2008/08/chip-ing-away-at-analysis.html (2011). 25

[140] MacQuarrie, K. L., Fong, A. P., Morse, R. H. & Tapscott, S. J. Genome-wide transcription factor binding: beyond direct target regulation. *Trends Genet* **27**, 141–8 (2011). 25

[141] MacQuarrie, K. L., Fong, A. P., Morse, R. H. & Tapscott, S. J. Genome-wide transcription factor binding: beyond direct target regulation. *Trends in Genetics* **27**, 141 – 148 (2011). URL `http://www.sciencedirect.com/science/article/pii/S0168952511000023`. 25

[142] Dowell, R. D. Transcription factor binding variation in the evolution of gene regulation. *Trends Genet* **26**, 468–75 (2010). 26, 43

[143] Pattin, K. A. & Moore, J. H. Role for protein-protein interaction databases in human genetics. *Expert Rev Proteomics* **6**, 647–59 (2009). 26

[144] Völkel, P., Le Faou, P. & Angrand, P.-O. O. Interaction proteomics: characterization of protein complexes using tandem affinity purification-mass spectrometry. *Biochem Soc Trans* **38**, 883–7 (2010). 26

[145] Prulj, N. s. Protein-protein interactions: making sense of networks via graph-theoretic modeling. *Bioessays* **33**, 115–23 (2011). 26

[146] Fields, S. & Song, O. A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245–6 (1989). 26

[147] Williamson, M. P. & Sutcliffe, M. J. Protein-protein interactions. *Biochem Soc Trans* **38**, 875–8 (2010). 26, 27

[148] De Las Rivas, J. & Fontanillo, C. Protein-protein interactions essentials: key concepts to building and analyzing interactome networks. *PLoS Comput Biol* **6**, e1000807 (2010). 27

[149] Bader, G. D. *et al.* Bind–the biomolecular interaction network database. *Nucleic Acids Res* **29**, 242–5 (2001). 27

[150] Xenarios, I. *et al.* Dip: the database of interacting proteins. *Nucleic Acids Res* **28**, 289–91 (2000). 27

[151] Breitkreutz, B.-J. J. *et al.* The biogrid interaction database: 2008 update. *Nucleic Acids Res* **36**, D637–40 (2008). 27

[152] Prasad, T. S. K., Kandasamy, K. & Pandey, A. Human protein reference database and human proteinpedia as discovery tools for systems biology. *Methods Mol Biol* **577**, 67–79 (2009). 27

[153] Hermjakob, H. *et al.* Intact: an open source molecular interaction database. *Nucleic Acids Res* **32**, D452–5 (2004). 27

[154] Kanehisa, M. *et al.* From genomics to chemical genomics: new developments in kegg. *Nucleic Acids Res* **34**, D354–7 (2006). 27

[155] Zanzoni, A. *et al.* Mint: a molecular interaction database. *FEBS Lett* **513**, 135–40 (2002). 27

[156] Güldener, U. *et al.* Mpact: the mips protein interaction resource on yeast. *Nucleic Acids Res* **34**, D436–41 (2006). 27

[157] Joshi-Tope, G. *et al.* Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res* **33**, D428–32 (2005). 27

[158] von Mering, C. *et al.* Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* **417**, 399–403 (2002). 27

[159] Remm, M., Storm, C. E. & Sonnhammer, E. L. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J Mol Biol* **314**, 1041–52 (2001). 27

[160] Lage, K. *et al.* A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat Biotechnol* **25**, 309–16 (2007). 28

[161] de Lichtenberg, U., Jensen, L. J., Brunak, S. & Bork, P. Dynamic complex formation during the yeast cell cycle. *Science* **307**, 724–7 (2005). 28

[162] Nichols, J. *et al.* Formation of pluripotent stem cells in the mammalian embryo depends on the pou transcription factor oct4. *Cell* **95**, 379–91 (1998). 28, 47

[163] Doetschman, T. C., Eistetter, H., Katz, M., Schmidt, W. & Kemler, R. The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. *J Embryol Exp Morphol* **87**, 27–45 (1985). 28, 48

[164] Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145 (1998). 28, 47

[165] Bratt-Leal, A. M., Carpenedo, R. L. & McDevitt, T. C. Engineering the embryoid body microenvironment to direct embryonic stem cell differentiation. *Biotechnol Prog* **25**, 43–51 (2009). 28, 48

[166] Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–56 (2005). 28

[167] Campbell, P. A., Perez-Iratxeta, C., Andrade-Navarro, M. A. & Rudnicki, M. A. Oct4 targets regulatory nodes to modulate stem cell function. *PLoS One* **2**, e553 (2007). 28

[168] Kim, J., Chu, J., Shen, X., Wang, J. & Orkin, S. H. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132**, 1049–61 (2008). 28

[169] Loh, Y.-H. H. *et al.* The oct4 and nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38**, 431–40 (2006). 28

[170] Mathur, D. *et al.* Analysis of the mouse embryonic stem cell regulatory networks obtained by chip-chip and chip-pet. *Genome Biol* **9**, R126 (2008). 28

[171] Matoba, R. *et al.* Dissecting oct3/4-regulated gene networks in embryonic stem cells by expression profiling. *PLoS One* **1**, e26 (2006). 28

[172] Babaie, Y. *et al.* Analysis of oct4-dependent transcriptional networks regulating self-renewal and pluripotency in human embryonic stem cells. *Stem Cells* **25**, 500–10 (2007). 28

[173] Ivanova, N. *et al.* Dissecting self-renewal in stem cells with rna interference. *Nature* **442**, 533–8 (2006). 28

[174] Sharov, A. A. *et al.* Identification of pou5f1, sox2, and nanog downstream target genes with statistical confidence by applying a novel algorithm to time course microarray and genome-wide chromatin immunoprecipitation data. *BMC Genomics* **9**, 269 (2008). 28

[175] Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nat Genet* **25**, 25–9 (2000). 29

[176] Baulcombe, D. Rna silencing. *Trends Biochem Sci* **30**, 290–3 (2005). 30

[177] Bramsen, J. B. & Kjems, J. Chemical modification of small interfering rna. *Methods Mol Biol* **721**, 77–103 (2011). 30

[178] Graveley, B. R. Molecular biology: power sequencing. *Nature* **453**, 1197–8 (2008). 33, 36, 39, 43

[179] Hardiman, G. Microarray platforms–comparisons and contrasts. *Pharmacogenomics* **5**, 487–502 (2004). 33, 34, 43

[180] Ozsolak, F. & Milos, P. M. Rna sequencing: advances, challenges and opportunities. *Nat Rev Genet* **12**, 87–98 (2011). 33

[181] JORDAN, B. Historical background and anticipated developments. *Annals of the New York Academy of Sciences* **975**, 24–32 (2002). URL `http://dx.doi.org/10.1111/j.1749-6632.2002.tb05938.x`. 33, 34

[182] Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. Quantitative monitoring of gene expression patterns with a complementary dna microarray. *Science* **270**, 467–70 (1995). 33

[183] Stoughton, R. B. Applications of dna microarrays in biology. *Annu Rev Biochem* **74**, 53–82 (2005). 34

[184] Dufva, M. Fabrication of dna microarray. *Methods Mol Biol* **529**, 63–79 (2009). 34

[185] Tanaka, A. *et al.* All-in-one tube method for quantitative gene expression analysis in oligo-dt(30) immobilized pcr tube coated with mpc polymer. *Anal Sci* **25**, 109–14 (2009). 34

[186] Bilban, M., Buehler, L. K., Head, S., Desoye, G. & Quaranta, V. Normalizing dna microarray data. *Curr Issues Mol Biol* **4**, 57–64 (2002). 35

[187] Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**, R80 (2004). 35

[188] Steinhoff, C. & Vingron, M. Normalization and quantification of differential expression in gene expression microarrays. *Brief Bioinform* **7**, 166–77 (2006). 35

[189] Bar-Joseph, Z. Analyzing time series gene expression data. *Bioinformatics* **20**, 2493–503 (2004). 35

[190] Bland, J. M. & Altman, D. G. Multiple significance tests: the bonferroni method. *BMJ* **310**, 170 (1995). 35

[191] Tsai, C.-A. A., Hsueh, H.-m. M. & Chen, J. J. Estimation of false discovery rates in multiple testing: application to gene microarray data. *Biometrics* **59**, 1071–81 (2003). 35

[192] Maxam, A. M. & Gilbert, W. A new method for sequencing dna. 1977. *Biotechnology* **24**, 99–103 (1992). 36

[193] Smith, L. M. *et al.* Fluorescence detection in automated dna sequence analysis. *Nature* **321**, 674–9 (1986). 36

[194] Schadt, E. E., Turner, S. & Kasarskis, A. A window into third-generation sequencing. *Hum Mol Genet* **19**, R227–40 (2010). 36, 43

[195] Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001). 36

[196] Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–51 (2001). 36

[197] Zhou, X. *et al.* The next-generation sequencing technology and application. *Protein Cell* **1**, 520–36 (2010). 37

[198] Wang, Z., Gerstein, M. & Snyder, M. Rna-seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57–63 (2009). 37, 38, 39, 41

[199] Illumina. Illumina protocol on mrna preparation for sequencing. 37

[200] Katz, Y., Wang, E. T., Airoldi, E. M. & Burge, C. B. Analysis and design of rna sequencing experiments for identifying isoform regulation. *Nat Methods* **7**, 1009–15 (2010). 37

[201] Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. ii. error probabilities. *Genome Res* **8**, 186–94 (1998). 38, 39

[202] Meyerson, M., Gabriel, S. & Getz, G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet* **11**, 685–96 (2010). 38

[203] Pareek, C. S., Smoczynski, R. & Tretyn, A. Sequencing technologies and genome sequencing. *J Appl Genet* (2011). 38, 43

[204] Kong, Y. Btrim: A fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. *Genomics* **98**, 152–3 (2011). 39

[205] Garber, M., Grabherr, M. G., Guttman, M. & Trapnell, C. Computational methods for transcriptome annotation and quantification using rna-seq. *Nat Methods* **8**, 469–77 (2011). 39, 42

[206] Trapnell, C., Pachter, L. & Salzberg, S. L. Tophat: discovering splice junctions with rna-seq. *Bioinformatics* **25**, 1105–11 (2009). 39

[207] Li, H., Ruan, J. & Durbin, R. Mapping short dna sequencing reads and calling variants using mapping quality scores. *Genome Res* **18**, 1851–8 (2008). 39

[208] Lassmann, T., Hayashizaki, Y. & Daub, C. O. Samstat: monitoring biases in next generation sequencing data. *Bioinformatics* **27**, 130–1 (2011). 39

[209] Wold, B. & Myers, R. M. Sequence census methods for functional genomics. *Nat Methods* **5**, 19–21 (2008). 41

[210] Edgren, H. *et al.* Identification of fusion genes in breast cancer by paired-end rna-sequencing. *Genome Biol* **12**, R6 (2011). 41

[211] Maher, C. A. *et al.* Transcriptome sequencing to detect gene fusions in cancer. *Nature* **458**, 97–101 (2009). 41

[212] Sultan, M. *et al.* A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **321**, 956–60 (2008). 41

[213] Griffith, M. *et al.* Alternative expression analysis by rna sequencing. *Nat Methods* **7**, 843–7 (2010). 41

[214] Trapnell, C. *et al.* Transcript assembly and quantification by rna-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511–5 (2010). 42, 43

[215] Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. & Gilad, Y. Rna-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* **18**, 1509–17 (2008). 42

[216] Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by rna-seq. *Nat Methods* **5**, 621–8 (2008). 43

[217] Rozowsky, J. *et al.* Peakseq enables systematic scoring of chip-seq experiments relative to controls. *Nat Biotechnol* **27**, 66–75 (2009). 43

[218] Vickaryous, M. K. & Hall, B. K. Human cell type diversity, evolution, development, and classification with special reference to cells derived from the neural crest. *Biol Rev Camb Philos Soc* **81**, 425–55 (2006). 47

[219] Shah, F. J. *et al.* Gene expression profiles of mouse spermatogenesis during recovery from irradiation. *Reprod Biol Endocrinol* **7**, 130 (2009). 67, 71, 73, 77, 78, 81, 82

[220] Cheng, C. Y. & Mruk, D. D. The biology of spermatogenesis: the past, present and future. *Philos Trans R Soc Lond B Biol Sci* **365**, 1459–63 (2010). 70

[221] Hermo, L., Pelletier, R.-M. M., Cyr, D. G. & Smith, C. E. Surfing the wave, cycle, life history, and genes/proteins expressed by testicular germ cells. part 1: background to spermatogenesis, spermatogonia, and spermatocytes. *Microsc Res Tech* **73**, 241–78 (2010). 70, 83

[222] Bellvé, A. R. *et al.* Spermatogenic cells of the prepuberal mouse. isolation and morphological characterization. *J Cell Biol* **74**, 68–85 (1977). 70, 84, 85

[223] Bellvé, A. R., Millette, C. F., Bhatnagar, Y. M. & O'Brien, D. A. Dissociation of the mouse testis and characterization of isolated spermatogenic cells. *J Histochem Cytochem* **25**, 480–94 (1977). 70, 84, 85

[224] Shinohara, T., Orwig, K. E., Avarbock, M. R. & Brinster, R. L. Remodeling of the postnatal mouse testis is accompanied by dramatic changes in stem cell number and niche accessibility. *Proc Natl Acad Sci U S A* **98**, 6186–91 (2001). 70

[225] OAKBERG, E. F. Duration of spermatogenesis in the mouse and timing of stages of the cycle of the seminiferous epithelium. *Am J Anat* **99**, 507–16 (1956). 70

[226] Delic, J. I., Hendry, J. H., Morris, I. D. & Shalet, S. M. Serum androgen binding protein and follicle stimulating hormone as indices of sertoli cell function in the irradiated testis. *Br J Cancer Suppl* **7**, 105–7 (1986). 71

[227] Jahnukainen, K., Ehmcke, J., Hou, M. & Schlatt, S. Testicular function and fertility preservation in male cancer patients. *Best Pract Res Clin Endocrinol Metab* **25**, 287–302 (2011). 71, 81

[228] Zhang, Z., Shao, S., Shetty, G. & Meistrich, M. L. Donor sertoli cells transplanted into irradiated rat testes stimulate partial recovery of endogenous spermatogenesis. *Reproduction* **137**, 497–508 (2009). 71

[229] Oakberg, E. F. A new concept of spermatogonial stem-cell renewal in the mouse and its relationship to genetic effects. *Mutat Res* **11**, 1–7 (1971). 71, 81

[230] van der Meer, Y., Huiskamp, R., Davids, J. A., van der Tweel, I. & de Rooij, D. G. The sensitivity of quiescent and proliferating mouse spermatogonial stem cells to x irradiation. *Radiat Res* **130**, 289–95 (1992). 71, 81

[231] van der Meer, Y., Huiskamp, R., Davids, J. A., van der Tweel, I. & de Rooij, D. G. The sensitivity to x rays of mouse spermatogonia that are committed to differentiate and of differentiating spermatogonia. *Radiat Res* **130**, 296–302 (1992). 71, 81

[232] OAKBERG, E. F. Gamma-ray sensitivity of spermatogonia of the mouse. *J Exp Zool* **134**, 343–56 (1957). 71, 81

[233] OAKBERG, E. F. Initial depletion and subsequent recovery of spermatogonia of the mouse after 20 r of gamma rays and 100, 300, and 600 r of x-rays. *Radiat Res* **11**, 700–19 (1959). 71, 81

[234] Ivell, R. & Spiess, A. N. *Analysing differential gene expression in the testis.* (Springer-Verlag, Berlin, 2002). 71

[235] Almstrup, K. *et al.* Analysis of cell-type-specific gene expression during mouse spermatogenesis. *Biol Reprod* **70**, 1751–61 (2004). 73, 77, 84, 85

[236] Shima, J. E., McLean, D. J., McCarrey, J. R. & Griswold, M. D. The murine testicular transcriptome: characterizing gene expression in the testis during the progression of spermatogenesis. *Biol Reprod* **71**, 319–30 (2004). 73, 77, 84, 85

[237] Acharya, K. K. *et al.* A novel tissue-specific meta-analysis approach for gene expression predictions, initiated with a mammalian gene expression testis database. *BMC Genomics* **11**, 467 (2010). 73, 77

[238] Dennis, G. *et al.* David: Database for annotation, visualization, and integrated discovery. *Genome Biol* **4**, P3 (2003). 73

[239] Colpi, G. M., Contalbi, G. F., Nerva, F., Sagone, P. & Piediferro, G. Testicular function following chemo-radiotherapy. *Eur J Obstet Gynecol Reprod Biol* **113 Suppl 1**, S2–6 (2004). 81

[240] Petersen, P. M. *et al.* Effect of graded testicular doses of radiotherapy in patients treated for carcinoma-in-situ in the testis. *J Clin Oncol* **20**, 1537–43 (2002). 81

[241] Shapiro, E. *et al.* Effects of fractionated irradiation of endocrine aspects of testicular function. *J Clin Oncol* **3**, 1232–9 (1985). 81

[242] Holm, M., Rajpert-De Meyts, E., Andersson, A.-M. M. & Skakkebaek, N. E. Leydig cell micronodules are a common finding in testicular biopsies from men with impaired spermatogenesis and are associated with decreased testosterone/lh ratio. *J Pathol* **199**, 378–86 (2003). 81

[243] Petersen, P. M., Daugaard, G., R, M. & Skakkebaek, N. E. Endocrine function in patients treated for carcinoma in situ in the testis with irradiation. *APMIS* **111**, 93–8; discussion 98–9 (2003). 81

[244] Rich, K. A., Kerr, J. B. & de Kretser, D. M. Evidence for leydig cell dysfunction in rats with seminiferous tubule damage. *Mol Cell Endocrinol* **13**, 123–35 (1979). 81

[245] Bang, A. K. *et al.* Testosterone production is better preserved after 16 than 20 gray irradiation treatment against testicular carcinoma in situ cells. *Int J Radiat Oncol Biol Phys* **75**, 672–6 (2009). 81

[246] Sarkar, P. S., Paul, S., Han, J. & Reddy, S. Six5 is required for spermatogenic cell survival and spermiogenesis. *Hum Mol Genet* **13**, 1421–31 (2004). 81, 93

[247] Guitton, N., Brouazin-Jousseaume, V., Dupaix, A., Jégou, B. & Chenal, C. Radiation effect on rat sertoli cell function in vitro and in vivo. *Int J Radiat Biol* **75**, 327–33 (1999). 81

[248] Kochar, N. K. & Bateman, A. J. Post-irradiation changes in sertoli cells. *J Reprod Fertil* **18**, 265–73 (1969). 81

[249] Schlecht, U. *et al.* Expression profiling of mammalian male meiosis and gametogenesis identifies novel candidate genes for roles in the regulation of fertility. *Mol Biol Cell* **15**, 1031–43 (2004). 82

[250] Sassone-Corsi, P. Unique chromatin remodeling and transcriptional regulation in spermatogenesis. *Science* **296**, 2176–8 (2002). 82

[251] Schmidt, E. E. & Schibler, U. High accumulation of components of the rna polymerase ii transcription machinery in rodent spermatids. *Development* **121**, 2373–83 (1995). 82

[252] Miller, D., Brinkworth, M. & Iles, D. Paternal dna packaging in spermatozoa: more than the sum of its parts? dna, histones, protamines and epigenetics. *Reproduction* **139**, 287–301 (2010). 82

[253] Godmann, M. *et al.* Dynamic regulation of histone h3 methylation at lysine 4 in mammalian spermatogenesis. *Biol Reprod* **77**, 754–64 (2007). 83

[254] Hermo, L., Pelletier, R.-M. M., Cyr, D. G. & Smith, C. E. Surfing the wave, cycle, life history, and genes/proteins expressed by testicular germ cells. part 2: changes in spermatid organelles associated with development of spermatozoa. *Microsc Res Tech* **73**, 279–319 (2010). 83

[255] Steger, K. Transcriptional and translational regulation of gene expression in haploid spermatids. *Anat Embryol (Berl)* **199**, 471–87 (1999). 83

[256] Baba, T., Azuma, S., Kashiwabara, S. & Toyoda, Y. Sperm from mice carrying a targeted mutation of the acrosin gene can penetrate the oocyte zona pellucida and effect fertilization. *J Biol Chem* **269**, 31845–9 (1994). 83

[257] Hurtado de Catalfo, G. E., de Alaniz, M. J. T. & Marra, C. A. Influence of commercial dietary oils on lipid composition and testosterone production in interstitial cells isolated from rat testis. *Lipids* **44**, 345–57 (2009). 84

[258] Rommerts, F. F. & Brinkman, A. O. Modulation of steroidogenic activities in testis leydig cells. *Mol Cell Endocrinol* **21**, 15–28 (1981). 84

[259] Vergouwen, R. P. *et al.* Postnatal development of testicular cell populations in mice. *J Reprod Fertil* **99**, 479–85 (1993). 84

[260] Ellis, P. J. I. *et al.* Modulation of the mouse testis transcriptome during postnatal development and in selected models of male infertility. *Mol Hum Reprod* **10**, 271–81 (2004). 84, 85

[261] Schmoll, H. J. *et al.* European consensus on diagnosis and treatment of germ cell cancer: a report of the european germ cell cancer consensus group (egcccg). *Ann Oncol* **15**, 1377–99 (2004). 89

[262] Feldman, D. R., Bosl, G. J., Sheinfeld, J. & Motzer, R. J. Medical treatment of advanced testicular cancer. *JAMA* **299**, 672–84 (2008). 89

[263] Looijenga, L. H. J., Gillis, A. J. M., Stoop, H. J., Hersmus, R. & Oosterhuis, J. W. Chromosomes and expression in human testicular germ-cell tumors: insight into their cell of origin and pathogenesis. *Ann N Y Acad Sci* **1120**, 187–214 (2007). 89

[264] Almstrup, K. *et al.* Embryonic stem cell-like features of testicular carcinoma in situ revealed by genome-wide gene expression profiling. *Cancer Res* **64**, 4736–43 (2004). 89, 94, 101

[265] Holstein, A. F. & Körner, F. Light and electron microscopical analysis of cell types in human seminoma. *Virchows Arch A Pathol Anat Histol* **363**, 97–112 (1974). 89

[266] Sonne, S. B. *et al.* Analysis of gene expression profiles of microdissected cell populations indicates that testicular carcinoma in situ is an arrested gonocyte. *Cancer Res* **69**, 5241–50 (2009). 89, 90, 92, 94, 100, 101

[267] Lai, E. C. Micro rnas are complementary to 3' utr sequence motifs that mediate negative post-transcriptional regulation. *Nat Genet* **30**, 363–4 (2002). 89

[268] Xie, X. *et al.* Systematic discovery of regulatory motifs in human promoters and 3' utrs by comparison of several mammals. *Nature* **434**, 338–45 (2005). 89

[269] Cho, W. C. S. Oncomirs: the discovery and progress of micrornas in cancers. *Mol Cancer* **6**, 60 (2007). 89

[270] Esquela-Kerscher, A. & Slack, F. J. Oncomirs - micrornas with a role in cancer. *Nat Rev Cancer* **6**, 259–69 (2006). 89

[271] He, L. *et al.* A microrna polycistron as a potential human oncogene. *Nature* **435**, 828–33 (2005). 89

[272] O'Donnell, K. A., Wentzel, E. A., Zeller, K. I., Dang, C. V. & Mendell, J. T. c-myc-regulated micrornas modulate e2f1 expression. *Nature* **435**, 839–43 (2005). 89

[273] Novotny, G. W. *et al.* Translational repression of e2f1 mrna in carcinoma in situ and normal testis correlates with expression of the mir-17-92 cluster. *Cell Death Differ* **14**, 879–82 (2007). 89, 101

[274] Giwercman, A., Cantell, L. & Marks, A. Placental-like alkaline phosphatase as a marker of carcinoma-in-situ of the testis. comparison with monoclonal antibodies m2a and 43-9f. *APMIS* **99**, 586–94 (1991). 90

[275] Saeed, A. I. *et al.* Tm4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34**, 374–8 (2003). 92

[276] Nielsen, J. E. *et al.* Germ cell differentiation-dependent and stage-specific expression of lancl1 in rodent testis. *Eur J Histochem* **47**, 215–22 (2003). 93

[277] Novotny, G. W. *et al.* Analysis of gene expression in normal and neoplastic human testis: new roles of rna. *Int J Androl* **30**, 316–26; discussion 326–7 (2007). 93, 102

[278] Wilson, K. D. *et al.* Microrna profiling of human-induced pluripotent stem cells. *Stem Cells Dev* **18**, 749–58 (2009). 93

[279] Yamada, Y. *et al.* Mir-96 and mir-183 detection in urine serve as potential tumor markers of urothelial carcinoma: correlation with stage and grade, and comparison with urinary cytology. *Cancer Sci* **102**, 522–9 (2011). 93

[280] Sonne, S. B. *et al.* Optimizing staining protocols for laser microdissection of specific cell types from the testis including carcinoma in situ. *PLoS One* **4**, e5536 (2009). 94

[281] Landgraf, P. *et al.* A mammalian microrna expression atlas based on small rna library sequencing. *Cell* **129**, 1401–14 (2007). 95, 96, 103

[282] Ro, S., Park, C., Sanders, K. M., McCarrey, J. R. & Yan, W. Cloning and expression profiling of testis-expressed micrornas. *Dev Biol* **311**, 592–602 (2007). 95

[283] Calabrese, J. M., Seila, A. C., Yeo, G. W. & Sharp, P. A. Rna sequence analysis defines dicer's role in mouse embryonic stem cells. *Proc Natl Acad Sci U S A* **104**, 18097–102 (2007). 95

[284] Mineno, J. *et al.* The expression profile of micrornas in mouse embryos. *Nucleic Acids Res* **34**, 1765–71 (2006). 95

[285] Thomson, J. M., Parker, J., Perou, C. M. & Hammond, S. M. A custom microarray platform for analysis of microrna gene expression. *Nat Methods* **1**, 47–53 (2004). 95

[286] Bentwich, I. *et al.* Identification of hundreds of conserved and nonconserved human micrornas. *Nat Genet* **37**, 766–70 (2005). 95

[287] Martinez, N. J. & Gregory, R. I. Microrna gene regulatory pathways in the establishment and maintenance of esc identity. *Cell Stem Cell* **7**, 31–5 (2010). 98, 101

[288] Bagga, S. *et al.* Regulation by let-7 and lin-4 mirnas results in target mrna degradation. *Cell* **122**, 553–63 (2005). 99

[289] Behm-Ansmant, I. *et al.* mrna degradation by mirnas and gw182 requires both ccr4:not deadenylase and dcp1:dcp2 decapping complexes. *Genes Dev* **20**, 1885–98 (2006). 99

[290] Tzur, G. *et al.* Comprehensive gene and microrna expression profiling reveals a role for micrornas in human liver development. *PLoS One* **4**, e7511 (2009). 99

[291] Zibert, J. R. *et al.* Micrornas and potential target interactions in psoriasis. *J Dermatol Sci* **58**, 177–85 (2010). 99

[292] Murray, M. J. *et al.* The two most common histological subtypes of malignant germ cell tumour are distinguished by global microrna profiles, associated with differential transcription factor expression. *Mol Cancer* **9**, 290 (2010). 101

[293] Combes, A. N. *et al.* Gonadal defects in cited2-mutant mice indicate a role for sf1 in both testis and ovary differentiation. *Int J Dev Biol* **54**, 683–9 (2010). 101

[294] Olive, V., Jiang, I. & He, L. mir-17-92, a cluster of mirnas in the midst of the cancer network. *Int J Biochem Cell Biol* **42**, 1348–54 (2010). 101

[295] LeBron, C. *et al.* Genome-wide analysis of genetic alterations in testicular primary seminoma using high resolution single nucleotide polymorphism arrays. *Genomics* **97**, 341–9 (2011). 102

[296] Hoei-Hansen, C. E. *et al.* Towards a non-invasive method for early detection of testicular neoplasia in semen samples by identification of fetal germ cell-specific markers. *Hum Reprod* **22**, 167–73 (2007). 102

[297] van Casteren, N. J. *et al.* Noninvasive detection of testicular carcinoma in situ in semen using oct3/4. *Eur Urol* **54**, 153–8 (2008). 102

[298] Murray, M. J. *et al.* Identification of micrornas from the mir-371 373 and mir-302 clusters as potential serum biomarkers of malignant germ cell tumors. *Am J Clin Pathol* **135**, 119–25 (2011). 102

[299] Lipton, A. *et al.* Serum timp-1 and response to the aromatase inhibitor letrozole versus tamoxifen in metastatic breast cancer. *J Clin Oncol* **26**, 2653–8 (2008). 107, 110

[300] Würtz, S. O., Schrohl, A.-S. S., Mouridsen, H. & Brünner, N. Timp-1 as a tumor marker in breast cancer–an update. *Acta Oncol* **47**, 580–90 (2008). 107, 110

[301] Johnston, S. R. D. New strategies in estrogen receptor-positive breast cancer. *Clin Cancer Res* **16**, 1979–87 (2010). 110, 122

[302] Würtz, S. O. *et al.* Tissue inhibitor of metalloproteinases-1 in breast cancer. *Endocr Relat Cancer* **12**, 215–27 (2005). 110

[303] Schrohl, A.-S. S. *et al.* Tumor tissue levels of tissue inhibitor of metalloproteinase-1 as a prognostic marker in primary breast cancer. *Clin Cancer Res* **10**, 2289–98 (2004). 110

[304] Jiang, Y., Goldberg, I. D. & Shi, Y. E. Complex roles of tissue inhibitors of metalloproteinases in cancer. *Oncogene* **21**, 2245–52 (2002). 110

[305] Willemoe, G. L. *et al.* Lack of timp-1 tumour cell immunoreactivity predicts effect of adjuvant anthracycline-based chemotherapy in patients (n=647) with primary breast cancer. a danish breast cancer cooperative group study. *Eur J Cancer* **45**, 2528–36 (2009). 110

[306] Liu, X.-W. W., Bernardo, M. M., Fridman, R. & Kim, H.-R. C. R. Tissue inhibitor of metalloproteinase-1 protects human breast epithelial cells against intrinsic apoptotic cell death via the focal adhesion kinase/phosphatidylinositol 3-kinase and mapk signaling pathway. *J Biol Chem* **278**, 40364–72 (2003). 110, 122

[307] Jäättelä, M., Benedict, M., Tewari, M., Shayman, J. A. & Dixit, V. M. Bcl-x and bcl-2 inhibit tnf and fas-induced apoptosis and activation of phospholipase a2 in breast carcinoma cells. *Oncogene* **10**, 2297–305 (1995). 111

[308] Holten-Andersen, M. N. *et al.* Quantitation of timp-1 in plasma of healthy blood donors and patients with advanced cancer. *Br J Cancer* **80**, 495–503 (1999). 111

[309] Lundholt, B. K., Briand, P. & Lykkesfeldt, A. E. Growth inhibition and growth stimulation by estradiol of estrogen receptor transfected human breast epithelial cell lines involve different pathways. *Breast Cancer Res Treat* **67**, 199–214 (2001). 112

[310] Ochsner, S. A. *et al.* Gems (gene expression metasignatures), a web resource for querying meta-analysis of expression microarray datasets: 17beta-estradiol in mcf-7 cells. *Cancer Res* **69**, 23–6 (2009). 114, 120, 121

[311] Wakeling, A. E. & Bowler, J. Ici 182,780, a new antioestrogen with clinical potential. *J Steroid Biochem Mol Biol* **43**, 173–7 (1992). 115

[312] Dowsett, M., Nicholson, R. I. & Pietras, R. J. Biological characteristics of the pure antiestrogen fulvestrant: overcoming endocrine resistance. *Breast Cancer Res Treat* **93 Suppl 1**, S11–8 (2005). 115, 116

[313] Frankel, L. B., Lykkesfeldt, A. E., Hansen, J. B. & Stenvang, J. Protein kinase c alpha is a marker for antiestrogen resistance and is involved in the growth of tamoxifen resistant human breast cancer cells. *Breast Cancer Res Treat* **104**, 165–79 (2007). 115, 119

[314] deGraffenried, L. A. *et al.* Inhibition of mtor activity restores tamoxifen response in breast cancer cells with aberrant akt activity. *Clin Cancer Res* **10**, 8059–67 (2004). 120, 122

[315] Frogne, T. *et al.* Antiestrogen-resistant human breast cancer cells require activated protein kinase b/akt for growth. *Endocr Relat Cancer* **12**, 599–614 (2005). 120, 122

[316] Nawaz, Z., Stancel, G. M. & Hyder, S. M. The pure antiestrogen ici 182,780 inhibits progestin-induced transcription. *Cancer Res* **59**, 372–6 (1999). 120, 122

[317] Rosenberg Zand, R. S., Grass, L., Magklara, A., Jenkins, D. J. & Diamandis, E. P. Is ici 182,780 an antiprogestin in addition to being an antiestrogen? *Breast Cancer Res Treat* **60**, 1–8 (2000). 122

[318] Hyder, S. M. & Stancel, G. M. Inhibition of progesterone-induced vegf production in human breast cancer cells by the pure antiestrogen ici 182,780. *Cancer Lett* **181**, 47–53 (2002). 122

[319] Robertson, J. F. R. *et al.* Activity of fulvestrant 500 mg versus anastrozole 1 mg as first-line treatment for advanced breast cancer: results from the first study. *J Clin Oncol* **27**, 4530–5 (2009). 123

[320] Cai, W., Zhang, Y. & Kamp, T. J. Imaging of induced pluripotent stem cells: From cellular reprogramming to transplantation. *Am J Nucl Med Mol Imaging* **1**, 18–28 (2011). 129

[321] Patel, M. & Yang, S. Advances in reprogramming somatic cells to induced pluripotent stem cells. *Stem Cell Rev* **6**, 367–80 (2010). 129