EPITHELIAL SPECIFIC TRANSCRIPTOME MAP OF THE HUMAN PROSTATE

By

Andrew Jonathan Symes

A dissertation submitted to University College London in conformity with the requirements for the degree of Doctor of Medicine (Research) MD (Res)

Prostate Cancer Research Centre
University College London
August 2010

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institution of learning.

I, Andrew Jonathan Symes, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

This work was undertaken by myself at the Prostate Cancer Research Centre between February 2004 and March 2006

Acknowledgements

First of all, I would like to thank my two supervisors, Dr Aamir Ahmed and Professor John Masters for supervising this thesis with such great interest. I thank them for their insight, guidance and inspiration in research, and for their consistent support and encouragement throughout all stages of my MD study.

I am indebted to the St Peter's Trust and the Covent Garden Cancer Research Trust (now the Prostate Cancer Research Centre) without whose generous grants and funding this project would not have been possible.

I am grateful to Dr Qin Wang for her help in introducing me to many of the molecular and cellular techniques that have formed the basis of my work. My special thanks to Nipurna Jina and Dr Mike Hubank, at the Gene microarray centre, Institute of Child Health, University College London, for all their help, expertise and advice for microarray studies. To Calum Thomson, University of Dundee, a warm thank you for all your skills with microscopy and sectioning. And to Joseph Nariculam, who constructed the prostate tissue array used in this study, my sincerest thanks.

I am grateful to Sharon Cole and Sapna Halder for their professional help in making administrative matters run smoothly. I also would like to thank my fellow colleagues, Isabelle Bisson, Tharani Nitkunan, Jo Nariculam, Corrina Kane, Charlotte Foley, Jason Constantinou and Magali Williamson, for making PCRC a pleasant place to work.

I would like to thank my wife Anya, who has stood by me throughout my MD, and has given me so much love and support to complete this work, as well as being a wonderful mother to our daughter Chloe. And to my mum Janet Symes, who continues to look after my family and me and who is an inspiration to us all.

I dedicate this dissertation to my father David Symes, a truly great man, who is sadly no longer with us. You are missed every day.

Abstract

The prostate has a zonal anatomy, with differing susceptibilities to disease (benign prostatic hyperplasia originates from the transition zone, prostate cancer largely arises in the peripheral zone). The molecular reasons for this are not understood. Previous prostate cancer microarray studies have used whole benign, diseased or tissue adjacent to the carcinoma as normal controls, for what is an epithelial disease. This study provides a gene expression profile of normal, non-diseased prostate, or a 'reference prostate gene expression profile'. This has been compared to prostate cancer to identify novel biomarkers of disease. This study also investigates zonal differences in gene expression between different anatomical zones of the prostate. I used normal, human donor prostate tissue, laser capture microdissection (LCM), and Affymetrix gene expression arrays to achieve these aims. Eight LCM prostate epithelial samples from 3 donor prostates were used. The gene expression data was validated by low density real-time PCR and immunohistochemistry on a prostate tissue microarray. Major differences in gene expression were discovered between whole tissue and LCM epithelium only prostate using homology tables. Novel prostate adenocarcinoma genes were identified using a publicly available LCM prostate cancer gene expression array dataset. 9318 genes showed significant differential expression in normal vs. cancer datasets. Three targets, MCM2, NR1D1 and ABCA1 were validated at the protein level. Expression of NR1D1 and ABCA1 were increased in cancer, suggesting they are novel epithelial biomarkers of prostate cancer.

An analysis of zonal differences in gene expression found significant differences between zones. Zonal specific markers included TGM4 (central zone), LPL (peripheral zone), and COL9A1 (transition zone).

This study provides: (i) a gene expression profile of the normal prostate epithelium (ii) novel, prostate adenocarcinoma specific gene and protein markers and (iii) the first gene expression profile of normal epithelium on the basis of zonal anatomy of the prostate.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

The prostate is a male reproductive organ, which has two functions: secretory and muscular. It produces about 30% of the volume of seminal fluid, providing nutrients for sperm. Contained within its secretions are proteases such as Prostate Specific Antigen (PSA) (Bostwick, 1994), which maintain semen fluidity mainly by acting as anticoagulants. The prostate contains smooth muscle surrounding the glands and these are used to force ejection of the prostatic fluid to mix into the seminal fluid during ejaculation. At a cellular level the prostate consists of a complex ductal system of epithelial cells embedded in a stromal matrix. The epithelial cells are responsible for exocrine and neuroendocrine functions with the stromal compartment consisting of smooth muscle cells and fibroblasts, whose secretion of growth factors is important in normal and pathological conditions of the prostate.

1.1 Anatomy of the Human Prostate

The name prostate was originally derived from the Greek word ‘prohistani’, meaning ‘to stand in front of’, and has been attributed to Herophilus of Alexandria who used the term in 335 B.C. to describe the organ located in front of the urinary bladder (Dobson, 1925). John Hunter was the first to describe the effects of a pathologically enlarged prostate in 1786 (Hunter, 1786). He described the prostatic middle lobe as ‘a valve obstructing the passage of urine’. Unlike many other organs in the body however, whose basic anatomies and functions have long been known, the anatomy and function of the prostate was not resolved until recently, and was the source of disagreement throughout the course of the 20th century.

1.1.1 The Prostate: Lowsley’s description

In 1912 Albert Lowsley published a detailed description of the anatomy of the human prostate, which would be widely adopted as the anatomy of the prostate for 40 years (Lowsley, 1912). By using serial sections and wax reconstructions of embryonic and foetal prostate glands, Lowsley observed the developing ductal system at all ages up to the time of birth. The prostate ducts were noted to bud from the urethra in five separate clusters, making possible the designation of discrete and separate lobes (Figure 1). The lobes were defined according to the site of origin of their buds form

the urethra and their subsequent direction of growth. The two lateral lobes were the largest, with their ducts arising both sides of the urethra and fanning out laterally to the capsule, thus making up the majority of the glandular tissue in contact with the prostate surface. The posterior lobe formed a midline strip lying between the lateral lobes and its buds arose near the prostatic apex, growing proximally to the bladder neck. The middle lobe buds branched directly proximally from the urethra at the veromontanum. They coursed between the ejaculatory ducts and the proximal urethra to the base of the prostate, where they had their greatest extent. The anterior lobe buds developed from the other duct origins and grew into the stroma anterior to the urethra. Lowsley noted that in all but a few cases the anterior lobe atrophied and disappeared by the time of birth. It is seldom referred to in descriptions of the adult prostate.

The histological appearance of the acinar and duct epithelium was not discussed by Lowsley, since the acinar system is not fully developed at birth, and epithelium remains simplified and non-descript until puberty. The type of epithelium in different lobes was assumed to be the same, and the prostate was thought to be histologically homogeneous.

Lowsley's own data dealt solely with foetal glands however he made anecdotal comments that carcinoma arose exclusively in the narrow strip of posterior tissue which coincided with the posterior lobe. Later authors confirmed the conclusion that carcinoma arose only in the posterior lobe, but they often defined the boundaries of that lobe differently (Huggins and Webster, 1948). Other investigators subsequently disagreed with Lowsley and challenged his findings (Franks, 1953). The point was made that no discrete lobe boundaries could be seen in the adult prostate and that carcinoma originated over an area much larger than the posterior lobe.

In 1939 Le Duc (Le Duc, 1939) injected opaque material into the duct system of several prostates and then cleared the tissue. He found that all of the ducts in the gland extended mainly laterally form the urethra. There was no duct system extending proximally in the posterior midline, as required by Lowsley's definition. It was concluded that the posterior lobe did not exist, at least in the adult. This work received little recognition.

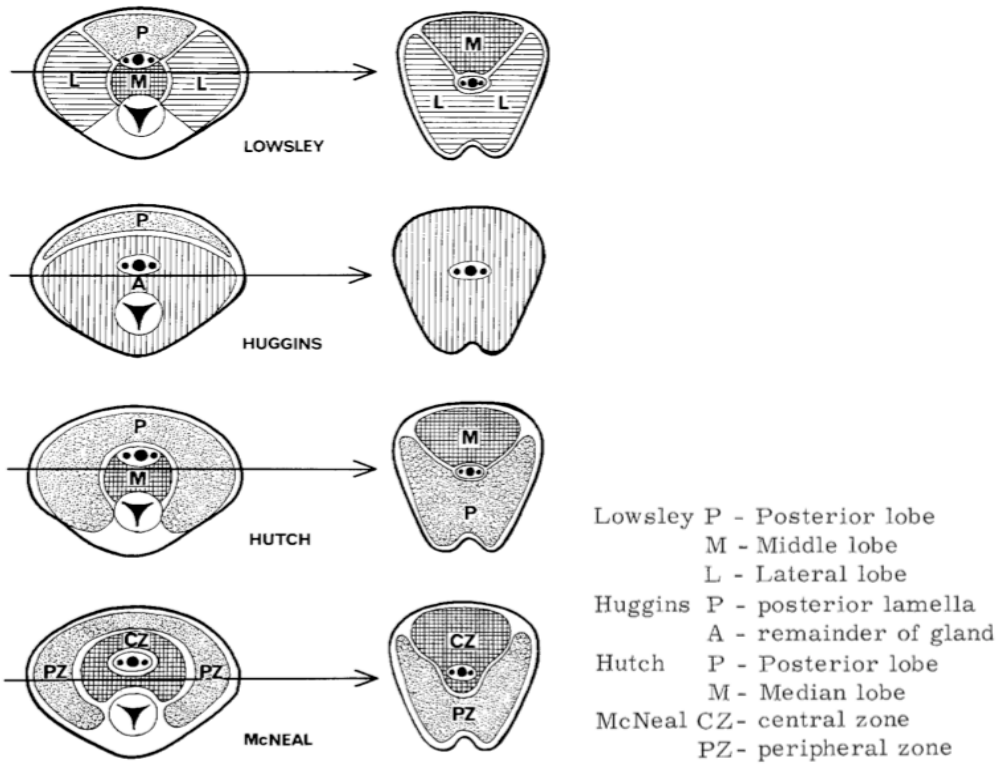


Figure 1 The evolution of knowledge of prostate anatomy (Blacklock, 1991)

1.1.2 The Prostate: Huggins' Description

In 1948 Huggins (Huggins and Webster, 1948) proposed the existence of a region that he called the posterior lobe but which had different boundaries to Lowsley's description (Fig. 1). It did not extend far enough into the prostate to contact the ejaculatory ducts. It could not be identified by any visible landmark: it was identified only by the effect of prolonged oestrogen administration to produce more profound atrophy here than in the rest of the gland. Like Lowsley, Huggins though that the posterior lobe was the selective site of origin of carcinoma, despite the fact that he was referring to different tissue than that specified by Lowsley. No further work was done on this hypothesis.

1.1.3 The Prostate: Franks' Description

In 1954, LM Franks was unable to identify lobe boundaries within the prostate, and his observations on cancer led him to conclude that carcinoma could arise anywhere in the territory of the posterior, lateral and middle lobes (Franks, 1953). He described this as the outer gland. Unlike Lowsley's description, he included the periurethral

glands as part of the prostate, and called this the inner gland. Franks also identified the inner gland as the exclusive site of origin of BPH.

1.1.4 The Prostate: Hutch's Description

In 1970, Hutch (Hutch, 1972) proposed, on the basis of gross dissections, that the middle lobe of Lowsley was part of the periurethral gland group (Figure 1). For many years it has been common colloquial usage among urological surgeons to refer to the midline and laterally projecting nodules of BPH as the middle and lateral lobes, respectively. Hutch's paper appears to have been an effort to confirm the validity of this concept. This formulation acknowledges that the entire area of Franks' outer gland (Lowsley's four lobes) is susceptible to carcinoma, but it refers to this entire area as the 'posterior lobe'. Thus, the middle and lateral lobes are relegated to the periurethral gland area. This interpretation is contrary to Lowsley's original concept, but his name has become attached to it. Furthermore, these 'lobes' are not reference points of normal anatomy but exist only in glands with BPH. This concept is an indiscriminate mixture of Lowsley, Franks, and pathologic anatomy. It represents a source of great confusion and can still remain a popular way to regard the anatomy of the prostate.

1.1.5 The Prostate: McNeal's Description

Until McNeal the prostate had been thought of as histologically homogeneous organ with simple anatomic structure. First in 1968 (McNeal, 1968) and then in 1981 (McNeal, 1981) McNeal published his concept of the prostate having a 'zonal anatomy'. Based on observations from prostatic tissue specimens from over 500 autopsies, he described an anatomically heterogeneous organ, composed of four separate regions, whose relationship to the urethra provided a central anatomic reference point (Figure 2, 3). Using serial histological blocks he built up a three-dimensional model of the anatomy. He found that the glandular tissue immediately surrounding the ejaculatory ducts was histologically different from the rest of the gland. The ducts of the region entered the urethra closely associated with the ejaculatory duct orifices and formed a base of tissue with its base at the base of the prostate, and apex at the veromontanum. The ejaculatory ducts ran through its centre and in young adults it made up approximately 25% of the prostate. He termed this region the 'central' zone.

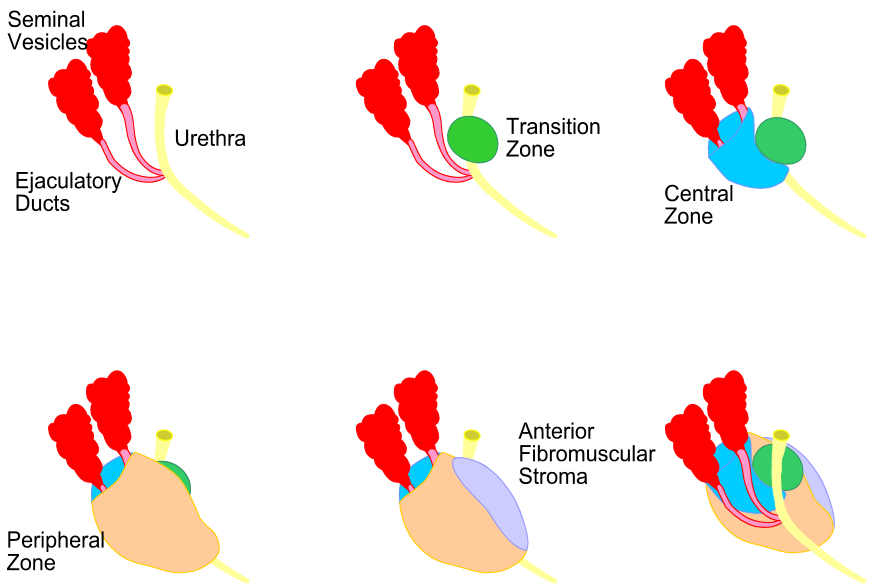


Figure 2 Zonal anatomy of the prostate (McNeal, 1981)

The remainder of the prostate was termed the 'peripheral' zone and contributed to over 70% of the prostate. The acinar tissue of the central zone consisted of large spaces of irregular contour, whose walls bore numerous ridges or septa projecting into the lumen. The epithelial cells appeared crowded, with nuclei of large size lying at different levels in a dark, granular cytoplasm. By contrast, the 'peripheral zone', was composed of small, round, regular acini with smooth walls.

McNeal proposed that the central zone might share a common embryological origin with the seminal vesicles from the Wolffian duct, given its close proximity and the similarity of its epithelium. The rest of the prostate was said to arise from the urogenital sinus. He, along with those before him, noted that prostate cancer arose much more frequently from the peripheral zone, and this raised the strong possibility of a difference in biological function between the zones.

In his work published in 1981, McNeal described the preprostatic sphincter, a sphincter of smooth muscle surrounding the urethra from the upper end of the veromontanum proximally to the bladder neck. Functionally this was thought to aid in continence, and further study revealed a small duct system that lay external to the sphincter, and was best developed along the distal half of its extent. He referred to this region as the 'transition' zone. It contributed less than 5% of the mass of the glandular prostate, with its histology and architecture closely resembling that of the peripheral

zone. Anatomically a unique feature of this region was the penetration of its ducts into the external layers of the sphincter. Large BPH nodules were seen to develop almost invariably in this zone, and thus its significance pathologically.

Thus greater than 70% of the prostate consisted of the peripheral zone, which formed a disc of tissue whose duct radiated laterally from the urethra lateral and distal to the veromontanum. Almost all carcinomas arose here.

The central zone constituted 25% of the prostate. Its lateral border fused with the proximal peripheral zone border, completing in continuity with the peripheral zone, a full disc of secretory tissue orientated in a coronal plane. Marked histological differences between central and peripheral zones suggested important biological differences.

The urethral segment proximal to the veromontanum was kinked anteriorly at a 35-degree angle to the distal segment. No major ducts arose in the proximal segment, but the lateral rows of peripheral zone orifices continued. Duct development was aborted there, producing only a small transition zone (5%) and several tinnier periurethral ducts. The development of these small ducts was possibly determined and limited by their intimate relationship to a periurethral smooth muscle sphincter that exists only proximal to the veromontanum. These small ducts were in a restricted area and were the exclusive site of nodular hyperplasia (BPH).

The anterior fibromuscular stroma formed the entire anterior surface of the prostate as a thick, nonglandular apron, shielding from view the anterior surface of the three glandular regions.

Evidence in support of this model has been derived from studies of histochemistry, protein expression (Colombel *et al.*, 1998, Kobayashi *et al.*, 1991, Krill *et al.*, 2001, McNeal *et al.*, 1988a, Reese *et al.*, 1986, Reese *et al.*, 1988, Reese *et al.*, 1992, Tsurusaki *et al.*, 2003), and histology (McNeal, 1988) of the prostate, as well as clinical observations using CT, MRI, Ultrasound (Hricak *et al.*, 1987, Mirowitz and Hammerman, 1992, Sommer *et al.*, 1986, Villers *et al.*, 1990), and histopathology (Colombo *et al.*, 2001, Erbersdobler *et al.*, 2002b, Erbersdobler *et al.*, 2002a, McNeal, 1968, McNeal *et al.*, 1988b).

1.1.6 The Prostate: Tissel's Description

In 1975, Tissel (Tisell and Salander, 1975) also reported histological heterogeneity within the prostate. Using blunt dissection on gross specimens, he was able to find

cleavage planes that separated three regions having microscopically different appearances. He designated these as the middle, lateral, and posterior lobes.

The middle lobe he described did not correspond in location to that of Lowsley. It completely surrounded the ejaculatory ducts as a wedge of tissue, with its apex at the veromontanum and its base against the base of the prostate. Tissel has since concurred that the lobe had boundaries and histological features identical to those described by McNeal for the central zone (McNeal, 1981).

Tissel's posterior lobe also did not correspond to that of Lowsley. It was a strip of tissue lying in a subcapsular location just beneath the entire posterior surface of the prostate. It did not extend into the gland far enough to contact the ejaculatory ducts and was separated from them by the middle and lateral lobes. Its location corresponded to that of the posterior lobe described by Huggins. In this region the histology was the same as that for the peripheral zone.

The lateral lobes described here were also different from those of Lowsley, in that there was a bridge of glandular tissue posterior to the ejaculatory ducts that united the two lateral glandular masses into a single lobe. In this region the histology resembled that of the peripheral zone more than the central zone, but the acinar spaces were of larger size.

It is difficult to understand why the terminology of Lowsley was used since the prostate of Tisell seems to be identical to that of McNeal, except for the description of a subdivision within the peripheral zone. (This is perhaps explained by the fact that Tisell founds this distinction clearer in glands from men aged over 40). Older prostates are a questionable reference point for normal anatomy, since deviations from the structure of normal young adults are already apparent in men in the fifth decade (McNeal, 1968, McNeal *et al.*, 1988b). By contrast, the lobes defined by Lowsley are quite different in anatomic boundaries from those of Tisell. These issues cannot be easily resolved since none of the work of McNeal was referred to in Tissel's paper.

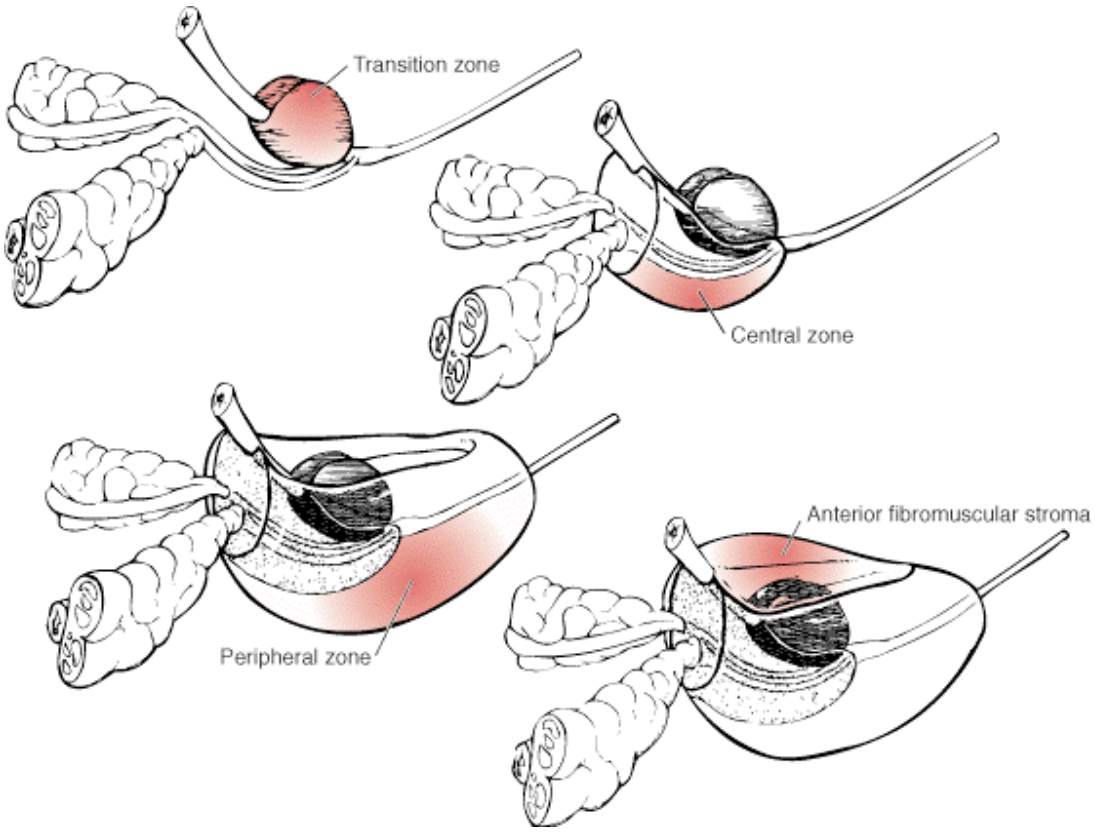


Figure 3 McNeal's zonal anatomy of the prostate (Campbell's Urology)

Although the gross anatomy of the prostate has been established there is little evidence for the molecular basis of these zones, and this is discussed subsequently (Chapter 1.10). The clinical significance of zonal anatomy is important in terms of development of both BPH and prostate cancer. Nodules of benign prostate tissue originate within and then expand the transition zone, distorting and compressing the adjacent peripheral zone. Prostate cancer, although it may affect any of the zones, is more prevalent (70%) in the peripheral zone (McNeal, 1968). The reasons for this geographic difference must lie at the molecular level, but remain undefined.

1.2 Normal function of the prostate

Sex accessory tissues include the prostate, seminal vesicles, ampullary glands, and bulbourethral glands, and they are believed to play a major but unknown role in the reproductive process. In the human, the sex accessory tissues produce extremely high concentrations of many important and potent biologic substances that appear in the

seminal plasma, such as fructose, citric acid, spermine, prostaglandins, zinc, proteins that include immunoglobulins, human kallekriens, semenogelins, and so forth, and specific enzymes such as proteases, esterases, and phosphatases (Wein *et al.*, 2006). At present, there is only limited knowledge of the physiologic function of any of these potent secretory products in the seminal plasma, with the exception of specific enzyme activities in the clotting and lysing process occurring with seminal plasma proteins that have unknown physiologic functions. Although the seminal plasma may not contain factors that are absolutely essential for fertilization, the secretions nevertheless may optimize conditions for fertilization by providing a buffered effect, by increasing sperm motility or survival, or by enhancing sperm transport in both the male and female reproductive tracts. The seminal plasma may extend viability of the sperm and decrease environmental shock.

1.3 Major prostatic diseases – benign prostatic hyperplasia (BPH) and prostate cancer

1.3.1 BPH

BPH develops almost exclusively in the transition zone (McNeal, 1968). It is considered to be a stromal disease (Rohr and Bartsch, 1980) although in some cases the epithelial component predominates. Prostatic enlargement by hyperplasia results in reduced compliance of the prostatic urethra and increasing outflow obstruction to urine (Wilson, 1980) that may result in secondary changes in the bladder: hypertrophy of the bladder wall, and the formation of trabeculae as well as diverticulae. Upper urinary tract dilatation may also result. Symptoms of urethral obstruction are due either to bladder filling or voiding problems and may include urine retention, nocturia, poor stream, urgency and frequency (Wein *et al.*, 2006). If left untreated, BPH can result in more severe symptoms, including urinary retention (acute and chronic), urinary tract infections, bladder stones, detrusor instability (sometimes resulting in incontinence), and obstructive uropathy. Operative relief can result in a reversal in the secondary changes in the bladder, provided they are not too advanced and chronic over distension of the detrusor has not developed. Approximately 10% of men over 50 will require intervention (surgical or medical) to relieve obstruction and 90% of men in their 80s have histopathological evidence of BPH (Wein *et al.*, 2006). Incidence rates increase from 3 cases per 1000 man-years at age 45-49 years (prevalence rate of 2.7%), to 38 cases per 1000 man-years by the age of 75-79 years

(prevalence rate of 24%) (Verhamme *et al.*, 2002). The clinical outcome of these statistics is ~40,000 resections per year in the UK (Wein *et al.*, 2006).

1.3.2 Prostate cancer

Prostate cancer is the most common malignancy in men and is the cause of considerable morbidity and mortality (Howe *et al.*, 2001). In the year 2010, 217,730 new cases of prostate cancer were diagnosed in the United States, with 32,050 American males dying of the disease (Jemal *et al.*, 2010). Early prostate cancer is often asymptomatic. It may be discovered during routine screening with the PSA blood test or during investigation for lower urinary tract symptoms. It may cause lower urinary tract symptoms similar to BPH, including as well, painful micturition and haematuria. It may also cause erectile dysfunction or painful ejaculation. Advanced prostate cancer may cause local symptoms secondary to invasion of adjacent structures such as ureter, bladder, and rectum. Alternately it may cause distant disease by metastases. It shows a disposition to spread to the skeleton and may cause fractures, hypercalcaemia, and anaemia (Wein *et al.*, 2006).

To date, the exact triggers for development of prostatic adenocarcinoma have not been elucidated, although epidemiological studies have shown links to familial, environmental and genetic contributors (Hughes *et al.*, 2005). A variety of growth factors and cytokines have been shown to influence the growth rate and development of prostatic cancers, and a number of proteins and genes have been identified whose regulation and/or function are altered between malignant and benign states (Hughes *et al.*, 2005). Several of these proteins and genes have been suggested as potential biomarkers and therapeutic targets for prostate cancer (Bradford *et al.*, 2006) but few have reached clinical practice.

Prostate cancer is classified according to the TNM (Tumor Node Metastasis) classification (Sobin, 2009) (Figure 4).

T - Primary tumour		
TX	Primary tumour cannot be assessed	
T0	No evidence of primary tumour	
T1	Clinically inapparent tumour not palpable or visible by imaging	
	T1a	Tumour incidental histological finding in 5% or less of tissue resected
	T1b	Tumour incidental histological finding in more than 5% of tissue resected
	T1c	Tumour identified by needle biopsy (e.g. because of elevated prostate-specific antigen [PSA] level)
T2	Tumour confined within the prostate[1]	
	T2a	Tumour involves one half of one lobe or less
	T2b	Tumour involves more than half of one lobe, but not both lobes
	T2c	Tumour involves both lobes
T3	Tumour extends through the prostatic capsule[2]	
	T3a	Extracapsular extension (unilateral or bilateral) including microscopic bladder neck involvement.
	T3b	Tumour invades seminal vesicle(s)
T4	Tumour is fixed or invades adjacent structures other than seminal vesicles: external sphincter, rectum, levator muscles, and/or pelvic wall	
N - Regional lymph nodes[3]		
	NX	Regional lymph nodes cannot be assessed
	N0	No regional lymph node metastasis
	N1	Regional lymph node metastasis
M - Distant metastasis[4]		
	MX	Distant metastasis cannot be assessed
	M0	No distant metastasis
	M1	Distant metastasis
		M1a Non-regional lymph node(s)
		M1b Bone(s)
		M1c Other site(s)

[1] *Tumour found in one or both lobes by needle biopsy, but not palpable or visible by imaging, is classified as T1c.*

[2] *Invasion into the prostatic apex, or into (but not beyond) the prostate capsule, is not classified as pT3, but as pT2.*

[3] *Metastasis no larger than 0.2 cm can be designated pN1 mi.*

[4] *When more than one site of metastasis is present, the most advanced category should be used.*

Figure 4 Tumour Node Metastasis (TNM) classification of prostate cancer (Sobin, 2009)

The Gleason score is the most commonly used system for grading adenocarcinoma of the prostate (Gleason and Mellinger, 1974). This scoring system was invented by Donald Gleason, a histopathologist working at the Minneapolis Veterans Association medical centre in 1962 (Egevad *et al.*, 2009). He was tasked with constructing a standardized scoring system for prostate cancer, that would allow effective communication between hospitals. It was twenty years later that this scoring system became widely accepted, and persists to this day. The Gleason score can only be assessed using biopsy material (core biopsy or operative specimens). Cytological preparations cannot be used. The Gleason score is the sum of the two most common patterns (grades 1-5) of tumour growth found. The Gleason score ranges between 2

and 10, with 2 being the least aggressive and 10 the most aggressive. In needle biopsy, it is recommended that the worst grade always should be included, even if it is present in < 5% of biopsy material (Amin *et al.*, 2005).

1.4 Genetics and molecular biology of prostate cancer

Direct genome modification and epigenetic influences are the two well described factors that may contribute to the development of prostate cancer. Approximately one quarter of all prostate cancer occurs in family clusters, with 9% of hereditary prostate cancer following an autosomal susceptibility trait (Carter *et al.*, 1993, Gronberg *et al.*, 1997). Several genes have subsequently been identified and implicated in hereditary prostate cancer including HPC1 (Gronberg *et al.*, 1997), and HPC2 (Rokman *et al.*, 2001).

Chromosomal loses, and occasionally gains, are frequently seen in prostate cancer. The most common loses occur in 8p and 13q (Bova *et al.*, 1993), with several genes in these regions identified as putative tumour suppressor genes. These include NKX3.1 and Retinoblastoma (He *et al.*, 1997, Phillips *et al.*, 1994) NKX3.1 is a homeobox gene that plays a key role in the regulation of prostate development (Bhatia-Gaur *et al.*, 1999). It is prostate specific tumour suppressor gene, with loss of a single allele predisposing to prostate carcinogenesis (Bhatia-Gaur *et al.*, 1999). The retinoblastoma gene is also a prostate tumour suppressor gene (Bookstein *et al.*, 1990). Regions of chromosome gain include amplification of the entire 8q and Xq11-13, which include the loci for MYC gene and the androgen receptor, both putative oncogenes (Nupponen *et al.*, 1998, Koivisto *et al.*, 1997).

More recently, epigenetic influences on the genome, defined as heritable changes in gene expression that occur without changes in DNA sequence (Wolffe and Matzke, 1999), have also been implicated in prostate carcinogenesis (Rennie and Nelson, 1998). The most common forms of epigenetic modification relate to DNA methylation. Oxidative damage may be one mechanism by which prostate carcinogenesis is initiated. Certain genes that protect against oxidative damage may be altered by methylation, for example glutathione-S-transferase P1 (GSTP1) (Florl *et al.*, 2004). Hypermethylation of the promoter of GSTP1 is related to its decreased expression in many prostate cancers (Lin *et al.*, 2001).

Numerous tumour suppressor genes and oncogenes have been implicated in prostate carcinogenesis, many of which are involved in varied biological processes, such as apoptosis, signal transduction, cell cycle regulation, cell adhesion and cohesion, and angiogenesis. These include, p53, Bcl-2, Clusterin, TGF-beta, Androgen Receptor, c-Myc, p27KIP1, pRb, E-cadherin, kallikreins, VEGF, AMACR and Hepsin (Shand and Gelmann, 2006)

1.5 Microarray technology

The central dogma of molecular biology states that genetic information flows from DNA to messenger RNA (mRNA) and from RNA to proteins which perform gene functions (Crick, 1970), and this process is called gene expression. Not all genes code for proteins, however. Some genes produce RNAs that aren't translated into proteins and are therefore called noncoding RNAs. In general the amount of RNA can thus be used an indicator of the level of gene expression. A microarray is "an ordered array of microscopic elements on a planar substrate that allows the specific binding of genes or gene products" (Schena *et al.*, 1995). Microarrays measure gene expression levels on a genomic scale simultaneously by monitoring the abundance of the intermediary mRNA (Schena *et al.*, 1995, Lockhart *et al.*, 1996). Over the last decade microarrays have become an increasingly important tool in biomedical and life sciences research.

The key idea of microarray technology is binding or hybridisation, which is the chemical process where the two complementary strands of DNA or RNA combine to form a double strand under certain conditions. The "microscopic elements" are single stranded nucleotide sequences, which are called probes, fixed to the surface of microarrays. "Genes or gene products" are mRNA or total RNA molecules isolated from the biological specimens termed targets. Targets are fluorescently labelled (e.g. with cyanaine 3, Cy3, which has a fluorescence emission wavelength of 570 nm corresponding to the green part of the light spectrum) and mixed in solution. The mixture of targets, known as sample, is then hybridised to the microarray. The RNA sequences of the targets bind to their complementary probes (hybridisation). After a certain time allowed for hybridisation, the arrays are washed to get rid of the extra sample and the arrays are scanned to obtain a two dimensional image. The intensity at each probe position indicates the amount of RNA molecules bound to the specific probe and provides a quantitative measurement of the expression level of the related gene.

A typical microarray experiment involves the following four main steps:

1. Isolate RNA from the tissue of interest and prepare fluorescently labelled targets.

2. Hybridise the labelled targets to the microarray.

3. Wash, process and scan the microarray.

4. Process the resulting image to obtain a quantitative measurement of the intensity for each probe.

Among currently available microarray technologies, there are three widely used classes, spotted cDNA microarrays first developed at Stanford (Schena *et al.*, 1995), synthetic bead arrays manufactured by Illumina (Kuhn *et al.*, 2004), and synthetic oligonucleotide microarrays mainly produced by Affymetrix (Lockhart *et al.*, 1996).

Spotted Microarrays

Spotted arrays (cDNA) are prepared robotically by applying precisely measured quantities of either PCR products generated from cDNA clones, or gene specific oligonucleotides, to glass slide (Brown and Botstein, 1999). Long oligonucleotides (40-70 nucleotides) are favoured by many allowing for high specificity and sequence identity. Spotted arrays are more flexible than synthetic arrays, however their synthesis requires skill in application and quality control. The costs are higher at initial outlay, but once fabricated they are cheaper. They provide excellent results at reasonable costs but have declined in use as other technology has become cheaper and more accessible.

Synthetic oligonucleotide arrays: Affymetrix microarray system

Affymetrix is one of the leading manufacturers of microarrays and its GeneChip® technology is hugely popular (Lockhart *et al.*, 1996, Auer *et al.*, 2009) (www.affymetrix.com/community/publications/index.affx). Using synthetic oligonucleotide technology, Affymetrix microarrays are high-density DNA probe arrays that contain millions of probe sequences. The expression level of each gene is measured by multiple probes on the microarray. In Affymetrix microarray technology, probes are oligonucleotides, which are short single-stranded nucleotide sequences (DNA or RNA) and usually include 25 nucleotides. According to the available sequence information, probes are chemically synthesized from DNA and RNA building blocks, nucleotides, at a specific location on the surface of arrays (Lockhart *et al.*, 1996). The precise location where each probe is synthesized is called a feature.

One single high-density Affymetrix array (HG-U133 Plus 2.0) with typical size 1.28cm × 1.28cm contains 1,300,000 distinct oligonucleotide features (Figure 5).

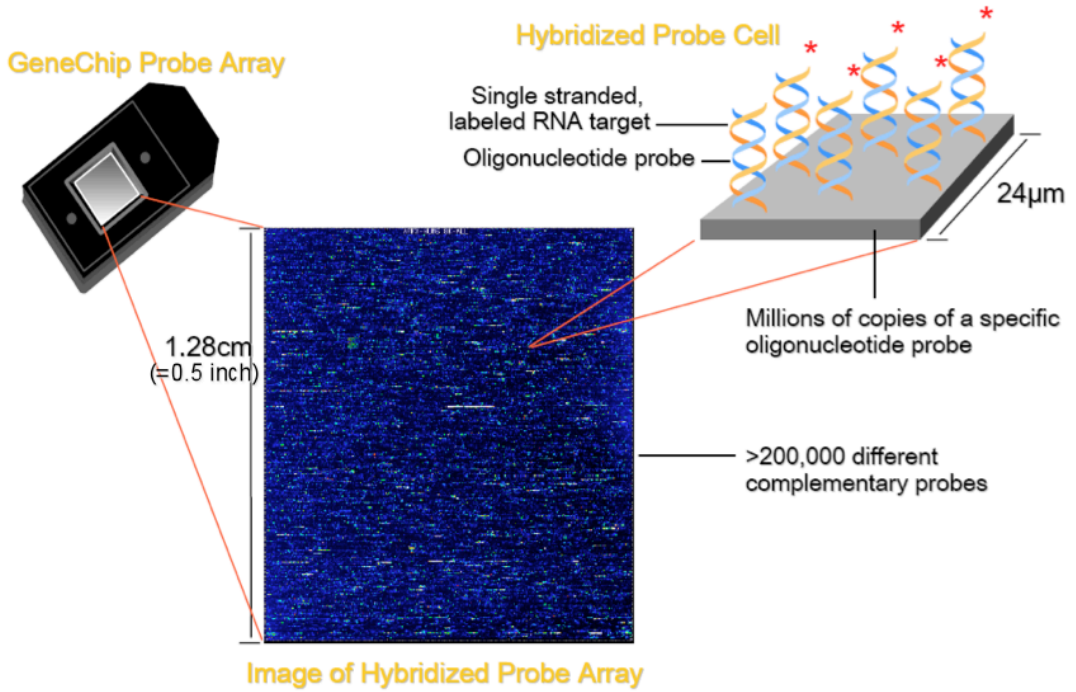


Figure 5 Affymetrix GeneChip® design

Apart from the synthetic oligonucleotides, another speciality of Affymetrix microarray technology is the redundancy in the probe design. The concept of redundancy is embodied in two aspects. One is that each gene corresponds to multiple probes on the array, another is that arrays contain pairs of probes for each of the RNA sequences being monitored. For each gene, the reference sequence comes from its related spliced mRNA, which contains only exons and flanking RNA. A subset of exon-specific probes is specifically chosen in order to detect the spliced mRNA in samples. The set of probes related to a particular gene is called a probe-set.

Two types of hybridisation occur on the array during the binding of targets to probes, specific hybridisation and non-specific hybridisation (Lockhart *et al.*, 1996), also known as specific binding and non-specific binding respectively. Specific hybridisation means that the double-stranded molecule is formed from two perfectly complementary strands, one from probe sequences and another from target sequences. Non-specific hybridisation, sometimes also called cross-hybridisation in the literature

(Casneuf *et al.*, 2007), refers to the hybridisation that occurs between two strands, which are not perfectly complementary.

Each probe on the array that is perfectly paired with its target sequence is called a perfect match (PM) probe. In order to identify the non-specific hybridisation, for each PM probe on the array there is a mismatch (MM) probe which has the identical nucleotide sequence as the PM probe except that the middle nucleotide is changed to the complementary one. A set of paired oligonucleotide probes, typically 25-mers, is designed for each gene. Each pair contains the canonical sequence, or perfect match probe, of the gene and also a deliberate mutation in the 13th position (middle) of the gene, or mismatch probe. The mismatch probe measures the degree of cross hybridization, or how much lower the detection signals for noise are. For example, A is changed to T and C is changed to G, and vice versa. There are 11-20 PM/MM probe-pairs contained in each probe-set. By design the MM probe detects the non-specific hybridisation on its complementary PM partner, since with only one base replaced the MM probe has a similar efficiency of binding to the non-specific target sequences as the PM counterpart. The MM probes therefore serve as internal controls for hybridisation specificity. In order to avoid minor defects in the hybridization image, probes are scattered throughout the surface of the arrays. During the experiment, fluorescent labeled RNA molecules are fragmented and hybridised to the array. The degree of hybridisation is assessed by monitoring fluorescent emission using a laser scanner. For each chip, a two dimensional image is created with each probe being identified by its coordinates on the array and measured for its fluorescent intensity. The measured intensity values represent the expression level of the related gene and coordinates on the array are stored in a cell intensity file (*.CEL) as the final results of the experiment. Each chip corresponds to a CEL file. For each type of chip related to each particular organism, Affymetrix Corp. provides an array layout description file (*.CDF). The CDF file describes the design of a chip defining which probes belong to which probe-sets. By looking up the CDF file the intensity values for each probe-set can be extracted. The relationships between probe-sets and genes are also provided by the manufacturer in the documentations.

Synthetic bead arrays: Illumina

Illumina has developed a novel bead array technology. A multicore optical ‘imaging’ fiber is etched such that a bead can fit into the resulting micron-sized etched wells on the tip of the fiber. Different oligonucleotide sequences are attached to each bead, and

thousands of beads can be self-assembled on the fiber bundle. A subsequent decoding process is carried out to determine which bead occupies which well. Complementary oligonucleotides present in the sample bind to the beads, and bound oligonucleotides are measured by using a fluorescent label (Oliphant *et al.*, 2002).

1.6 Process of Biological Study Using Microarrays

A typical microarray experiment is usually motivated by a biological question, like "which genes show changes in expression between a healthy tissue and a diseased one", or a more specific biological hypothesis, like "a certain group of genes are responsible for the development of a particular disease". It follows that an answer to the question or the verification of the hypothesis may help with the diagnosis and treatment of the particular disease.

Large amounts of gene expression data are generated from microarray experiments. In order to obtain meaningful biological information for the organism being studied, multiple levels of analyses are performed on the primary data. The first stage of the analysis is probe-level analysis, which summarises the raw signal intensities of the individual probe pairs, to obtain a single expression value for each gene. The probe-level analysis should provide reliable measurements of gene expression levels, which can be used in the high level analysis. The next stage of analysis (high level analysis) performs various tasks on the measured gene expression resulting from the probe-level analysis (Quackenbush, 2001, Slonim, 2002). These tasks include detecting differential gene expression or identifying patterns of gene expression between two conditions (e.g. normal vs. diseased, treated vs. untreated etc), but primarily depend upon the biological questions posed for the experiment.. Differential gene expression constitutes the most basic aim of a microarray experiment involving two or more microarrays and can be achieved by comparing 2 different GeneChips® in an Affymetrix experiment. Gene expression patterns can be discovered by "unsupervised" approaches and gene functions can be predicted by "supervised" approaches (Bassett *et al.*, 1999). The unsupervised approaches include principal component analysis and clustering. These methods are applied on data solely and do not involve any previous knowledge of experimental parameters such as treatment effect with a drug. Clustering could be performed in different ways but in common practice genes that have similar functions are organised into the same cluster or stay

close together in the visual representation resulting from principal component analysis. With known biological information, unknown gene functions can be revealed by examining the known classes they fall in by supervised analysis approaches (Pan, 2006). Inferring gene regulatory networks is another important goal of high level analysis of gene expression data (Segal *et al.*, 2003, Friedman, 2004).

At the final stage of the discovery process, biological conclusions are drawn based on the results obtained from the high level analysis. As a result, either new biological knowledge is discovered or the original hypothesis is falsified. Every step in the discovery process is vital to subsequent steps. Within a carefully designed and performed experiment, analysis of the experimental data plays an important role in making sound biological conclusions.

1.7 Use of Microarrays in Cancer Research

In the past several years, microarray technology has been widely used in cancer research and many studies have used this technology to identify candidate gene expression signatures to predict the diagnostic category or prognostic stage of a cancer patient (Sorlie *et al.*, 2001), 14-32].

One of the common problems in clinical cancer research is the fact that histopathological identification and classification of cancer can be quite challenging. Morphologically indistinguishable cancers may belong to clinically distinct classes even though they arise from the same origin. The ability to classify unknown samples into different categories may offer great potential for more accurate and systematic cancer diagnosis.

Golub *et al.* (Golub *et al.*, 1999) first demonstrated the use of gene expression profiling for cancer diagnosis in a study using DNA microarrays to study the gene expression of 6,817 genes in 72 human acute leukaemia tumour samples. In this study, using unsupervised learning analysis, leukaemia tumour samples were clustered into two clusters of known subtypes of leukemia – acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL) – based solely on gene expression profiling. Using supervised learning analysis, a weighted gene voting classifier, built from a subset of informative genes chosen based on their correlation with the class distinction between AML and ALL, was successfully used to assign a group of unknown samples into the correct category. The accuracy of the classifier was assessed by both cross-validation on the initial training data and independent test on

an independent set of samples (for explanation see Appendix). This study demonstrated the feasibility of cancer classification based solely on gene expression.

A variety of studies have subsequently used gene expression profiling for cancer classification (Welsh *et al.*, 2001b, Bittner *et al.*, 2000, Perou *et al.*, 2000, Bhattacharjee *et al.*, 2001, Hedenfalk, 2002, Ramaswamy *et al.*, 2001)(Belbin *et al.*, 2002, Thomas *et al.*, 2001). Bittner *et al.* (Bittner *et al.*, 2000) reported the discovery of a subset of cutaneous melanomas identified by mathematical analysis of gene expression in a series of 31 melanomas samples. Both the hierarchical clustering of the 31 melanoma samples and the non-hierarchical cluster affinity search technique (CAST) algorithm identified the identical major cluster of 19 melanomas, which had very similar gene expression profiles (for explanation see Appendix). In another study, Hedenfalk *et al.* (Hedenfalk, 2002) demonstrated that in hereditary breast cancers, the gene expression profiles of tumours with BRCA1 mutations, tumours with BRCA2 mutations, and sporadic tumours differed significantly from each other. Statistical analyses were used to identify a set of 176 genes, which could accurately distinguish tumours with BRCA1 mutations from tumours with BRCA2 mutations on the basis of gene expression profiles of these genes.

Diagnostic cancer classification studies have been performed in a wide range of cancers, including breast cancer (Perou *et al.*, 2000), lung cancer (Bhattacharjee *et al.*, 2001), prostate cancer (Singh *et al.*, 2002, Welsh *et al.*, 2001b, Stuart *et al.*, 2004, Lapointe *et al.*, 2004), bladder cancer (Dyrskjot *et al.*, 2003), head and neck cancer (Belbin *et al.*, 2002), and ovarian cancer (Welsh *et al.*, 2001b, Ono *et al.*, 2000).

For classification of multiple categories of human cancers, several recent studies explored the potential use of multiclass tumour classifiers built from microarray gene expression data to discriminate different kinds of tumour classes based on tissue origin (Ramaswamy *et al.*, 2001, Tibshirani *et al.*, 2002, Liu *et al.*, 2005). Many publications report on cancer classification problems where the number of classes is rather small. For example, the classification problem of Golub et al. comprised only two classes (Golub *et al.*, 1999). When many 'classes' are being considered (e.g. cancer stage and biological behaviour), these tools are referred to as multiclass tumor classifiers. In an effort to classify multiple common adult cancers purely by molecular classification, Ramaswamy *et al.* (Ramaswamy *et al.*, 2001) collected gene expression profiles of 16,063 genes and expressed sequence tags for 90 normal tissue samples and 218 tumour samples spanning 14 tumour types. The accuracy of a multiclass

classifier based on a support vector machine algorithm (see Appendix) was evaluated by both cross-validation and independent test data. The overall classification accuracy was approximately 78%. They also found that poorly differentiated cancers could not be accurately classified according to their tissues of origin, indicating that they have a very different gene expression patterns compared to their well-differentiated counterparts.

Different classification methods have been proposed for molecular classification of cancers based on gene expression profiles. These methods include classification trees (Zhang *et al.*, 2001), nearest-neighbor classifiers (Giordano *et al.*, 2001), linear discriminant analysis (LDA) (Shen *et al.*, 2006), support vector machines (SVMs) (Ramaswamy *et al.*, 2001), and artificial neural networks (ANNs) (Khan *et al.*, 2001).

1.8 Microarrays in Prostate Cancer

Spotted cDNA and oligonucleotide microarrays have both been used to study gene expression profiles in prostate cancer. Numerous potential diagnostic markers for prostate cancer and of prostate cancer progression have been identified. There is variation between the most over- and under-expressed genes between studies, however some genes come up consistently in independent studies using different platforms.

1.8.1 Some important genes identified using microarray analysis in prostate cancer

Multiple studies have identified hepsin (HPN aka TMPRSS1) as overexpressed in prostate cancer compared to BPH (Dhanasekaran *et al.*, 2001, Luo *et al.*, 2001, Magee *et al.*, 2001, Welsh *et al.*, 2001a). Hepsin encodes a membrane bound serine protease which activates the hepatocyte growth factor/c-Met (HGF/c-Met) pathway by cleaving HGF (Kirchhofer *et al.*, 2005). This pathway leads to enhanced cell migration of prostate cancer cells via the production of matrix metalloproteinases (MMPs) and urokinase type plasminogen activator (uPA) (Fujiuchi *et al.*, 2003). The role of HPN in disease progression maybe over once dissemination of cancer outside the prostate has happened (Vasioukhin, 2004). HPN expression is lower in metastatic prostate cancer with forced over-expression of the gene in metastatic prostate cancer cell lines weakening the proliferation and colony-forming potential of the cells (Srikantan *et al.*, 2002).

A-methylacyl-CoA racemase (AMACR) has also been identified as an overexpressed gene in more than one microarray study (Zheng *et al.*, 2002, Dhanasekaran *et al.*, 2001, Welsh *et al.*, 2001a, Luo *et al.*, 2001, Luo *et al.*, 2002). These studies demonstrated significant over-expression of AMACR not only in prostate cancer samples but also in HGPIN, suggesting a potential role in early prostate cancer development. AMACR functions in the peroxisomal beta oxidation of branched-chain fatty acid molecules and has been implicated in the link between high meat high fat diets and the increased incidence of prostate cancer observed by many epidemiological studies (Mobley *et al.*, 2003). High levels of protein expression have been shown almost exclusively in malignant and pre-malignant cells using tissue microarrays (Rubin *et al.*, 2002)(Luo *et al.*, 2002). However normal prostate samples located adjacently to prostate cancer have also shown increased expression of AMACR, leading to the possibility and hypothesis of a cancer field effect (Ananthanarayanan *et al.*, 2005). AMACR is occasionally used clinically to reach a definite diagnosis of prostate cancer, when traditional histopathology and immunohistochemistry have yielded only an 'atypical' diagnosis (Das *et al.*, 2005). Epigenetic events implicated in prostate cancer have also been demonstrated in normal prostate tissue adjacent to prostate cancer (Mehrotra *et al.*, 2008). This study showed methylation of several genes in both histologically normal prostate tissue and prostate cancer.

1.8.2 Classification of prostate cancer based upon microarray analysis

In addition to the identification of individual genes microarray studies have shown that it is possible to distinguish prostate cancer from normal prostate tissue, and to distinguish organ confined cancer from metastatic disease (Dhanasekaran *et al.*, 2001, Luo *et al.*, 2001, Lapointe *et al.*, 2004, Welsh *et al.*, 2001a). Primary prostate cancer maybe clustered into subtypes with distinct clinical properties, even when pathological determinants, such as Gleason score and tumour stage, are similar (Luo *et al.*, 2002, Lapointe *et al.*, 2004). Microarrays have made it possible to identify molecular signatures (consisting of small clusters of four to five genes), which can predict patients likely to progress within one year of radical prostatectomy (Glinsky *et al.*, 2004). This study also identified a set of 70 genes that could be used to predict the aggressiveness of cancer with high accuracy (93% sensitivity and 87% specificity).

Meta-analysis, whereby the results of several studies are combined and analysed using statistical methods, has also been used to compare prostate microarray studies and assess the validity and relevance of the original findings. Meta-analysis can be useful to identify genes 'missed' in individual analyses due to small sample size and/or data filtering. Meta-analysis may find completely new sets of significant genes as shown by Xu *et al.* 2005, who identified a pair of genes, HPN and STAT6 (signal transducer and activator of transcription 6) by meta-analysis of five microarray studies (Xu *et al.*, 2005). These proved robust markers for distinguishing primary prostate cancer from benign samples.

The development of advanced bioinformatics has also led to the creation of online resources such as the Oncomine Cancer Microarray Database (http://www.oncomine.org) which allow meta-analysis on multiple experiments / datasets (Rhodes *et al.*, 2007a, Rhodes *et al.*, 2007b). This has recently led to the discovery of a gene fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer (Tomlins *et al.*, 2005).

1.9 Field effect theory

The concept "field effect in cancer" originated in 1953 from the histopathological observations of Slaughter *et al.* (Slaughter *et al.*, 1953) regarding the occurrence of multiple primary oral squamous cell carcinomas and their local recurrences. The development of modern molecular technologies has extended the field effect concept by exploring the molecular abnormalities in tissues that appear histologically normal. The possibility of a field effect has been explored and documented in other studies. Yu *et al.*, 2004, examined the differences in gene expression between healthy donor prostates, prostate cancer, and normal prostate adjacent to prostate cancer (Yu *et al.*, 2004). They showed that histologically normal prostate tissue from cancer patients shares many of the same molecular characteristics of prostate cancer. A similar microarray study also showed significant up regulation of proliferation related genes including transcription factors, signal transducers and growth regulators, in both tumour and adjacent normal tissue, when compared with donor tissue (Chandran *et al.*, 2005).

1.10 Molecular Differences between the anatomical zones of the prostate

Several studies have attempted to examine molecular differences between the three anatomical zones of the prostate. A major motivation has been to explain the differing propensity for diseases.

Most studies have used the basic over or under-expression to identify differences in gene expression. For example, pepsinogen II (Reese *et al.*, 1986), tissue plasminogen activator (Reese *et al.*, 1988), lactoferrin (Reese *et al.*, 1992), and lipochrome pigment (Leung and Srigley, 1995) have all been to shown at higher levels in the central zone than the peripheral or transition zone.

However to date only a few large scale gene expression analysis of the prostate zones using microarray technology have been performed (Stamey *et al.*, 2003, van der Heul-Nieuwenhuijsen *et al.*, 2006, Noel *et al.*, 2008). Stamey et *al.* attempted to address which zone of prostatic tissue was the best control in a microarray study of peripheral zone derived prostate cancer (Stamey *et al.*, 2003). They attempted to answer this question in their pioneering study by comparing gene expression in prostate cancers with 'normal' prostate tissue from areas of BPH, and central and peripheral zone prostate using the GeneChip HuGeneFL (6800 genes – Affymetrix, Santa Clara, CA). All tissue was harvested following radical prostatectomy and the tissue from the 3 zones was separated out after being identified by visual inspection. Substantially different gene expression profiles were found depending on which of the three zonal tissues was used as a control. Peripheral zone (PZ) was least able to identify or differentiate Gleason score 4/5 prostate cancer from normal tissue implying genetic similarity. CZ was considered to be the best at differentiating between normal and cancerous tissue, however inter-zonal gene expression analyses were not performed.

Van der Heul-Nieuwenhuijsen *et al.* 2006, used custom cDNA microarrays to compare the gene expression profiles of peripheral and transition zone normal whole prostate taken from 5 radical prostatectomy specimens (van der Heul-Nieuwenhuijsen *et al.*, 2006). 346 differentially expressed genes were identified, with 199 more highly expressed in the peripheral zone, and 147 more highly expressed in the transition zone. They also compared these gene lists with gene lists generated from prostate cancer microarray studies (Singh *et al.*, 2002, Dhanasekaran *et al.*, 2001, Lapointe *et al.*, 2004). They found a prominent overlap between genes expressed in their 'normal' prostate PZ and genes over expressed in prostate cancer.

Noel *et al.* 2008, used Affymetrix HGU133 Plus 2 gene chips to compare differential gene expression between normal peripheral and transition zone tissues obtained from radical prostatectomy specimens for prostate cancer (Noel *et al.*, 2008). They hoped to identify genes whose zonal-specific preferential expression might be associated with susceptibility or resistance to prostate cancer. Whole tissue specimens were used for RNA extraction from 3 prostates. Forty-three genes were identified as differentially regulated in the peripheral zone compared with the transition zone, 33 under expressed and 12 over expressed. Genes associated with neurogenesis, signal transduction, embryo implantation and cell adhesion were expressed at a higher level in the peripheral zone. Those over expressed in the transition zone were associated with neurogenesis development, signal transduction, cell motility and development. The authors discussed their findings but were unable to correlate their findings with disease predisposition.

Several proteome analyses of prostate tissue have also been described (Alaiya *et al.*, 2001, Meehan *et al.*, 2002, Ahram *et al.*, 2002), but did not evaluate the zonal differences within the prostate. Lexander *et al.* 2005, addressed this issue by comparing protein expression in the three zones using 2D gel electrophoresis (Lexander *et al.*, 2005). Seventeen radical prostatectomy specimens were divided up into zones for proteomic analysis using 2 dimensional gels and mass spectrometry. Ten proteins with significant zonal expression were identified, 8 under expressed in the CZ compared with the PZ and TZ, and 2 proteins were over expressed in the central zone. Proteins included: cytokeratin 8, laminin A/C, tropomysin and vimentin). Interestingly there were no significant differences in protein expression between the transition and peripheral zones.

1.11 Aims and objectives

There are some major issues with the majority of prostate microarray studies performed to date. Although various microarray profiles of the diseased prostate exists (Singh *et al.*, 2002, Dhanasekaran *et al.*, 2001, Welsh *et al.*, 2001a), no such information is available for the prostate obtained from normal, non-diseased prostate. Prostate adenocarcinoma is a disease of the epithelium, however, to date no study has performed global gene expression analyses with truly ‘normal’ prostate epithelium. Thus we have some idea of which genes may be dysregulated in cancer, but there is no benchmark for ‘normal prostate’ against which these studies could be compared. In

addition almost all the studies have been performed on RNA isolated from whole tissue with extensive contribution of heterogenous cell population, particularly the supporting mesenchyme. This may have a dual effect of diluting the expression of genes that are inevitably expressed in both the epithelium and the mesenchyme and exaggerating the effect of those that may be mesencyhyme specific. In addition the concept of a 'field effect' has been demonstrated (Yu *et al.*, 2004). Although the studies utilizing whole tissues for microarray analysis have made a major contribution and advances in our understanding of prostatic diseases, it is essential that a normal transcriptome map of the prostate epithelium is available. The establishment of this forms the basis of this dissertation, This should then not have the problems of dilution of gene expression by other cell types, and will be obtained from non-diseased normal human prostate.

The main aim of this dissertation was to provide the first transcriptome map of the normal human prostate epithelium by gene expression profiling using Affymetrix gene chip arrays. Further aims were to use the normal human prostate transcriptome map:

1. To compare with whole tissue transcriptome to identify discrepancies that could be introduced by using gene expression profiles from a heterogeneous cell population
2. To compare with an epithelial specific prostate cancer transcriptome to identify prostate adenocarcinoma specific genes
3. To delineate putative differences in gene expression between the three anatomical zones of the prostate

The specific objectives were:

1. To obtain normal prostates from organ donors aged between 15 and 35, and divide them into respective prostatic zones by visual examination.
2. To make 6-8μm tissue sections of the prostate samples
3. To precisely dissect prostate epithelial cells from the surrounding stroma using laser capture microdissection.
4. To isolate RNA from microdissected epithelial cells and perform global gene expression analysis using Affymetrix GeneChips.
5. To validate microarray 'targets' using real-time PCR.

6. To analyze this data using unsupervised and supervised protocols in pursuance of the aims described above.
7. To investigate potential novel epithelial biomarkers at the protein level using immunohistochemistry and low density tissue array

Chapter 2 Core techniques

2.1 Introduction

2.1.1 RNA assessment

Messenger RNA (mRNA) makes up 1-3% of total RNA, whilst ribosomal RNA (rRNA) makes up more than 80% of total RNA (tRNA) (Alberts, 2008). The majority of rRNA in humans is composed of 28S and 18S rRNA species. Traditionally mRNA quality has been assessed by denaturing gel electrophoresis of tRNA. Mammalian 28S and 18S are approximately 5kb and 2kb in size, and their ratio is approximately 2.7:1. A ratio of 2:1 is accepted as the standard for intact RNA. The assumption is that the quality and quantity of rRNA is reflective of the underling mRNA. Agarose gel assessment (Figure 6) is prone to error as a result of variations in electrophoresis conditions, amount of RNA loaded, and saturation of ethidium bromide fluorescence (Imbeaud *et al.*, 2005).

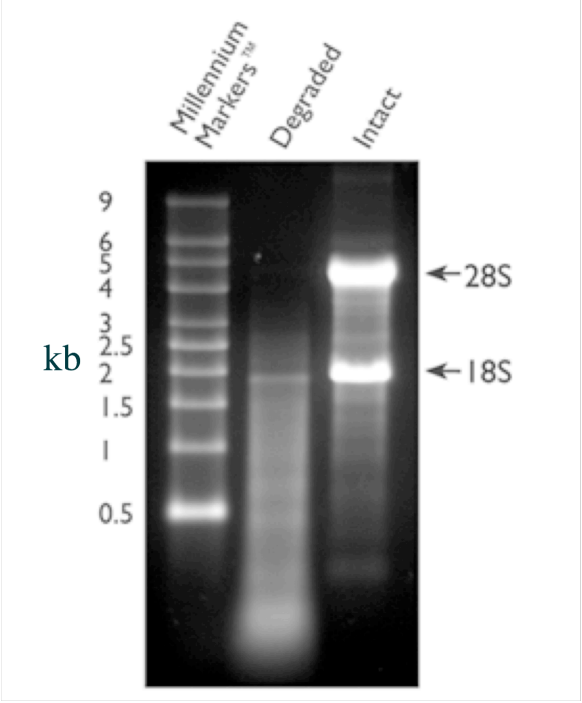


Figure 6 Denaturing gel electrophoresis of total RNA from primary cell cultured prostate epithelial cells. Two µg of degraded total RNA and intact total RNA were run beside Ambion's RNA Millennium Markers™ (kb) on a 1.5% denaturing agarose gel. The 18S and 28S ribosomal RNA bands are clearly visible in the intact RNA sample. The degraded RNA appears as a lower molecular weight smear. (Courtesy of Ambion)

An improvement on gel electrophoresis, and now the industry standard, is the 2100 Bioanalyzer (Agilent). This uses a combination of microfluidic technology, capillary electrophoresis, and fluorescent dyes that bind to nucleic acid to evaluate RNA concentration and integrity. It is quick, reproducible, accurate and requires only very small amount of sample, thus allowing assessment of RNA quality in limited samples.

The 28S:18S rRNA ratio is calculated by integrating the areas of 18S and 28S rRNA peaks and then dividing the area of the 18S rRNA peak into the area of the 28S rRNA peak (Figure 7). Whilst a ratio of 2 is optimal, and reflects high quality, it is rare for RNA isolated from human samples or small cell numbers. In part this reflects the inherent instability of 28S rRNA in relation to 18S rRNA, but also different tissues will demonstrate differing ratios and may never achieve the ideal ratio of 2 (Imbeaud *et al.*, 2005). When a sample is classified as degraded on 28S rRNA profile, its mRNA and 18S rRNA may still be intact (Figure 8).

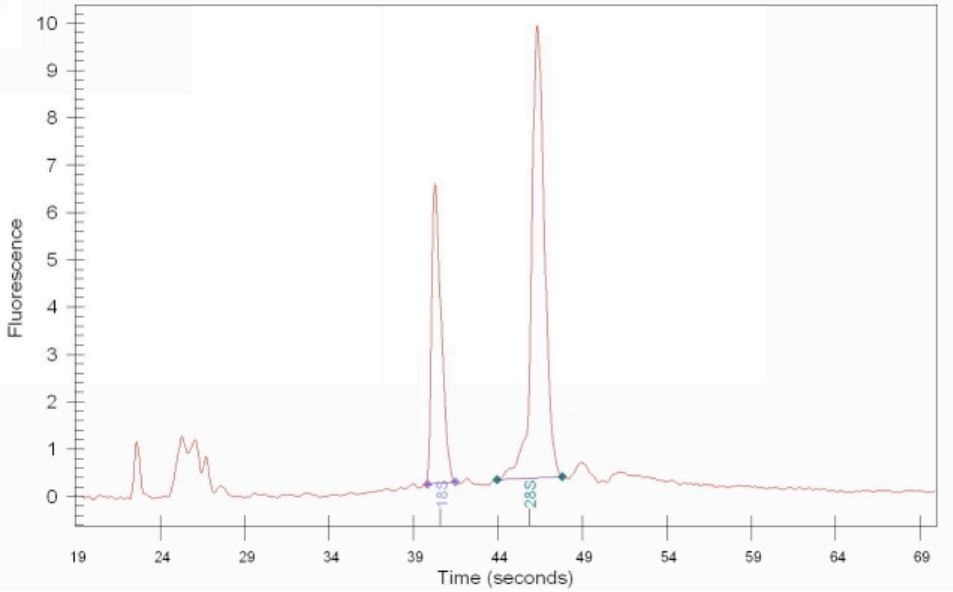


Figure 7 Intact total RNA with 28S:18S ratio of 2:1. The RNA was isolated from whole human prostate (fresh frozen – courtesy of Agilent). The 28S and 18S rRNA bands, hallmarks of good-quality total RNA, show distinct peaks in this example, at run times of about 52 s and about 44 s, respectively. (The additional peak at a run time of 24 s characterizes an alignment marker that is an integral part of the BioAnalyzer assay protocol).

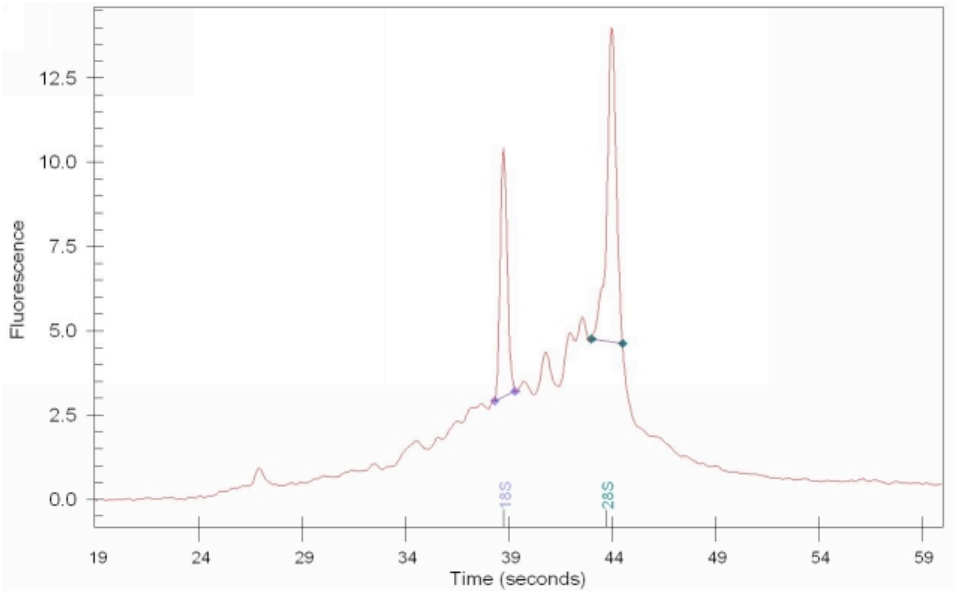


Figure 8 Partially degraded total RNA with 28S:18S ratio of 1.4:1 . Although the 28S and 18S bands are visible there is high baseline implying partial degradation of the RNA. This is reflected in the low ratio. The RNA was isolated from whole human prostate (fresh frozen – courtesy of Agilent).

The Agilent 2100 bioanalyzer (Agilent Technologies) was used to provide RNA assessment for quantification and quality control, with 2 differing chips used dependant on expected RNA amounts. The RNA pico chip was used for laser microdissected specimens and the RNA nano chip for cell cultured specimens. The RNA picochip has a qualitative range of 50-5000 pg/μL for total RNA assay, and the RNA nano chip has qualitative range of 25-500 ng/μL for total RNA assay. Both chips were used according to manufacturer's specifications. The RNA integrity number (RIN) is a software tool available on the Agilent 2100 bioanalyser designed to estimate the integrity of total RNA samples. An integrity number is automatically assigned to the total RNA sample. Sample integrity is no longer determined by the ratio of the ribosomal bands, but by the entire electrophoretic trace of the RNA sample. This includes the presence or absence of degradation products. In this way, interpretation of an electropherogram is facilitated, comparison of samples is enabled and repeatability of experiments is ensured. The assigned RIN is independent of sample concentration, instrument and analyst and is becoming a de facto standard for RNA integrity (Imbeaud *et al.*, 2005).

2.1.2 RNA from fresh and archival tissue samples

Quality of RNA is a critical issue for the investigation of gene expression from tissue samples and has a major impact upon conclusions derived from such studies (Schuchhardt *et al.*, 2000, Group, 2004). Freshly procured human tissue, snap frozen in liquid nitrogen, provides the best guarantee of RNA integrity for use in gene expression analysis using microarrays (Group, 2004). Archived pathology specimens are normally formalin-fixed and paraffin embedded (FFPE) and while useful for immuno-histopathological studies they have limited use in gene expression analysis. Formalin fixation causes potential fragmentation and chemical modification of RNA (linkage to protein and addition of CH2OH groups to amino groups of bases) (Masuda *et al.*, 1999) and is therefore an impediment in use of such samples for microarray analysis. While Affymetrix do produce the X3P array specifically designed for FFPE specimens (www.affymetrix.com), traditional global expression analysis studies have used fresh frozen tissue.

2.1.3 Factors involved in RNA degradation from tissue samples

RNA degradation is an important concern in any gene expression study design. Several studies have investigated the use of fixatives on RNA stability (Vincek *et al.*, 2003, Guo and Catchpoole, 2003, Florell *et al.*, 2001), but few have examined the effect of time and handling on tissue specimens after acquisition and before freezing / fixation. The gold standard for handling of fresh tissue specimens is immediate freezing of tissue specimens in isopenthane / liquid nitrogen and subsequent storage at -80^0C (Srinivasan *et al.*, 2002), however this is not always possible in the clinical setting. There are several commercially available fixatives, such as RNALater (Ambion), which have been shown to preserve RNA effectively for microarray studies (Mutter *et al.*, 2004, Grotzer *et al.*, 2000).

Micke *et al.* (Micke *et al.*, 2006) examined the effect of different handling conditions on RNA quality in fresh tissue specimens. They tested 4 different conditions and measured the effects on RNA quality using the Agilent Bioanalyser and quantitative Real-Time PCR. Fresh human tonsil tissue and colon were obtained from the operating theatre. Specimens were cut into cubes with the control tissues immediately frozen in isopenthane / dry ice and stored at -80^0C. The remaining samples were either stored at room temperature, on ice, in cold 0.9% saline solution, or in RNALater. At 0.5, 1, 3, 6, and 16 hours pieces were removed and stored at -80^0C.

Interestingly they found little in the way of RNA degradation at room temperature, even up to 16 hours, as evidenced by 28S:18S ribosomal ratios on the Bioanalyser. They postulated that this surprising finding was the result of intact cellular structure. To address the question of possible gene expression changes they performed Real-Time PCR on six genes (cfos, HIF1alpha, Bcl2, PCNA, TGFbeta1, and SMAD7) known to be responsive to cellular stress. These genes were most stable when stored on ice, and least stable when stored in RNALater, and they concluded that storage on ice until freezing was the optimum method.

Dash *et al.* (Dash *et al.*, 2002) examined the effect of warm ischaemia time on global gene expression in radical prostatectomy specimens. Samples were maintained at room temperature at 0, 0.5, 1, 3, and 5 hours before processing and hybridisation to cDNA microarrays. Less than 0.6% of genes showed statistically significant differential expression at 1 hour, and did not include genes previously thought to be associated with prostate carcinogenesis, such as Hepsin, AMACR, fatty acid synthase, and PTEN. There was little overall increase in the gene expression variability present with ischemia time at any of the time points. Of the genes identified showing significant differences, several have been previously identified as showing increased expression with ischaemic stress.

RNA expression represents the tissue response to insults such as ischaemia, and is sensitive to degradation, and as such these must be considered when performing gene expression studies on fresh tissue.

2.1.4 Laser Capture Microdissection

Laser capture microdissection (LCM) was developed at the National Cancer Institute at the National Institutes of Health (Bethesda, MD, USA) and was initially described by Emmert-Buck and colleagues, 1996 (Emmert-Buck *et al.*, 1996). Their technique involved placing a thin thermoplastic transparent film over a tissue section, visualizing the tissue microscopically, and selectively adhering the cells of interest to the film with a focused pulse from an infrared laser. The film, with the attached tissue was then removed and placed into a buffer for processing. They were subsequently able to demonstrate successful polymerase chain reaction amplification of DNA and RNA.

Two different laser capture microdissection techniques are now in common use. The first, named laser capture microdissection, is the technique developed by Emmert-

Buck *et al.* (Emmert-Buck *et al.*, 1994) and has been commercialized by Arcturus Bioscience Inc. – PixCell II (Mountain View, CA, USA). It is widely used (Fend and Raffeld, 2000). A tissue section is mounted on a glass slide and stained in accordance with the requirements for subsequent downstream applications. The slide is replaced on an inverted microscope and a transparent cap with thermoplastic film placed on top of the section. A low power infrared laser beam is directed at the cells from above. The laser is positioned over the cells of interest and when activated, the thermoplastic film is melted above them. When the cap is lifted the selected cells come with it and the film can then be placed into buffer for extraction of DNA, RNA, or protein (Figure 9).

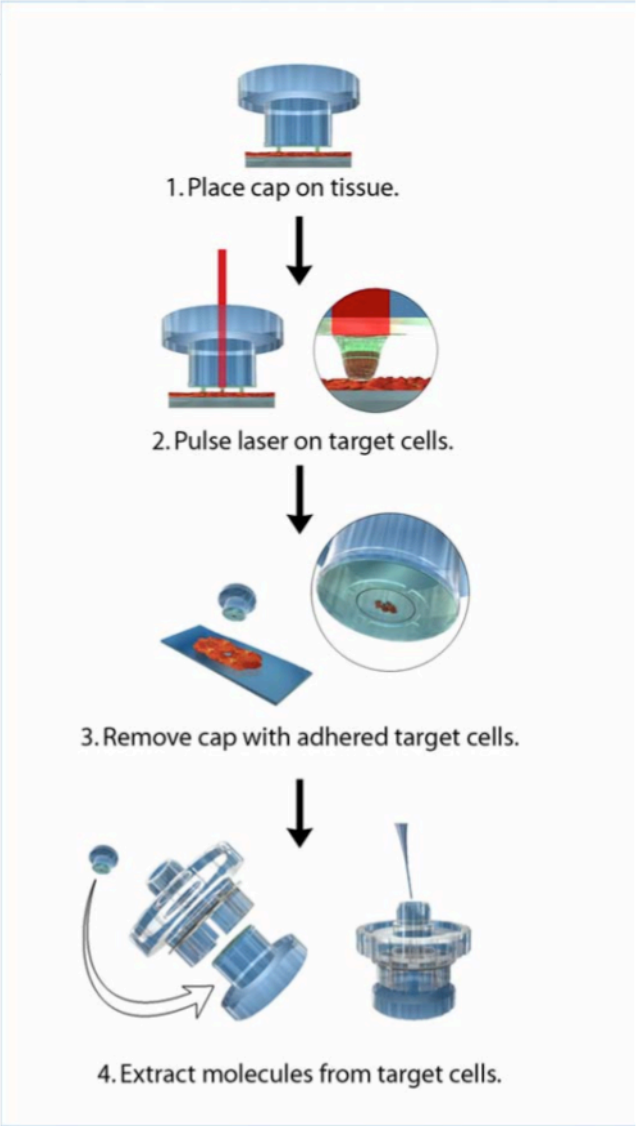


Figure 9 Laser capture microdissection (Courtesy of Arcturus)

The second technique uses two steps: laser ablation and laser pressure catapulting. This technology is made by P.A.L.M. (Positioning and Ablation with Laser Microbeams) Microlaser technologies – PALM MicroBeam (Carl Zeiss MicroImaging GmbH, Bernreid, Germany) and is known as 'laser microdissection and pressure catapulting' (LMPC). Tissue is mounted on nuclease free polyethylene naphthalate (PEN) membrane slides and stained. A highly focused laser beam (pulsed nitrogen laser – wavelength 337nm) ablates the outline of the selected cells precisely. Because the tissue is resting on a membrane, the area of interest is detached from the rest of the sample. Next a second laser is used to 'catapult' the isolated cells, using the high photonic pressure force of the laser, into an eppendorf cap placed directly above the slide, and filled with buffer. The cells are then spun down and can be used for downstream applications. As there is no contact or heat DNA, RNA, protein, and live cells can be recovered intact (Figure 10).

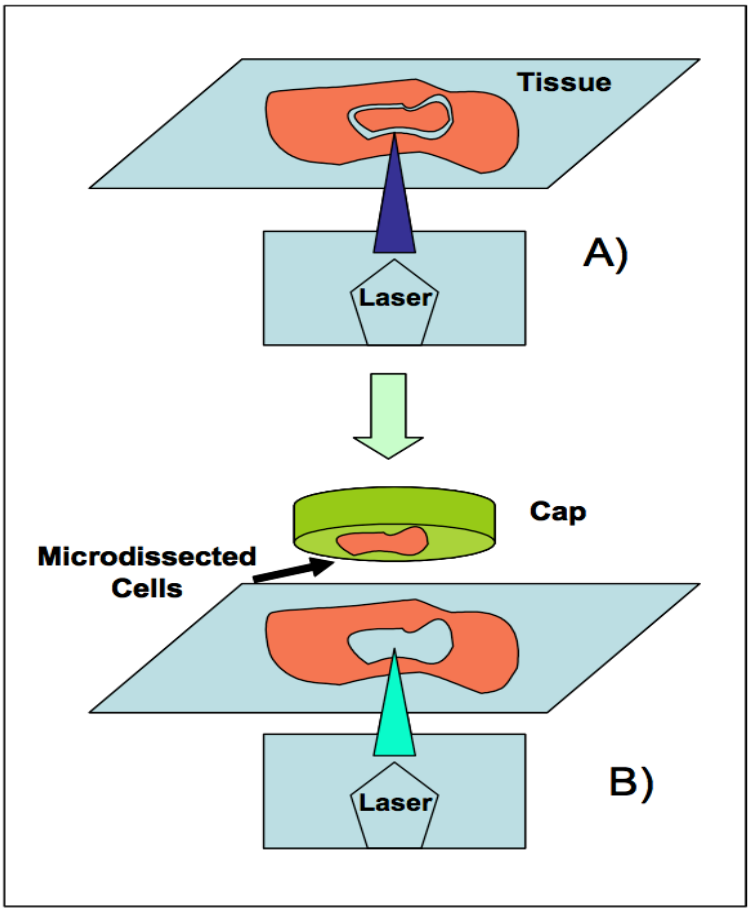


Figure 10 Laser Microdissection and Pressure Catapulting (Courtesy of PALM)

Both of these techniques offer accurate, rapid, and 'no touch' methods to obtain high quality DNA, RNA and protein, from specific cellular compartments. Which system is used depends on local availability and expertise.

Many different types of molecular analyses have been successfully performed on cells procured by LCM including RT-PCR amplification (Emmert-Buck *et al.*, 1996), loss of heterozygosity (LOH) (Takeshima *et al.*, 2001), microsatellite instability (Lu *et al.*, 2003), cDNA microarrays (Alevizos *et al.*, 2001), and proteomics (Grubb *et al.*, 2003).

2.1.5 RNA Amplification and Microarray Analysis

Expression profiling requires microgram quantities of RNA, e.g. Affymterix gene chips require 5μg of RNA (15μg cRNA) for each chip. Human tissue specimens are often small (obtained through fine needle aspiration or needle core biopsy) and display heterogeneity of cell type. For example, for secretory glands, the proportion of epithelia to corresponding mesenchyme could be 5:1 (Bartsch *et al.*, 1979, Bartsch *et al.*, 1987). If gene X is expressed at a high level only in the mesenchyme but is not expressed in the epithelial cells, it would still be represented as expressed in RNA isolated from whole tissue. Gross quantities of specimen do not therefore reflect the cell type of interest and may yield distorted expression profiles in microarray analysis. Laser capture microdissection can be used to obtain specific cellular samples, even single cells, but the amount of RNA extracted is small to use directly for microarray studies, with contemporaneous techniques.

To obtain sufficient quantity of RNA for microarray hybridisation after LCM, an amplification method is necessary. Two techniques are available. Polymerase chain reaction (PCR) results in exponential amplification and is used to quantitate individual genes. Its use for global expression profiling is limited as PCR amplifies long, GC rich sequences with less efficiency than short, AT rich sequences. This results in skewed amplification levels from gene to gene which is carried on from round to round, thus exacerbating the difference (Polz and Cavanaugh, 1998). Van Gelder and Eberwine's (Van Gelder *et al.*, 1990) method uses *in vitro* transcription (IVT) with T7 RNA polymerase resulting in amplified RNA (aRNA). This is known as linear amplification. RNA is primed for cDNA synthesis by a polyT oligonucleotide containing the 17bp sequence for the T7 RNA polymerase promoter. After second stage synthesis to cDNA the template is then transcribed by a highly

concentrated T7 polymerase, resulting in an approximate 2000-fold amplification of antisense RNA that can be used for hybridisation analysis (Phillips and Eberwine, 1996). One round of amplification is sufficient for many microarray studies, however, when dealing with very small amounts of starting material (<1-100ng) it is often necessary to perform a second, and sometimes third, round of amplification. This has enabled the transcriptome of a single cell to be analysed (Eberwine *et al.*, 1992, Cheetham *et al.*, 1997). Further rounds of amplification require the use of random primers in reverse transcribing the output RNA from the preceding round. The polyT-T7 oligonucleotide is used again to prime second strand synthesis and the resulting template is again transcribed by T7 RNA polymerase, creating a second amplification of aRNA. The amplification potential is enormous (>10^6 fold) (Feldman *et al.*, 2002). The potential drawback of more rounds of amplification comes with the use of random primers, which result in 30% shortening of the transcript after each round (Eberwine *et al.*, 1992). This results in a 3' bias to the aRNA, although the effect is limited as many microarrays are designed using the 3' sequences of genes (Luzzi *et al.*, 2003). The length of amplified RNA ranges from 200 to 6000 nucleotides for the first round of amplification and 100 to 3000 nucleotides for the second round when random primers are used (Feldman *et al.*, 2002). The fidelity of IVT has been extensively tested by gene expression analysis, real time PCR and statistical testing comparing estimates of gene expression in amplified versus non-amplified RNA (Wang *et al.*, 2000).

Luo *et al.* (Luo *et al.*, 1999) were first to demonstrate that it was possible to combine, LCM, linear based RNA amplification, and cDNA microarray analysis. They were able to compare the expression profiles of different subtypes of neuronal cells on 477-element cDNA arrays, using only 1000 neurons per sample. Later that year, Sgroi *et al.* (Sgroi *et al.*, 1999) combined LCM and cDNA arrays to profile breast cancer samples taken at different stages of disease and run on arrays with more than 8000 targets. Since then numerous studies have combined LCM and microarrays in expression profiling small highly pure, homogenous samples in a variety of different tissues (Alevizos *et al.*, 2001, Kitahara *et al.*, 2001, Ma *et al.*, 2003, Khanna *et al.*, 2004).

2.1.5 Affymetrix microarrays

Affymetrix Human Genome U133 Plus 2.0 Arrays were used in this study. They

contain roughly 54,000 different transcripts representing nearly 38,500 genes and expressed sequence tags (EST). These transcripts are made up of synthetic oligonucleotide probes and may be full length (well-curated RefSeq and Ensembl transcripts as well as Genbank transcripts annotated as "complete CDS" [CDS = coding sequence, or the portion of a gene or mRNA which actually codes for a protein]), partial (moderately supported RefSeq and Ensembl transcripts and Genbank transcripts not annotated as "complete CDS"), EST's (EST-only supported transcripts), or predicted (Computationally predicted transcripts). Hence the discrepancy in number of targets and genes seen on Affymetrix GeneChips. The Gene Microarray Centre at the Institute of Child Health, University College London, performed amplification, labeling, hybridization, washing, and scanning according to Affymetrix specifications (www.affymetrix.com).

2.1.6 Nomenclature - GeneSpring analysis

1. Gene list - A subset of genes. Analysis results are saved as gene lists.
2. Experiment - A collection of microarray samples that are analyzed together
3. Sample - Data from one microarray chip
4. Condition - A grouping of one or more samples (e.g. prostate or zone)

2.2 Materials – Fresh Normal Human Prostate

Whole prostates were collected from cadaveric organ donors at the time of transplant organ harvest. Donors were young men, aged between 15-35, whose relatives had agreed to the use of their organs for transplant and the prostate for research purposes. This group was selected, with donors post-pubescent but still young, therefore with developed prostates likely to be free of clinically silent, undiagnosed diseases. Samples were anonymised and therefore no demographic data is available. Ethical approval was obtained from the Joint UCL/UCLH Committees on the Ethics of Human Research. Six prostates were obtained as shown in Table 1.

PROSTATE	Date of retrieval
A	270800
B	200102
C	120902
D	011002
E	120104
F	301104

Table 1 shows the prostates that were collected from cadaveric organ donors at the time of transplant organ harvest. The date of retrieval is indicated and prostates are distinguished alphabetically. These prostates were kindly harvested by the North Thames transplant retrieval team.

Immediately after harvest of transplantable organs, the prostate was retrieved intact (approximately within 30- 60 min of aortic clamping) and preserved on ice until return to our institution (~240 min). Each prostate was cut into approximately 5mm slices from apex to base. Small blocks of tissue (about $5mm^3$) were visually selected from each zone and snap-frozen in liquid nitrogen with subsequent storage at –80°C. Peripheral zone was isolated from the most peripheral and posterolateral aspect of the gland. Transition zone was isolated from the area identified immediately lateral to the urethra (peri-urethral). Central zone was isolated from the base of the prostate, in an area adjacent to the ejaculatory ducts. Further blocks ($5mm^3$) were set aside for primary tissue culture, and the remaining tissue was fixed in formalin for paraffin-embedding and routine histopathological analysis. Representative slides were examined by a histopathologist (Mr Alex Freeman) to rule out prostatic diseases.

2.3 Methods

2.3.1 Cryosectioning

Each zone from each prostate was sectioned using a Leica cyrostat at -25^0C. The cryosections were performed by Calum Thomson at the University of Dundee, according to the following protocol.

1. Cryomolds were labeled and set up on crushed dry ice and cooled.

2. The cryomolds were filled two thirds with embedding media (OCT – tissue-tek) allowing the media to become cooled and viscous.
3. Frozen tissue blocks were placed in the cooled media, centered, pushed flat against the bottom of the mold, and allowed to freeze on dry ice.
4. Further embedding media was then placed in the cryomold to ensure complete coverage of tissue, which was then frozen at -80^{0}C until sectioning.
5. The cryostat was cooled to -25^{0}C and cleaned with 100% ethanol. RNase Zap® (Ambion), a solution that completely removes RNase contamination from glass and plastic surfaces, was not used at it solidifies at this temperature. (www.ambion.com/catalog/CatNum.php?AM9780)
6. The frozen OCT-embedded tissue was removed from its cryomold and attached to the metal specimen stage with OCT media, and allowed to equilibrate for 10 minutes.
7. A fresh disposable blade was used for each zone and prostate.
8. 8μm sections were prepared and transferred to RNase free PALM membrane slides (at room temperature) ensuring they lay flat on the slide without folding, curling or wrinkling (Figure 11). A maximum of 3 sections were placed on each slide.
9. Once the first sections were mounted the slide was kept in the cryostat whilst the remaining sections were mounted.
10. The slides were then stored in a clean slide box and stored on dry ice or in -80^{0}C freezer with silica desiccant. The slides were at no time allowed to thaw.
11. Sections were processed within 4 weeks of sectioning.

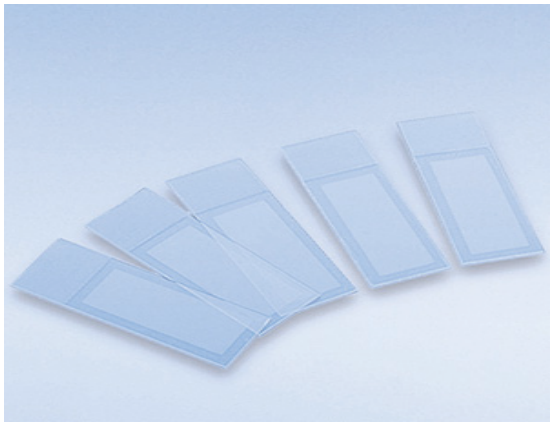


Figure 11 PEN nuclease free membrane slides – PALM

2.3.2 Staining

It is often difficult to recognize histological details after frozen sectioning and therefore a staining method is required. RNA integrity and quality is paramount for microarray studies and routine Haematoxylin and Eosin (H&E) staining will not suffice. A variety of different histological staining methods have been described including methyl green pyronin (MGP), Nissel staining (NS), immunoflourescent stains, and modified H&E, and numerous commercial kits exist. Some have been shown to significantly affect the integrity of cellular RNA with respect to microarray studies (Wang *et al.*, 2006).

We modified an H&E staining technique: (http://www.palm-microlaser.com/dasat/images/0/100270-md-faerb.pdf), (http://dir.nichd.nih.gov/lcm/LCMTAP.htm). Subsequent to H&E staining, the sample must be dehydrated to reduce RNA degradation: ethanol and xylene were used for this purpose. The advantage of chemical dehydration over vacuum desiccation is that it allows immediate microdissection. Fresh solutions were used for each zone of each prostate and only RNase free materials were used. Samples were stained immediately prior to microdissection (Figure 12). Staining method was as follows:

1. Membrane slide and section placed in 70% ethanol (ice cold 4°C) for 1 min
2. Slide placed in RNase free water for 15 sec to remove embedding medium
3. Section stained in Mayer's Hematoxylin for 15s to 1 min, depending on staining required
4. Dipped in RNase free water 10 times (15s)
5. Dipped in 70% ethanol 20 times (30s)
6. Dipped in 96% ethanol 20 times (30 s)
7. Stained in Eosin Y (alcoholic with phyloxine) for 15s
8. Dipped in 95% ethanol for 15s
9. Dipped in 95% ethanol for 15s
10. Dipped in100% ethanol for 30s
11. Dipped in 100% ethanol for 30s
12. Placed in Xylene for 1min (without agitation)
13. Placed in Xylene for 5min (without agitation)

These last 2 steps were carried out in a fume hood.

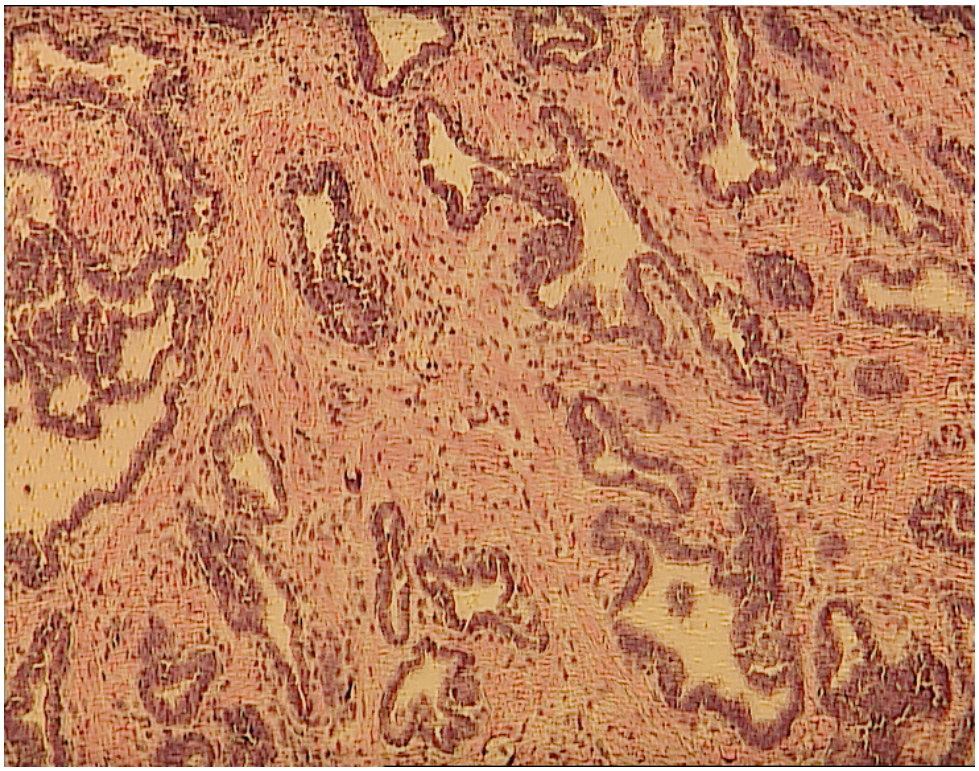


Figure 12 Prostate D peripheral zone after cryosectioning and modified H&E staining. The prostatic acini and stroma are distinguishable at 20x magnification.

Preliminary experiments were performed to investigate whether small amounts of RNA could be obtained using this sectioning / staining protocol. To do this, four sections from each prostate were stained, scrapped and pooled into RLT buffer for RNA extraction. RNA was isolated using the RNeasy® Mini Kit [Qiagen, Basel, Switzerland] and eluted into 30μL RNase free water and stored at -80^{0}C. Quality and quantity was assessed using the Agilent bioanalzer system (Agilent Technologies). Results show that good quality and quantity (nanograms) RNA could be isolated (Figure 13).

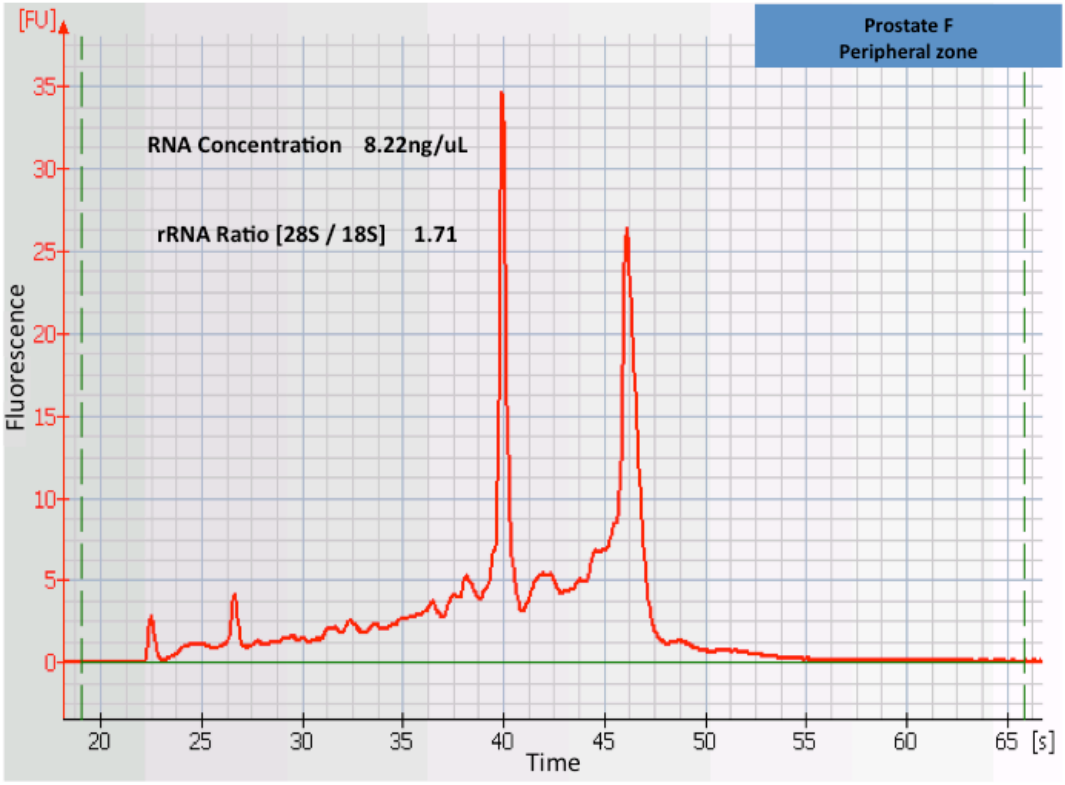


Figure 13 Electropherogram of pooled peripheral zone RNA (stromal and epithelial) from prostate F.

2.3.3 Laser Capture Microdissection

The PALM MicroBeam system (Figure 14), based at St Georges Hospital Medical School, London, was used to perform LMPC on each zone from the three prostates. Following staining the sections were allowed to dry for 5 minutes before being mounted on PALM PEN (polyethylene naphthalate) slidesMicroBeam system. PALM RoboSoftware version 4.0 was used to draw around individual prostatic acini, including basal and luminal cells, with the automated software then used to cut around and catapult the sample into a RNase free eppendorf containing 20µl RLT buffer (RNeasy® Mini Kit [Qiagen, Basel, Switzerland]) (Figure 15). Microdissection times were kept less than 20 minutes, from the start of the H&E staining and dehydration protocol to storage of laser captured sample, to minimize RNA degradation. Approximately 10 – 20 acini of differing sizes were collected in each 20-minute session. Each acinus was estimated by manual counting to contain ~ 200 cells (up to 4000 cells harvested in each session). After brief centrifugation (1000x*g*) in a

benchtop Eppendorf centrifuge at 4^{o}C, the samples were stored on dry ice until return to the laboratory.

Figure 14 PALM MicroBeam based at the Medical Biomics Centre, St Georges Hospital Medical School, London

Laser microdissection was carried out successfully for all 3 zones of prostate D and E, however the transition zone of prostate F was found to contain no epithelial acini and no epithelial sample from this zone could be isolated.

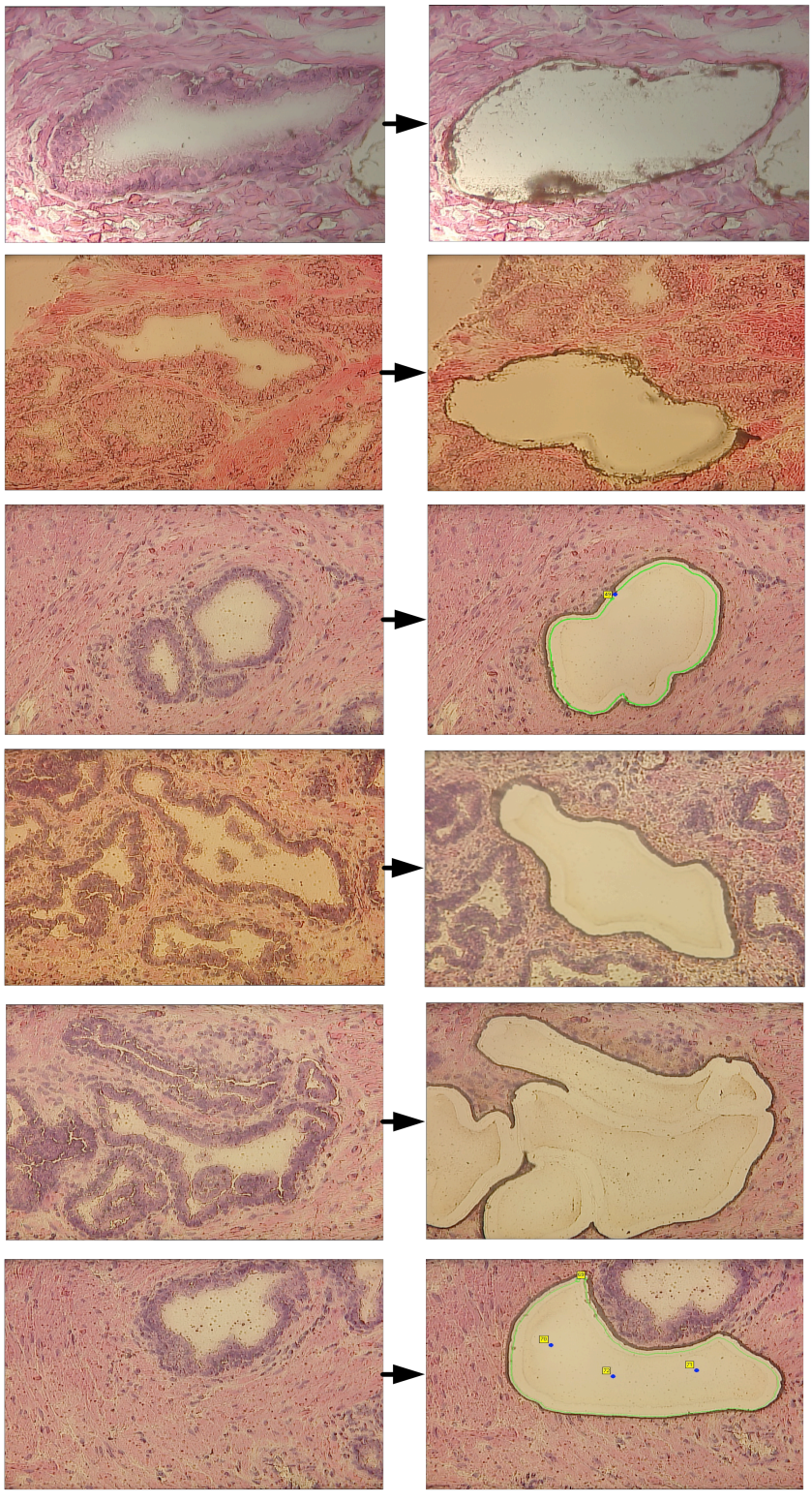


Figure 15 Representative images are shown pre and post laser capture microdissection and pressure catapulting using the PALM microbeam system. Dissection of epithelium and adjacent stroma is shown.

2.3.4 RNA Isolation

RNA isolation was carried out using the RNeasy® Micro Kit [Qiagen, Basel, Switzerland]. The final sample was eluted into 12µL of RNase free water and stored at -80^{0}C. One microlitre was used to assess RNA quality and quantity on Agilent pico chips. Representative electropherograms are shown in Figure 16. Table 2 and Figure 17 demonstrate the quantity and quality of RNA that was isolated.

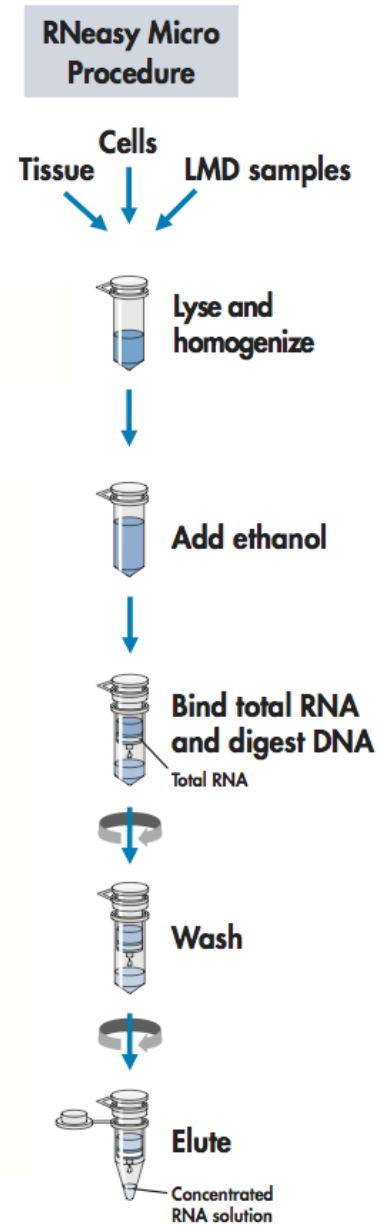


Figure 16 RNeasy® Micro procedure, that combines the selective binding properties of a silica based membrane with guanidine-thiocyanate-containing lysis buffer, to promote selective binding of RNA. (www.qiagen.com/hb/RNeasyMicro)

Prostate	Zone	RNA concentration	Total RNA amount	A260/280 Ratio	RIN Number
E (120104)	CZ	745pg/µL	8.94ng	1.2	6.8
E (120104)	TZ	4162pg/µL	49.94ng	0.7	6.1
F (301104)	PZ	2856pg/µL	34.27ng	1.7	8.3
F (301104)	CZ	2629pg/µL	31.55ng	0.9	7
E (120104)	PZ	147pg/µL	1.76ng	1.1	8
D (011002)	TZ	2008pg/µL	24.1ng	0.8	7.2
D (011002)	CZ	3126pg/µL	37.51ng	0.7	6.8
D (011002)	PZ	710pg/µL	8.52ng	0.7	6.3

Table 2 The Agilent Bioanalyser was used to assess RNA quality and quantity, which are shown. Low nanogram quantities were isolated, with RIN numbers (section 3.4) sufficient for subsequent microarray analyses (1 degraded – 10 totally intact)

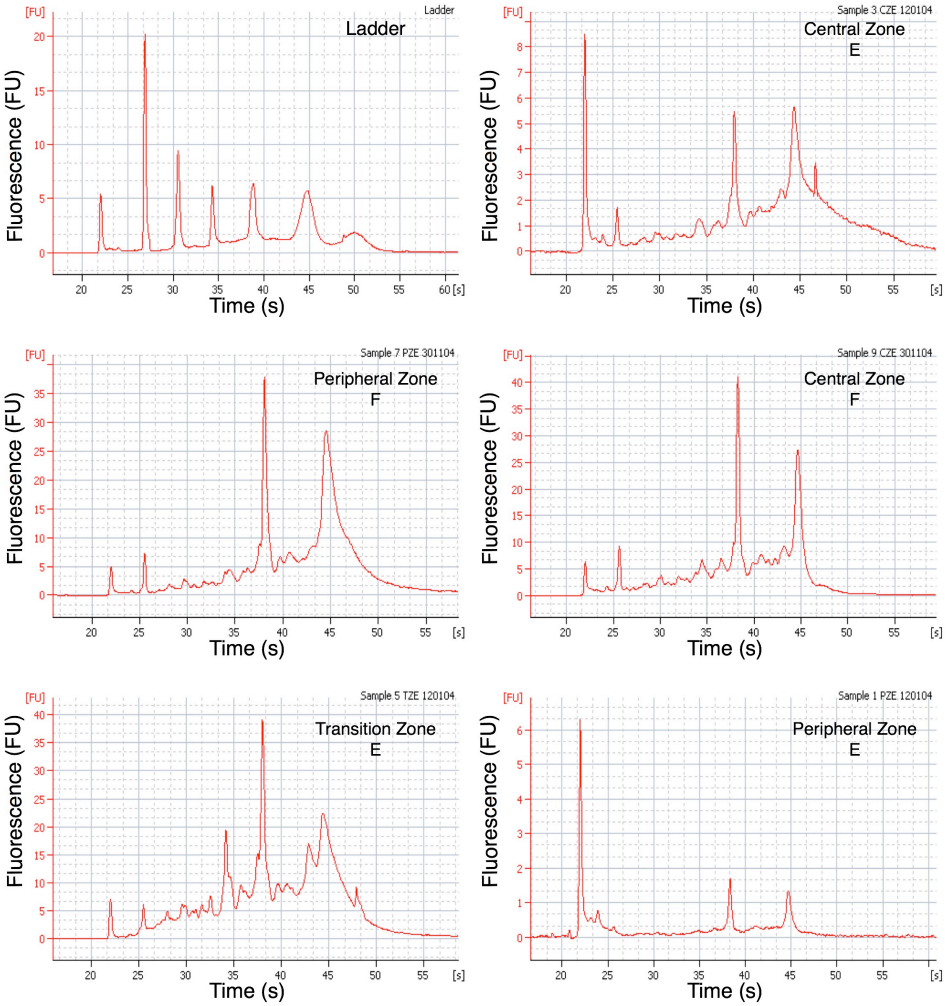


Figure 17 RNA quality and quantity was analysed using Agilent RNA 6000 Chips on an Agilent 2100 Bioanalyzer. The quality and quantity of RNA is shown in Table 3.1. RNA ladder and representative electropherograms are shown demonstrating the integrity of RNA.

2.3.5 RNA amplification

RNA was amplified using MessageAmp^{TM} II aRNA Amplification Kit (Figures 18, 19) based on the RNA amplification protocol developed in the Eberwine laboratory (Van Gelder *et al.*, 1990). Two rounds of amplification were performed and amplified RNA was assessed using the Agilent Bioanalyzer (Figures 20, 22). The results indicate that RNA could be amplified at least 10-fold to yield sufficient quantities (over 15μ g) of amplified RNA for oligoarray and real-time PCR experiments. A similar approach has been previously used in other tissues to determine gene expression profiles (King *et al.*, 2005, Stanbrough *et al.*, 2006).

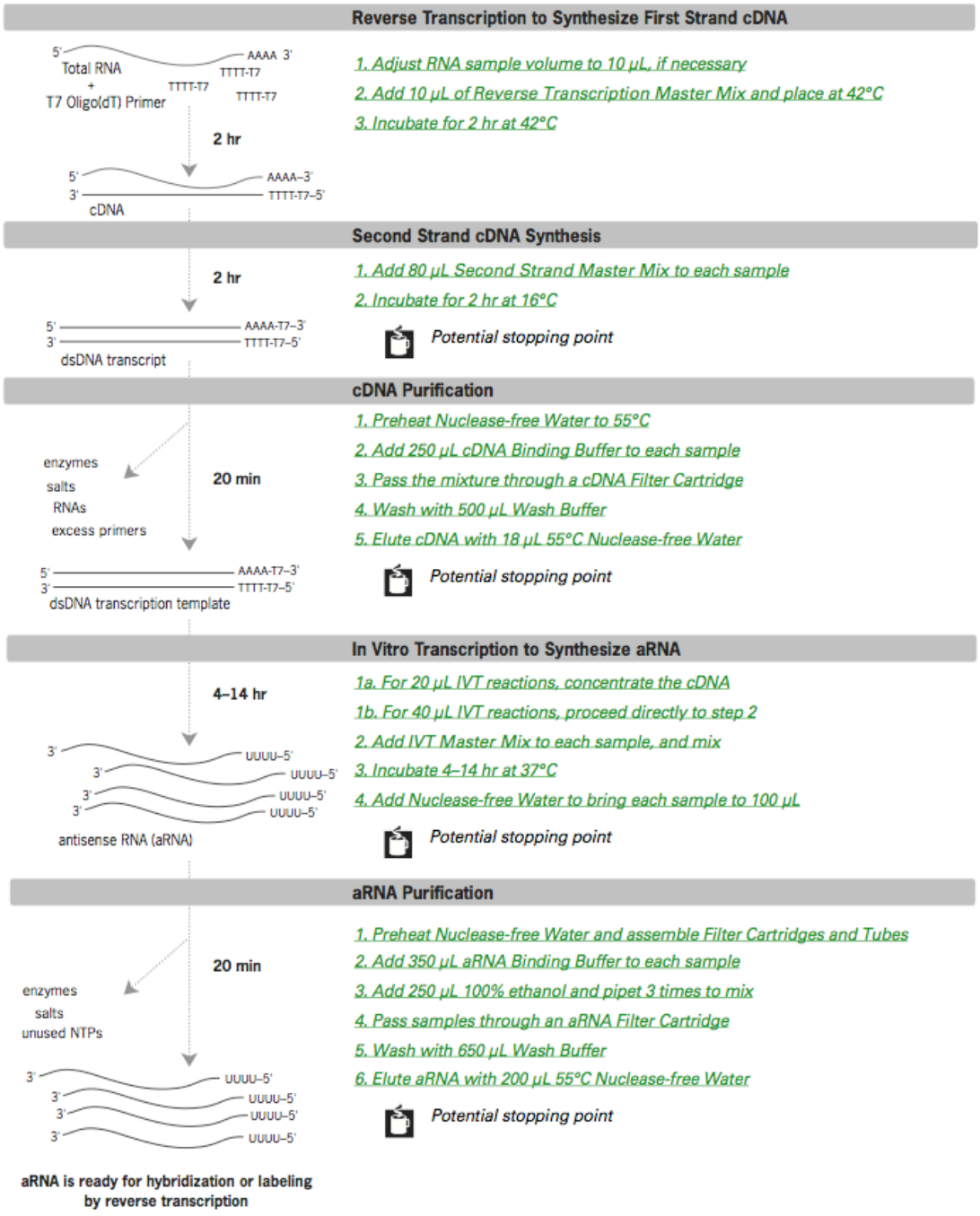


Figure 18 MessageAmp™ II aRNA Amplification Procedure (1st round). This scheme is taken from the MessageAmp™ II Protocol (http://www.ambion.com/techlib/prot/fm_1751.pdf)

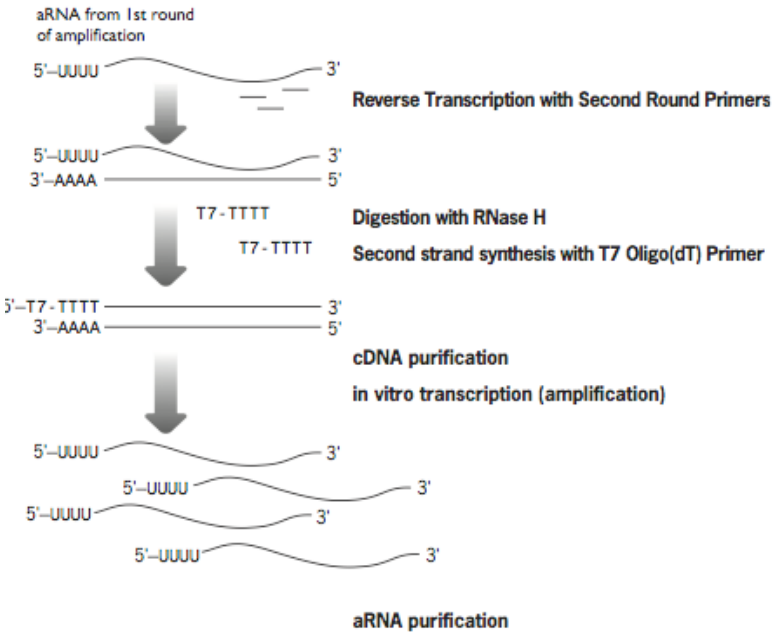


Figure 19 MessageAmp™ II aRNA second round amplification. (http://www.ambion.com/techlib/prot/fm_1751.pdf)

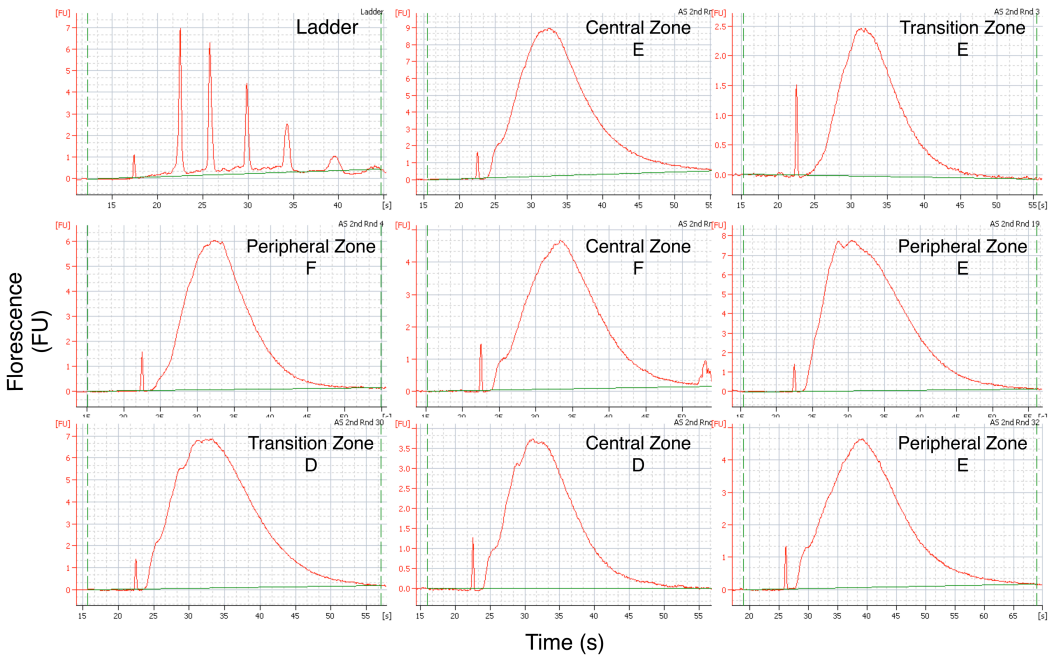


Figure 20 Electropherograms are shown demonstrating successful RNA amplification after one round with MessageAmp™ II aRNA. The expected aRNA profile should be a distribution of sizes 250–5500 nucleotides (nt) with most of the aRNA between 1000–1500 nt (Figure 21).

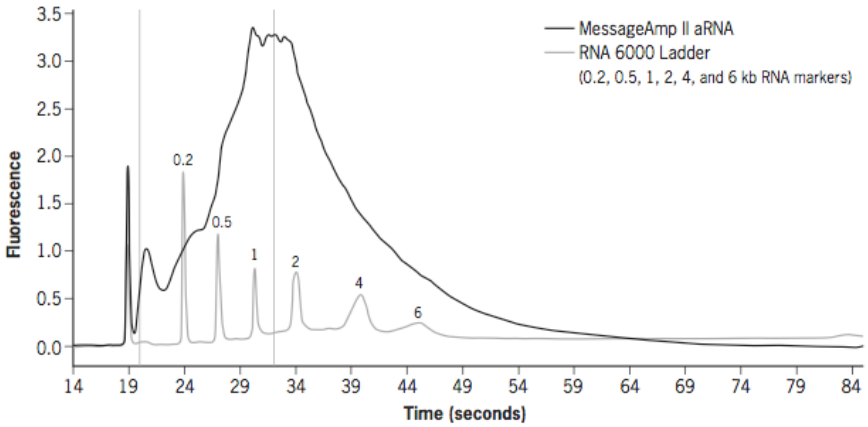


Figure 21 This electropherogram displays the aRNA size distribution from 1 round of amplification of 1 µg of Control RNA using biotin labeled NTPs in a 40 µL IVT reaction that was incubated for 14 hr. Using the Agilent bioanalyzer mRNA smear assay, 50% of the aRNA is calculated to be larger than 1470 nt. (The vertical grey lines mark 65 and 1470 bases. The area outside these lines represents 50% of the area under the curve produced by the product of the positive control reaction.) (www.ambion.com/techlib/prot/fm_1751.pdf)

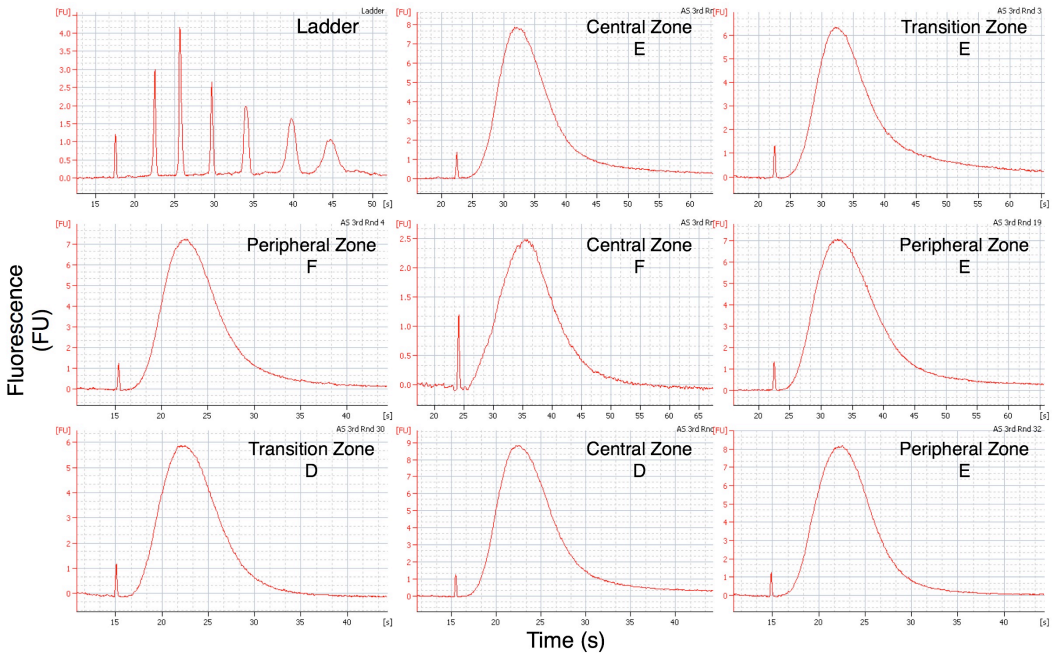


Figure 22 Electropherograms show successful RNA amplification after second round with MessageAmp™ II aRNA. Second round amplification products are typically shorter than first round amplification products. (www.ambion.com/techlib/prot/fm_1751.pdf)

2.3.6 Affymetrix microarray hybridisation

The Gene Microarray Centre at the Institute of Child Health, University College London, performed amplification, labeling, hybridization, washing, and scanning according to Affymetrix specifications (www.affymetrix.com). Briefly biotinylated cRNA (20 μg) was fragmented and added to a hybridization mixture. Affymetrix U133 plus 2.0 GeneChip arrays (Affymetrix, Santa Clara, CA) containing ~54000 known human transcripts and ESTs, were used to create expression profiles. Hybridization to these chips was performed overnight at 45^{o}C for 16h using the GeneChip Hybridization Oven 640 (Affymetrix). Washing and staining (streptavidin–phycoerythrin) was done in the GeneChip Fluidics Station 400 (Affymetrix). Images were acquired using the Affymetrix probe array scanner. PC, Affymetrix chip hybridization. Twenty micrograms of cRNA were fragmented by incubating in a buffer containing 200mmol/L Tris-acetate, pH 8.1, 500 mmol/L KOAc, and 150 mmol/L MgOAc at 95°C for 35 minutes. The fragmented cRNA were then hybridized with a pre-equilibrated Affymetrix chip at 45°C for 14 to 16 hours. After the hybridization cocktails were removed, the chips were then washed in a fluidic station with low-stringency buffer (sodium chloride, sodium phosphate dibasic, and EDTA; 0.01% Tween 20; 0.005% antifoam) for 10 cycles (two mixes/cycle), and stringent buffer (100 mmol/L MES, 0.1 M NaCl and 0.01% Tween 20) for four cycles (15 mixes/cycle), and stained with Strepto-avidin Phycoerythrin (SAPE; Molecular Probe, Eugene, OR). This was followed by incubation with biotinylated mouse antiavidin antibody, and restained with SAPE. The chips were scanned in a GeneChip Scanner 3000 (Affymetrix Inc) to detect hybridization signals.

2.3.7 Quality control and normalisation

There are many different techniques and steps to perform quality control (Q.C.) when running Affymetrix chips. Chip hybridization Q.C. involves 5 steps:

1. Validate RNA Quality using Agilent Bioanalyser (see 3.2.4, 3.2.5)
2. Determine quantity of cRNA (performed at Institute of Child Health, UCL). Good hybridization signals require approximately 20μg of labeled probe.
3. Confirm size of fragmented cRNA. Probe fragmentation results in better hybridization to oligonucleotide arrays. Assessed on Agilent Bioanalyser (Institute of Child Health, UCL).
4. Confirm hybridization quality using control sequences on test array (optional)

5. Confirm hybridization quality using control sequences on species array. GeneChip arrays contain sets of PM and MM oligonucleotides complementary to the 5' and 3' regions of housekeeping genes. Good cRNA probes hybridize to both oligo sets from the same gene yielding 3'/5' signal ratios between 1.0 and 3.0. They also generate background fluorescence of less than 100 units and detect the presence of 100 pM CreX, 25 pM BioD, 5 pM BioC and often 1.5 pM BioB in the hybridization cocktail (Figure 3.16 (A)). (BioB, bioC, and bioD are genes of the biotin synthesis pathway from the bacteria E. coli, and cre is the recombinase gene from P1 bacteriophage. A ready-prepared mixture of these biotinylated controls at staggered concentrations can be added with labeled eukaryotic cRNA samples to hybridize onto GeneChip probe arrays. Signal intensities obtained on these genes provide information on how well the hybridization, washing, and staining procedures have performed).

Examination of Chip images is also an important step in Q.C. Data Analysis Array images (.dat files) were digitized by using MAS version 5 (Affymetrix). DAT images generated by Affymetrix probe array scanner were assessed for quality control (Figure 23). A small spot was seen on the chip for transition zone prostate E, reflecting a possible design flaw with the chip. Care was therefore taken when considering the downstream analysis of this sample.

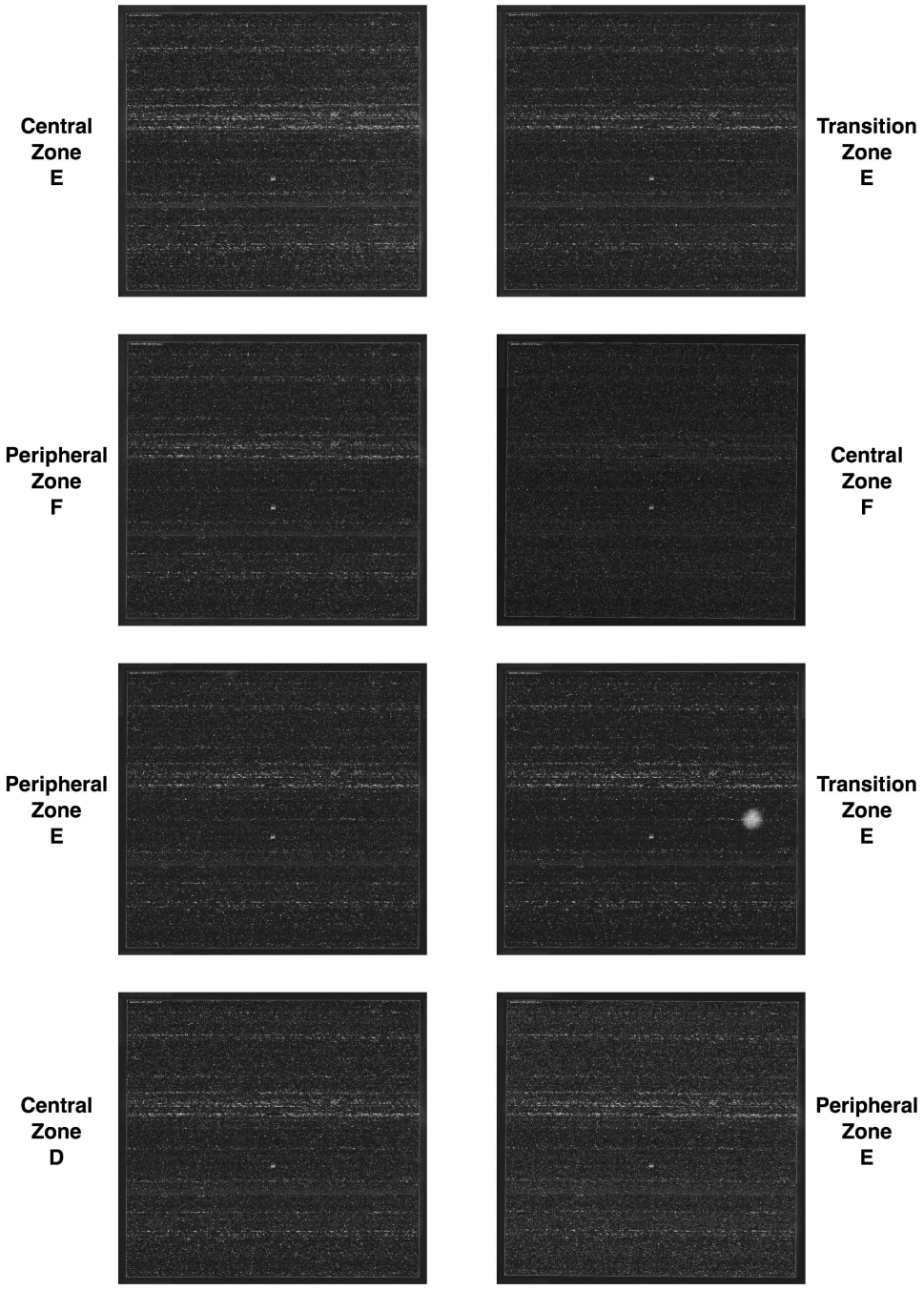


Figure 23 The raw image data from the Affymetrix probe array scanner (.dat files). Note the small defect, seen here as a white spot, on transition zone E chip. This was discussed with Affymetrix who felt it represented a design flaw.

RNA is degraded from the 5' end of a sequence; therefore intensities of probes at the 3' end of a probe set are higher than those at the 5' end. An RNA degradation plot shows the mean intensity for each position within probesets. High slopes indicate degradation, however more important than the slope is the agreement between arrays.

Figure 24 (B) shows equivalence in RNA degradation between arrays.

Gene expression values were generated from the resulting raw numerical data (.cel files) by Affymetrix GCOS publishing software. CEL files were imported into GeneSpring 7.2 (Silicon Genetics) and data was normalized using GCRMA (see Chapter 2.4.1). An additional normalisation step was taken (per gene: normalize to median) ensuring the expression value for each gene across the different conditions was centered on 1. This was achieved in GeneSpring by dividing the expression value by the median value of the expression values for that gene across the conditions (prostates). This ensures that genes that do not change across conditions get a normalized expression value of 1, allowing for easy visual detection of differentially expressed genes. The resultant Gene List generated (a group of genes with a common property) forms the starting point for further downstream analysis, with the formation of subsequent lists. After normalization the gene list contained 54,675 genes (total number of probes available on HG-U133 Plus 2.0 array).

Final quality control involved generation of Box and MvA plots (Figures 24 C, 25) within GeneSpring, enabling construction of a correlation coefficients table (Table 3). The correlation coefficients were calculated for all pair-wise comparisons of the samples in the experiment. Samples representing the same experiment condition should be more similar to each other than to samples representing a different experimental condition

Sample	D PZ	D CZ	D TZ	E PZ	E CZ	E TZ	F PZ	F CZ
D PZ	1.00	0.95	0.93	0.92	0.95	0.96	0.93	0.92
D CZ	0.95	1.00	0.98	0.93	0.96	0.97	0.96	0.93
D TZ	0.93	0.98	1.00	0.93	0.96	0.95	0.94	0.91
E PZ	0.92	0.93	0.93	1.00	0.95	0.92	0.01	0.87
E CZ	0.95	0.96	0.96	0.95	1.00	0.96	0.94	0.92
E TZ	0.96	0.97	0.95	0.92	0.96	1.00	0.94	0.92
F PZ	0.93	0.96	0.94	0.91	0.94	0.94	1.00	0.96
F CZ	0.92	0.93	0.91	0.87	0.92	0.92	0.96	1.00

Table 3 Correlation coefficients table as part of quality control showing good concordance between chips.

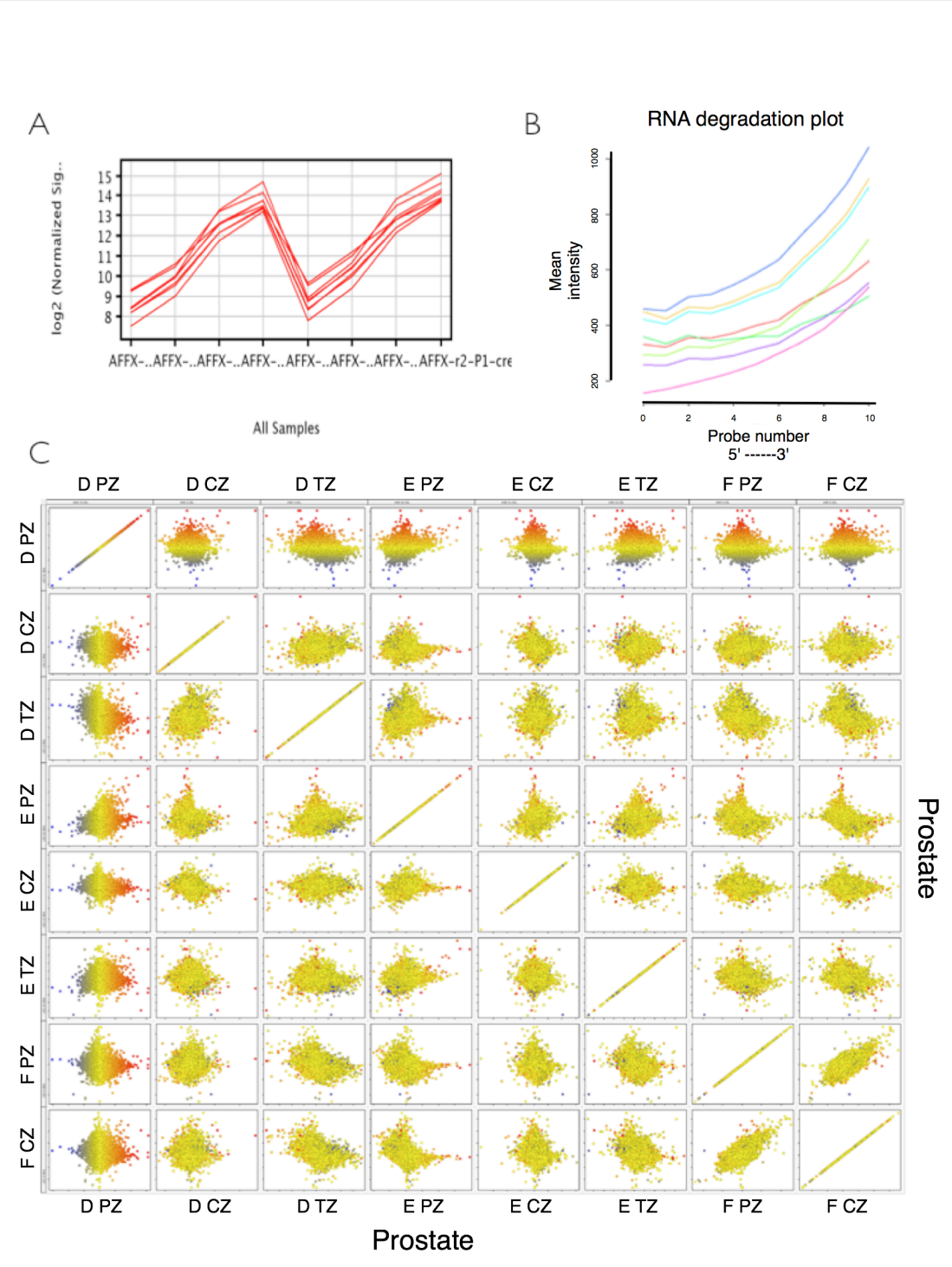


Figure 24 Quality control analysis was performed in GeneSpring 7.2 demonstrating A: equal hybridisation controls, B: RNA degradation plots and C: MvA plots. RNA degradation plot shows the mean expression from the 5' to the 3' end of the mRNA. Every chip is represented with a single line. In an ideal situation the lines would be flat, but most important is that the slopes and profiles should be as similar as possible. B shows similar curves amongst samples.

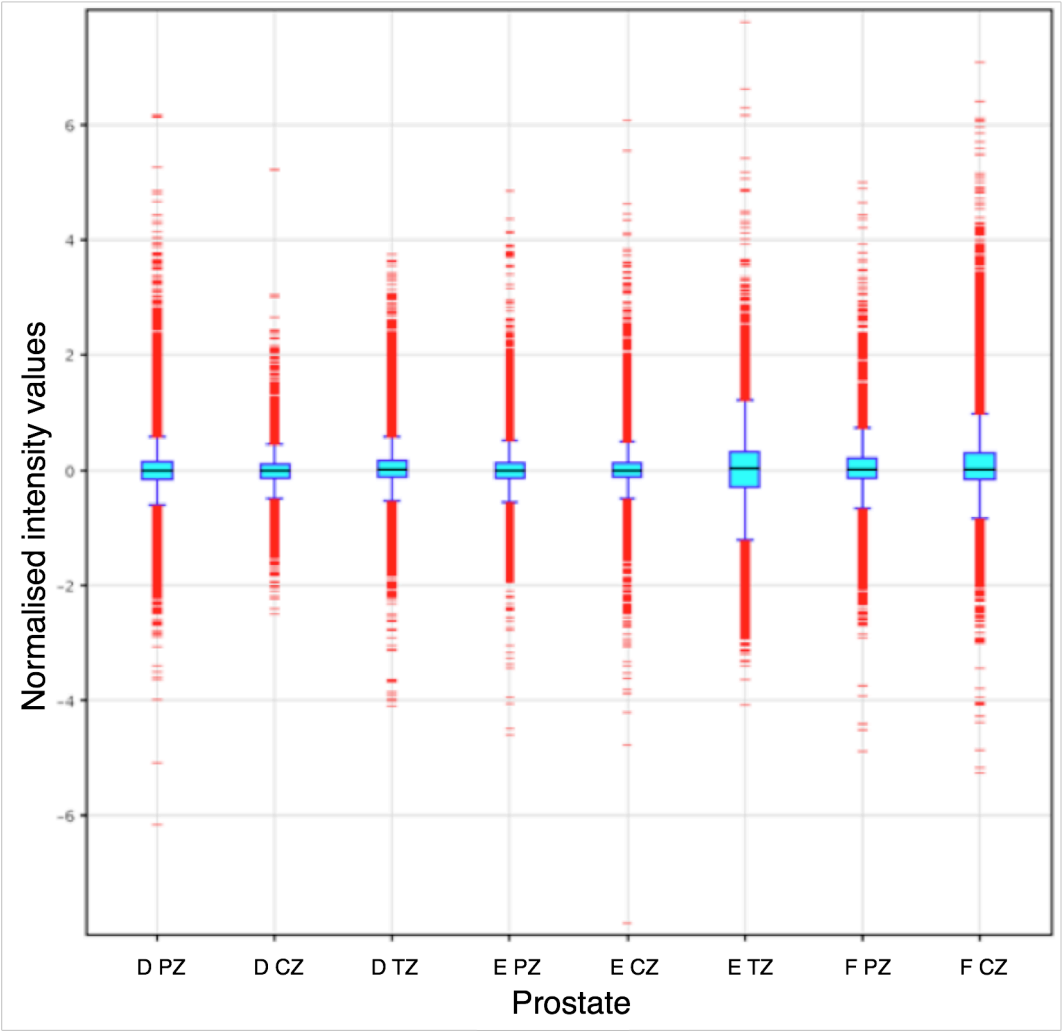


Figure 25 Box whisker plot showing distribution of normalized expression data

2.4 Statistical Analysis – Method

2.4.1 Normalisation

All statistical analysis was performed by myself using GeneSpring software. A variety of probe summarization algorithms exist (MAS 5.0, perfect match only, RMA, GCRMA) which perform three tasks: background correction, normalisation, and probe summarization (the conversion of probe level values to probeset expression values in a robust manner). These, along with other statistical methods, are discussed at length in the Appendix. The GCRMA method (GeneSpring 7.2 Addendum) was used in this study. GCRMA (Robust Multi-chip Average, with GC-content background correction) is a method of normalizing and summarizing probe-level intensity measurements from Affymetrix GeneChips. Starting with the probe-level data from a set of GeneChips, the perfect match (PM) values are background-corrected, normalized and finally summarized resulting in a set of expression

measures. The three steps of the process are outlined below (GCRMA Probe Summarization).
(http://www.chem.agilent.com/cag/bsp/products/gsgx/downloads/pdf/GCRMA_Probe_Summarization.pdf)

Background Correction: The background correction used in GCRMA is designed to account for background noise, as well as non-specific binding. Probe affinity is modeled as a sum of position-dependent base effects, and can thus be calculated for each PM and MM value, based on its corresponding sequence information. The correction is motivated by the assumptions that observed PM and MM values consist of optical noise, non-specific binding noise, and signal. Optical noise is assumed to be normal, and logged non-specific binding noise from PM-MM pairs assumed to be bi-variate normal. Using the data on a single array, the corresponding model parameters can be estimated. Each PM value is then adjusted by subtracting a shrunken MM value that has been corrected for its affinity.

Normalization: Normalization is necessary so that multiple chips can be compared to each other, and analyzed together. It is motivated by the assumption that all n chips should have approximately the same distribution of PM values. The normalization used in RMA is quantile normalization. This is a generalisation of the idea behind quantile-quantile plots to more than two dimensions. The quantiles for each PM value are plotted in n dimensions, and projected onto the diagonal. The final result is that the PM values on each chip will have the same distribution.

Summarization: Once the probe-level PM values have been background-corrected and normalized, they need to be summarized into expression measures, so that the result is a single expression measure per probe-set, per chip. The summarization used is motivated by the assumption that observed log-transformed PM values follow a linear additive model containing a probe affinity effect, a gene specific effect (the expression level) and an error term. For RMA, the probe affinity effects are assumed to sum to zero, and the gene effect (expression level) is estimated using median polishing. Median polishing is a robust model fitting technique that protects against outlier probes.

2.4.2 Filtering

Filtering of the dataset was performed using fold change as a discriminator, and was performed in GeneSpring. Although this is an arbitrary cut off it serves to reduce the

dimensionality of the dataset thus making downstream analysis more manageable.

2.4.3 Unsupervised analysis

Hierarchical clustering was performed using the 'centroid' clustering method to generate gene and condition trees using GeneSpring (Silicon Genetics – Agilent). In this method the distance between 2 clusters is the distance between the averages of the data points under one branch and the averages of the data points under another.

Principal components analysis is a method that reduces data dimensionality by performing a covariance analysis between factors. Principal components analysis produces a set expression patterns known as principal components. Linear combinations of these patterns can be assembled to represent the behavior of all of the genes in a given data set. PCA was applied on conditions to explore potential correlations between samples.

2.4.4 Supervised analysis

Welsh t-test with p value cut off of 0.05 was used calculate differential gene expression. The Benjamini-Hochberg false discovery rate was applied to the data (Benjamini and Hochberg, 1995).

2.4.5 Interpreting biological meaning

Expression microarray experiments usually result in large gene lists of significance to biological conditions. These may range in size from several hundreds to thousands of genes. Analysing these large gene lists is important to understand biological meaning, however it represents an extremely challenging task. Several publicly available tools exist to aid researchers in performing these tasks including GSEA (Subramanian *et al.*, 2007, Subramanian *et al.*, 2005), Onto-express (Khatri *et al.*, 2004), GoMiner (Zeeberg *et al.*, 2005) and DAVID (Dennis *et al.*, 2003, Hosack *et al.*, 2003). They all adopt a common core strategy to systematically map a large number of interesting genes in a list to the associated biological annotation (e.g., gene ontology terms), and then statistically highlight the most overrepresented (enriched) biological annotation out of thousands of linked terms and contents. Each tool has distinct features and strengths (Khatri and Draghici, 2005).

I used a web-based software, DAVID 2008 (Database for Annotation, Visualization and Integrated Discovery, http://david.abcc.ncifcrf.gov/ provided by the National

Institute of Allergy and Infectious Diseases, National Institutes of Health (Bethesda, MD, USA)) to provide functional annotation tools to help understand the biological meaning in our differentially expressed genes (Dennis *et al.*, 2003). Based on the Fishers Exact test, DAVID sorts the large list of genes generated from typical microarray experiments into the categories from dozens of annotation systems such as Genbank accession, OMIM, KEGG pathway, GeneOntology, and functional annotation clustering etc. DAVID incorporates enrichment analysis, the principle foundation of which is that if a biological process is abnormal in a given study, the co-functioning genes should have a higher potential (enriched) to be selected as a relevant group by the high-throughput screening technologies. Huang *et al.* 2009, provide an excellent review / overview of DAVID, its methodology, statistics, and applications (Huang da *et al.*, 2009).

Two functions of DAVID were used in this study, Functional Annotation Clustering and pathway analysis.

Functional Annotation Clustering: Functional Annotation Clustering groups and displays similar annotations / genes together. The grouping algorithm is based on the hypothesis that similar annotations should have similar gene members. DAVID uses an agglomeration method, which groups related genes or terms into functional groups (biological modules) based on the similarity distance measure. A gene or term may participate in more than one functional group, unlike other clustering techniques such as Hierarchical, K-means, or self-organizing maps. Clusters (functional gene groups) are produced which may be viewed as a heatmap and to determine which are more significant an 'enrichment score' is allocated (Huang da *et al.*, 2007). The enrichment score is the geometric mean of the EASE Scores (Fischers exact test) associated with each enriched annotation term that belonging to the cluster. The geometric mean is a relative score instead of an absolute p value, therefore minus log transformation is applied on the geometric mean. The group enrichment scores are intended to order the relative importance of the gene groups instead of as absolute decision values. A higher score for a group indicates that the group members are involved in more important (enriched) roles. However, all gene groups are potentially interesting despite lower rankings. An enrichment score of less than 0.05 translates to 1.3 on the minus log scale.

Pathway Analysis: DAVID was also used to input significantly over represented genes (Fischer's test) on to static pathway maps generated by Kyoto Encyclopedia of

Genes and Genomes (KEGG). The KEGG PATHWAY database (http://www.genome.ad.jp/kegg/pathway.html) is a collection of manually drawn pathway maps representing current knowledge on the molecular interaction and reaction networks for: metabolism, genetic information processing, environmental information processing, cellular processes, human diseases, and drug development.

2.5 Validation - Quantitative Real-Time PCR using novel technology

2.5.1 Background and principles

Quantitative reverse transcription polymerase chain reaction with real time monitoring (qRT-PCR) is the most sensitive and reproducible method to quantify messenger RNA abundance (Canales *et al.*, 2006). mRNA transcripts are reverse transcribed into cDNA using oligo(dt), random or gene specific primers, and the cDNA's of interest are then exponentially amplified by PCR using gene specific primers. The concentration of amplicon in the reaction is monitored with fluorophore-conjugated hybridisation probes or DNA-intercalating dyes (reporter dye), which are released by the polymerase enzyme.

Validation of microarray datasets at an RNA, protein and functional level is necessary to control for false discovery and to ensure that gene expression translates into tissue effects. A number of factors, including array-array variability, statistical analysis of microarray data, inherent false positive and negative rates arising from laboratory processes, and general limits of fluorescent microarray technology often lead to results that fail validation by downstream techniques. qRT-PCR has become a bedrock of microarray studies and validation, and is an integral part of the process. High throughput profiling / validation of many genes with qRT-PCR is often not practical, as it is extremely labour intensive, expensive, requires high technical skill and there may be intra / inter-assay variation. Recently Applied Biosystems developed a novel qRT-PCR technique using microfluidic technology – the TaqMan® Low Density Array (https://www2.appliedbiosystems.com/), which circumvents these problems.

The TaqMan® LDA is a low to medium throughput microfluidic card (array) that allows for 1 – 8 samples to be run in parallel against 12 – 384 TaqMan® gene expression assay targets, that are pre-loaded into each of the wells on the card. Hundreds of user specified real-time PCR reactions can be performed simultaneously.

LDA's require minimal amounts of sample and are simple to load and run on the Applied Biosystems 7900HT Fast Real-Time PCR system. In addition they require no optimization and are highly reproducible.
(http://www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/general documents/cms_040595.pdf). LDA's have been used successfully to examine gene expression profiles in a variety of tissues (Tenedini *et al.*, 2004, Marionneau *et al.*, 2005, Abruzzo *et al.*, 2005, Jost-Albrecht and Hofstetter, 2006).

2.5.2 Method

Fluorescent real-time PCR (TaqMan, Applied Biosystems) was used to confirm differences in gene expression. 300ng of amplified RNA (set aside after amplification, and before biotinylation – Affymetrix) was reverse transcribed using Omniscript Reverse Transcriptase (Qiagen) with 10μM random nonamers (Stratagene) and 1μM Oligo-dT primers (Qiagen). TaqMan probes and primers for real-time PCR were purchased as a pre-developed assay system (ABI) and GAPDH was used as endogenous control. Supplemental file chapter 2 shows the genes chosen with corresponding Applied Biosystems assay ID. Quantitative PCR was performed using ABI Prism 7900 with microfluidic cards (Applied Biosystems) according to manufacturer's protocols, with each reaction containing 6ng of reverse transcribed RNA in 2 μl reaction mix. The following cycling parameters were employed: 48°C for 30 min, 95°C for 10 min, followed by 40 cycles of 95°C for 15 sec and 60°C for 15 sec. Each sample was tested in quadruplicate and results were analyzed using sequence detector software (SDS v2.2 – Applied Biosystems). Relative quantitation was performed using the $2^{-\Delta\Delta}C_T$ method (Lival and Schmittgen 2001).

2.5.3 Relative Quantification using the Comparative C_T Method

Two common methods are used to analyze data from qRT-PCR. The absolute quantification method determines the input copy number by relating the PCR signal to a standard curve. Relative quantification relates the PCR signal of the target transcript in a treatment group to that of another sample such as an untreated group. The $2^{-\Delta\Delta}C_T$ method is a way of analyzing the relative changes in gene expression from qRT-PCR experiments (Livak and Schmittgen, 2001). This second comparative C_T method was used in this thesis. This involved comparing the C_t values of the samples of interest

against each other. The C_t values of these were then normalized to an appropriate endogenous housekeeping gene (GAPDH). The comparative C_t method is also known as the $2^{-[delta][delta]Ct}$ method, where $[delta][delta]C_t = [delta]C_{t,sample} - [delta]C_{t,reference}$. Here, $[delta]C_{T,sample}$ is the C_t value for any sample normalized to the endogenous housekeeping gene and $[delta]C_{t,\ reference}$ is the C_t value for the other sample of interest, also normalized to the endogenous housekeeping gene.

2.6 Summary points

1. RNA was isolated from precisely from pure epithelial normal human prostate using laser capture microdissection and RNA extraction
2. This RNA was amplified in two rounds to allow hybridization to Affymetrix HG-U133 Plus 2.0 arrays

Chapter 3. Gene expression profile of laser microdissected normal human prostate epithelium: comparison with laser microdissected cancer prostate epithelium

3.1 Introduction

The prostate is a heterogeneous organ consisting of a large stromal compartment with a smaller epithelial compartment (5-10%). It has a zonal anatomy with differing glandular appearances, putative differences in function between the zones and a likely different embryological origin (McNeal, 1981). Prostate cancer, an adenocarcinoma, is characterized by basal cell layer disruption, basement membrane disruption, progressive loss of secretory differentiation markers (e.g. prostatic acid phosphatase and PSA), increasing nuclear and nucleolar abnormalities, increasing proliferative potential, and increasing variation in DNA content (aneuploidy). It grows in an unrestrained pattern, infiltrating into surrounding normal stroma and epithelium, as well as benign prostatic hyperplasia. One of the difficulties therefore in studying normal and cancer prostate has been in isolating pure epithelial samples for DNA and RNA analysis. The use of primary cell culture, where the prevailing environment may promote growth of only epithelial or stromal cells, and cell lines is well established. For example LnCaP, a prostate cancer cell line, is derived from the supraclavicular lymph node of a man with metastatic prostate cancer (Horoszewicz *et al.*, 1983). However the gene expression pattern of cultured cells is unlikely to reflect the true environment within the prostate. In addition many of the cancer cell lines are derived from metastatic disease, thus missing many of the early changes in gene expression and limiting the discovery of new biomarkers and putative therapeutic targets.

Various studies (Dhanasekaran *et al.*, 2001, Singh *et al.*, 2002, Welsh *et al.*, 2001a, Luo *et al.*, 2001, Stuart *et al.*, 2004, Yu *et al.*, 2004) have described global gene expression in prostate cancer and compared these with non-diseased prostate tissue, obtained from cancer-containing prostates, as ‘normal’ controls to gain insights into changes associated with disease. These studies present a complex picture of the gene expression profile of human prostate. However, almost all of these studies have been performed using whole tissue samples containing a mixture of epithelial and the predominant stromal component. It is therefore difficult to ascertain how much distortion could be introduced by using heterogenous cell types as starting material

for this kind of analysis, particularly in the context of prostate adenocarcinoma. For example, to identify genes that are dysregulated in prostate adenocarcinoma it is essential to have a knowledge of the expression profile of normal, epithelial, cells from the prostate in young men, before the onset of the changes that may occur due to age and disease. This is important not only for an understanding of prostate cancer at the molecular level, but also for the identification of diagnostic or prognostic markers and for the design of effective therapies.

Normal adjacent tissue to prostate tumours is a commonly used 'control' used in differential gene expression studies of prostate cancer. This is because it is readily available from radical prostatectomy specimens and has been shown to possess genetic changes in prostate cancer that have subsequently been validated by other techniques (Dhanasekaran *et al.*, 2001). There is however evidence of a cancer 'field effect' in histological appearing normal tissue around several cancers, including lung (Franklin *et al.*, 1997), oesophagus (Prevo *et al.*, 1999), cervix (Chu *et al.*, 1999), and urinary tract (Takahashi *et al.*, 1998). Several studies have confirmed a field effect in prostate cancer (Ananthanarayanan *et al.*, 2005, Yu *et al.*, 2004, Chandran *et al.*, 2005). Previous prostate cancer microarray studies are therefore limited in this respect and likely to miss important changes in prostate carcinogenesis. The use of benign prostate tissue, BPH, as a control is also popular (Stamey *et al.*, 2001), however BPH represents a disease process (see above) and is unlikely to truly reflect the molecular events in prostate cancer. Prakash *et al.* 2002 used DNA microarrays to assess differential gene expression between normal transition zone and BPH (Prakash *et al.*, 2002). They defined a profile of 511 genes that differentiated the 2 tissues. Thus the use of BPH as a control tissue for prostate cancer molecular profiling is flawed. BPH is a proliferative disease of the stroma and use of BPH tissue as a reference sample when investigating prostate cancer gene expression profiling will yield a distorted picture of differentially expressed genes in cancer.

Historically several methods have been used to try and overcome this problem, including gross dissection to enrich specific cell populations (Radford *et al.*, 1993), irradiation of unwanted areas to destroy genetic material (Shibata *et al.*, 1992), and manual microdissection (Emmert-Buck *et al.*, 1994). These processes were laborious, time consuming, and technically difficult, especially in obtaining high quality genetic material for down stream molecular applications, especially for gene expression

analysis that requires RNA isolation. Laser capture microdissection has overcome these problems (Emmert-Buck *et al.*, 1996).

To date prostate cancer microarray studies have not used disease free, normal prostatic tissue or epithelial specific prostatic tissue for comparison with diseased samples to identify molecular changes. There is then clearly a need for a reference normal epithelial, and mesenchymal, prostatic database to compare the transcriptome of prostatic adenocarcinoma with. I therefore hypothesized that to study epithelial tumours, it is vital to be able to isolate pure epithelial samples free of contamination from the surrounding stroma.

3.1.1 Gene Expression Profiling of the Prostate using LCM and Microarray technology

A disease is a comparative state that can only be defined in relation to the steady or 'normal' state. Gene expression technology provides a great opportunity to define the molecular basis of human prostatic diseases. This however requires availability of gene expression profile from normal, disease free tissue. As prostate adenocarcinoma is manifested in the proliferative overgrowth of prostate epithelium, it is therefore important that a transcriptome map of as pure a population of epithelial cells as possible is described.

Several studies have used LCM, linear amplification, and microarrays in the study of prostate cancer (Tomlins *et al.*, 2007, Stanbrough *et al.*, 2006, Petrovics *et al.*, 2005, Luzzi *et al.*, 2003, Ernst *et al.*, 2002, Chen *et al.*, 2003). A number of putative oncogenes / tumour suppressor genes have been identified using these methods including: Hepsin (Chen *et al.*, 2003), Mapsin (Chen *et al.*, 2003), AMACR (Petrovics *et al.*, 2005, Ernst *et al.*, 2002), MKP1 (Febbo *et al.*, 2006), ERG (Petrovics *et al.*, 2005), DD3 (Petrovics *et al.*, 2005), GSTP1 (Petrovics *et al.*, 2005) SGP28 (Ernst *et al.*, 2002), and low density lipoprotein-phospholipase (Ernst *et al.*, 2002). These studies have used either 'normal adjacent prostate' or benign prostatic hyperplasia as the tissue for comparison. The concept of a field effect has been demonstrated (Ananthanarayanan *et al.*, 2005, Yu *et al.*, 2004) and thus neither of these tissues represent the true transcriptome of normal human prostate. Tomlins *et al.* (Tomlins *et al.*, 2007) are the only group to have included in their study the epithelial compartment of 3 prostates from organ donors presumed free of disease. They were

grouped together with benign and normal adjacent prostate and not analysed separately, thus clouding their usefulness.

For the reasons described above, but foremost to provide a transcriptome map of the normal human prostate epithelium, I decided to investigate the gene expression profile of precisely dissected prostate epithelium from normal human prostate using Affymetrix GeneChip HG-U133 Plus 2.0 Arrays. Two hypotheses were tested:

1. RNA isolated from whole tissue is not representative of either the stromal or epithelial cell gene expression.
2. Existing prostate cancer microarray studies, comparing compromised normal (and invariably whole) tissue with cancer tissue arrays, give a 'distorted' picture of the gene expression profile changes in prostate cancer compared to non-diseased, normal tissue.

A normal gene expression profile of human prostate epithelium has therefore been established, and was validated using low-density real time PCR. This transcriptome was compared to the expression profiles of 'normal' whole prostate tissue used in several previous prostate cancer microarray studies using comparative microarray analysis.
Furthermore I have tested my dataset with a laser microdissected prostate cancer epithelial dataset (Febbo *et al.*, 2005) to identify potential new gene targets involved in prostate carcinogenesis. Three targets (MCM2, NR1D1 and ABCA1) were also chosen for further investigation at the protein level using a prostate tissue array in an unbiased, semi-automated particle analysis method.

3.2 Materials and Methods

3.2.1 Data analysis – gene expression patterns in human prostate epithelium

Affymetrix CEL files were normalised in GeneSpring 7.2 with GCRMA. Subsequent normalisation included setting of signal values less than 0.01 to 0.01, total chip normalisation to the 50th percentile, and normalisation of each gene to the median. These normalizations allowed for the visualization of data based on relative abundance at any given time point rather than compared with a specific control value. Following this genes with raw expression values less than 50 were filtered out, with

further filtering as required performed according to fold change. Differential gene expression was calculated using parametric t-test (Welch) $p<0.05$. Where specified, a false discovery rate correction was applied ($p<0.01$) using the Benjamini and Hochberg method (Benjamini and Hochberg, 1995). Hierarchical clustering was performed in GeneSpring 7.2.

3.2.2 Validation of oligoarray data and quantitation of selected targets by real-time PCR

Quantitative PCR (Q-PCR) was used to verify differential expression. Custom TaqMan low density arrays were designed for the desired genes on Microfluidic cards (Applied Biosystems) and PCR performed on Applied Biosystems 7900HT Fast Real-Time PCR System (section 2.8). Cycle thresholds (Ct) were calculated using SDS 2.2 software (Applied Biosystems). (In a real time PCR assay a positive reaction is detected by accumulation of a fluorescent signal. The Ct (cycle threshold) is defined as the number of cycles required for the fluorescent signal to cross the threshold (i.e. exceeds background level). Ct levels are inversely proportional to the amount of target nucleic acid in the sample (i.e. the lower the Ct level the greater the amount of target nucleic acid in the sample)).

3.2.3 Comparative analysis of normal LCM prostate transcriptome

Comparative microarray analyses were performed with several publicly available prostate microarray datasets using the homology table function of GeneSpring 7.2. Publicly available datasets were downloaded and imported into GeneSpring. A brief description of each study used for the comparative analysis is given below.

3.2.4 Publicly available human prostate gene expression array datasets for RNA isolated from whole tissue

Yu *et al.* 2004 performed a comprehensive gene expression analyses on RNA isolated from 152 whole tissue human prostate samples, including prostate cancer, prostate tissues adjacent to cancer, and donor prostate tissue free of disease (Yu *et al.*, 2004). The Affymetrix (Santa Clara, CA) U95av2, U95b and U95c chip sets were used. The donor prostate group was used for comparative analyses described here, and included 23 prostates obtained at the time of organ donation in brain-dead men, with ages ranging from 13 to 63 years old. There was no clinical or histological evidence of

prostatic or genitourinary disease. Yu *et al.* (Yu *et al.*, 2004) selected samples from the peripheral zones were that contained at least 60% glandular tissue and were processed, within 30 minutes after removal. Total RNA was extracted using the Qiagen RNeasy kit (Qiagen, San Diego, CA) and sufficient quantities were extracted so as not to require additional amplification. CEL files for this study were obtained from the Gene Expression Omnibus website (NCBI), (GSE6604 – expression data from normal prostate tissue free of any pathological alteration). These were directly imported into GeneSpring and normalized using GCRMA as previously described (chapter 2).

Singh *et al.* 2002 (Singh *et al.*, 2002), analyzed global gene expression in 52 prostate cancers and 50 'non-tumour' prostate samples. These were whole tissue. These 'non-tumour' samples were referred to as normal and were taken from histologically normal adjacent areas to prostate cancer. U95av2 arrays (Affymetrix, Santa Clara, CA) were used to generate gene expression profiles after total RNA extraction using the TRIzol® method (TRIzol®, invitrogen), generation of labeled cRNA, fragmentation and hybridization according to Affymetrix protocols. CEL files for these 'normal' prostate samples were obtained from the Broad Institute (http://www-genome.wi.mit.edu/MPR/prostate). These were imported into GeneSpring and normalised using GCRMA (chapter 2).

Welsh *et al.* 2001 (Welsh *et al.*, 2001a), analyzed gene expression in 25 prostate cancer tissues, 9 non-malignant prostate tissues, and 21 cell line samples. Sharp dissection was used to isolate cancerous and normal adjacent prostate tissue, which was sharp dissected. This was whole tissue. RNeasy (Qiagen) was used to extract total RNA from each sample and sufficient quantities were obtained to generate labeled cRNA with hybridisation to Affymetrix U95av2 arrays according to standard Affymetrix protocol. Scanned image files were analyzed with Genechip (Affymetrix) and scaled to an average hybridization intensity of 200. The dataset was obtained from http://public.gnf.org/cancer/prostate/ and was subsequently imported directly into GeneSpring. Normalisation was not performed as the dataset was already in scaled and normalised format (.txt).

3.2.5 Publicly available human prostate cancer gene expression array datasets for RNA isolated from for laser capture microdissected tissue

Febbo *et al.* 2006 (Febbo *et al.*, 2006), used laser capture microdissection to isolate prostate cancer before and after the neoadjuvant administration of imatinib mesylate to assess the impact on global gene expression of the drug. These men had intermediate or high risk prostate cancer, as based on PSA, Gleason score, and clinical staging (Table 4). Thirteen prostates were sampled before treatment using prostate biopsy specimens. LCM was performed using the Arcturus PixCell II system and RNA extracted using the Absolutely RNA Nanoprep kit (Stratagene). Two rounds of subsequent RNA amplification were performed using a modified protocol created by adapting several previously published methods. Labelled cRNA was then hybridized to U133A microarrays (Affymetrix, Santa Clara, CA). CEL files for the pre-treatment group (supplemental data http://clincancerres.aacrjournals.org/cgi/content/full/12/1/152/DC1) were imported into GeneSpring and normalized with GCRMA (chapter 2).

Patient number	PSA	Biopsy Gleason score	Clinical stage (DRE)
1	6	4 + 3 = 7	T1c
2	4	4 + 3 = 7	T2a
3	10.8	3 + 4 = 7	T1c
4	5.98	3 + 4 = 7	T1c
5	8.5	3 + 3 = 6	T2c
6	6.3	3 + 4 = 7	T2c
7	4.6	3 + 4 = 7	T2a
8	7.6	4 + 3 = 7	T1c
9	7.3	3 + 4 = 7	T2b
10	7.9	3 + 3 = 6	T1c
11	8.46	4 + 4 = 8	T1c

Table 4 Patient characteristics Febbo *et al.* (Febbo *et al.*, 2006)

As these datasets are from different generation Affymetrix chips, with fewer genes in different places on the chip, it is critical that they be ‘aligned’ with my chip, for laser

capture microdissected normal prostate, to provide a meaningful comparison. To do this I used homology tables (GeneSpring) to compare my dataset with other datasets. Homology tables in GeneSpring allow the comparison of expression results from one array (or genome) to another. A gene list or experiment can be translated, once a homology table is created, from one genome to a gene list of the corresponding genes in the other genome. In these experiments homology tables were used to translate HGU133 Plus 2.0 gene lists to HGU133A and HGU95av2 chips. Raw data were simultaneously normalized to standardize different generations of Affymetrix gene chips (*e.g.* HGU95A and HGU133 Plus 2.0). Differential gene expression and filtering on fold change could then be calculated as previously described (Chapter 2).

3.2.6 Protein expression in prostate tissue arrays

Patient selection, disease state and construction details of tissue blocks are given elsewhere (Nariculam *et al.*, 2009, Wang *et al.*, 2010). The tissue array was constructed by Joesph Nariculam, Prostate Cancer Research Center, UCL (a representative picture is shown in Supplemental Figure 3 - Appendix). Briefly, tissue blocks were constructed using archival formalin-fixed, paraffin-embedded radical prostatectomy specimens from the 82 patients with pathological stage pT3a or b and pre-operative PSA stage of >3. A urological pathologist (Mr Alex Freeman) examined all radical prostatectomy specimens. 5-6μm sections were cut from the tissue arrays onto coated slides and dried overnight at 60°C, prior to performing standard antigen retrieval. Immunostaining was performed by Philipa Munson, UCL Diagnositics, UCL, using standard 3,3-diaminobenzidine staining protocol and 0.5-2μg/ml of NR1D1 (ab56754), ABCA1 (ab53117) and MCM2 (ab53136) primary antibodies (Abcam) on an automated Bond maXTM machine (Vision BioSystems) using the Bond polymer detection system kit (containing post primary antibodies), according to manufacturer's protocol, at high contrast (DS9173) (Wang *et al.*, 2010).

Images of each tissue core were digitally acquired at 20x magnification at standardised settings. The images (3840 x 3072 pixels) for NR1D1 and ABCA1 antibodies were acquired individually with a Nikon DXM 1200 digital imaging system (image resolution = 3840 x 3072 pixels) attached to a Nikon Diaphot. Images for MCM2 staining were acquired using a Olympus FluoView FV 1000 microscope equipped with an automated stage (20x objective, image resolution 624 x 600); a grid map was created, images acquired in a series and were stitched together using

Olympus montaging software (performed by Marta Eillertsen).

A reproducible, automated method (Wang *et al.*, 2010) was employed to quantify the DAB signal on benign and malignant human prostate tissue cores using ImageJ software (Rasband, 1997). Macros were written to execute the following sequence of events for acquired jpeg images: 1. Open image 2. Convert to 16-bit image 3. Set threshold (see Figure 34 for details for each antibody) 4. Analyze particle (Size 0.5-Infinity, Circularity 0.00-1.00) 5. Save image 6. Save particle information (count, total area, average size and area fraction) into an excel spreadsheet (rsb.info.nih.gov/ij/docs/pdfs/examples.pdf). Units are default ImageJ setting (pixels). For all 3 proteins tested, expression was observed to be largely epithelial (see results) and analysis was also restricted to the epithelial expression; set threshold parameters were chosen after manual analysis of random cores for the subsequent quantitation of the signal. A contiguous spreadsheet for all the usable cores (between 242 and 289 cores) for normal vs. cancer comparison, for different antibodies was constructed and statistical analysis using ANOVA and Student's t-test was performed.

3.3 Results

3.3.1 Basic analysis of gene expression profiling in normal prostate

Eight LCM epithelium only samples from 3 different prostates were analyzed for gene expression analysis. The results of basic, exploratory analysis involving principal component analysis (PCA) and identification of genes that were present or absent in the normal prostate epithelium is given in Figure 26 and 27 (gene lists supplemental files 1A & B). PCA was performed using GeneSpring 7.2 software on condition (each prostate LCM sample) and on all genes (54,675 transcripts). PCA on conditions show principal components 1 and 2 with 38.1% and 14.6% variance. PCA on genes indicates that smaller number of components account for majority of variation (Figure 27). Using detection algorithm and significance cut off ($p<0.05$) we identified 11657 genes to be expressed (present) in the 8 epithelial samples and 21,124 genes were absent. Unsupervised clustering of the 8 epithelial samples did not reveal a significantly different gene expression pattern (Figure 28) between these samples. This therefore represents the first epithelial transcriptome map of the human prostate. This epithelial transcriptome map was used to test the hypothesis described in (section 3.1.1) (i) by comparing gene expression profile from my data from LCM normal prostate epithelial cells and those from whole tissue studies (see 3.2.4) and (ii)

by comparing my data with the Febbo study (Febbo *et al.*, 2006) describing gene expression profile of LCM prostate cancer epithelium, to identify novel, differentially expressed genes in prostate cancer.

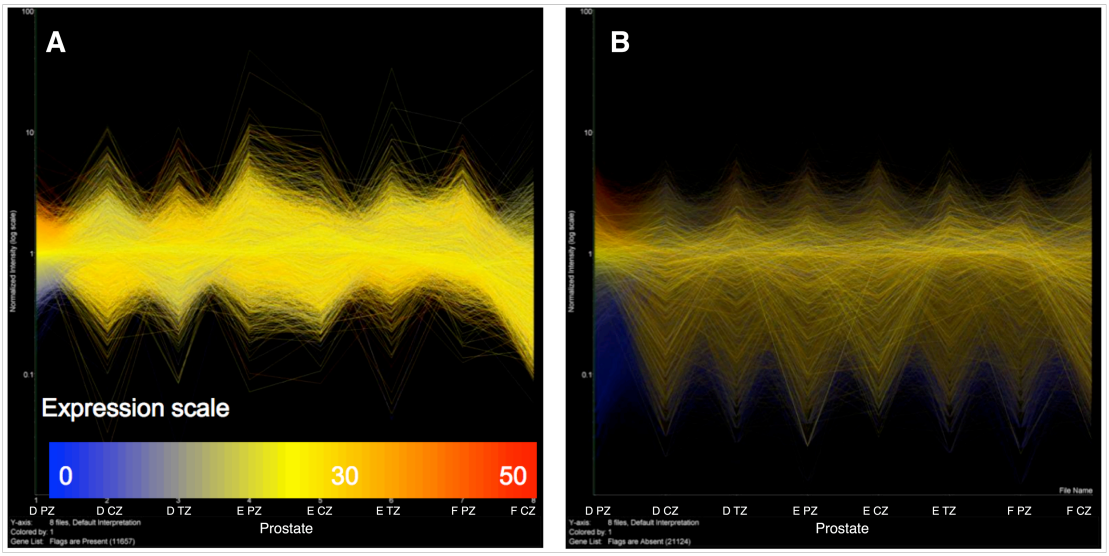
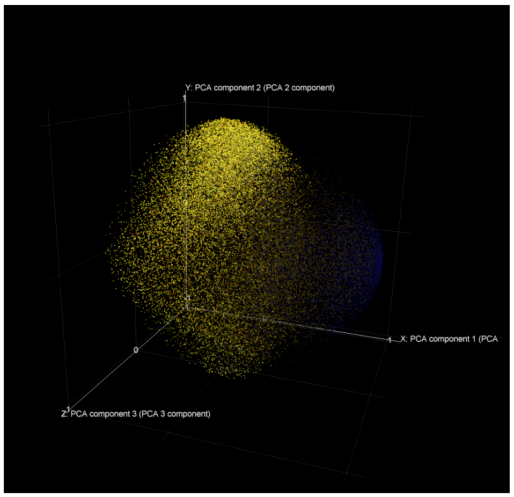


Figure 26 Using detection algorithm and significance cut off ($p<0.05$) 11657 genes were identified to be expressed (A = present) in all the 8 epithelial samples and 21,124 genes were absent (B).



Principal component	Percent Variance	Cumulative Percent
Principal component #1	38.14%	38.14%
Principal component #2	14.60%	52.74%
Principal component #3	10.11%	62.85%
Principal component #4	9.21%	72.06%
Principal component #5	7.67%	79.73%
Principal component #6	7.46%	87.19%
Principal component #7	6.68%	93.87%
Principal component #8	6.13%	100.00%

Figure 27 Principal component analysis on all genes of the expression data from normal, LCM prostate epithelium. PCA analysis was performed on all genes of the expression datasets from 8 Affymetrix samples from 3 donor prostate using basic protocols in GeneSpring GX 7.2 software.

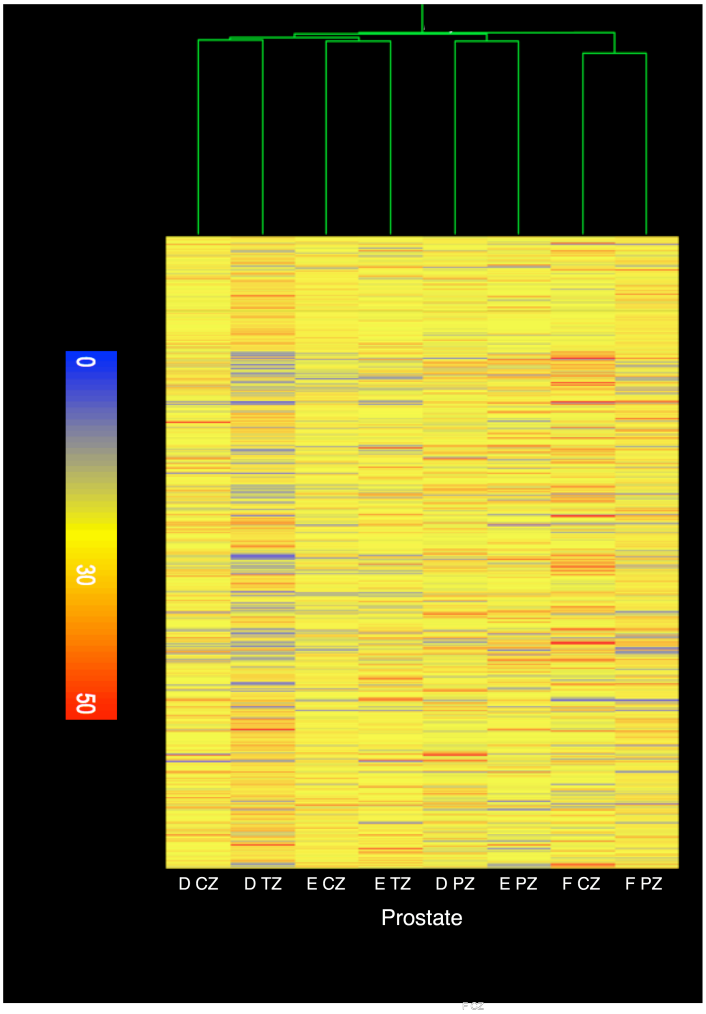


Figure 28 Hierarchical clustering on conditions (samples) did not reveal different gene expression profiles amongst the samples. This analysis was performed using the 'centroid' clustering method. In this method, the distance between two clusters is the distance between the averages of the data points under one branch and the averages of the data points under another. This method is sometimes referred to as the "average-linkage" method.

3.3.2 Validation of microarray data with real-time PCR

Randomly selected targets were used on a microfluidic card real-time PCR array for verification. 30 genes were randomly selected from the gene list (Table 5, Supplemental file 1A). For example, genes KLK3 (also called prostate specific antigen), B2M, IDH1 were all found to be present in the human prostate samples (supplemental file 1A) using real-time PCR. A complete list with average Ct values ± SD is given in Table 5.

Gene	Ct	SD
B2M	18.5854	1.704342
IFITM1	22.24602	1.425877
PAPD4	23.08227	15.84894
MAF	25.07318	1.522375
IMMP1L	25.3862	1.524564
PPAP2A	25.81274	2.92865
RPL17	26.45161	1.136174
MT01	27.20571	1.812181
NDUFA12	27.4901	1.291011
DARS	27.70086	0.670677
RARRSE1	27.85288	2.449404
PIAS2	28.15801	5.563861
PRPF4B	29.42128	0.739768
EIF1AY	29.59145	1.513877
IDH1	29.63436	0.572494
PLCB1	29.69154	3.045368
KLK3	30.55266	2.197981
WDR27	30.56812	3.318825
SCCPDH	30.62914	1.078141
RALB	30.80814	0.736271
PTPRA	30.81518	0.953013
SNAPC5	30.8319	1.1374
PPIG	31.61786	1.187032
STAT1	31.62738	1.125823
DKFZ	31.66869	1.660908
GAPDH	32.17615	0.77264
NBEA	32.39767	2.821052
COMMD3	32.4684	2.154918
CTBP2	33.33053	1.253607
GCOM1	35.15904	0.555935

Table 5 QPCR using Applied Biosystems Low Density arrays. Randomly selected targets were used on a microfluidic card real-time PCR array for verification. 30 genes were randomly selected from the gene list for Figure 27 A (genes present in all 8 epithelial samples). Each target was measured in quadruplicate on the array and data analyzed using SDS 2.2 software (Applied Biosystems).

3.3.3 Whole tissue vs. epithelium only microarrays of human prostate

Most of the available gene expression data utilizes RNA isolated from whole prostate tissue (Singh *et al.*, 2002, Welsh *et al.*, 2001a, Yu *et al.*, 2004). A major problem with the whole tissue studies is the dilution of epithelial mRNA as stromal tissue mass of the prostate is five times greater than the epithelial component. I hypothesized that using whole tissue gene expression profile of 'normal' prostate (whether obtained from disease adjacent or non-malignant prostate) will yield a distorted picture of epithelial gene expression. To test this hypothesis, I performed a direct comparison between: condition 1: all of my LCM epithelial samples (epithelium only) and condition 2: other whole tissue datasets (whole tissue) using homology table function in GeneSpring 7.2 software. To assess this, I obtained raw data files for some whole tissue prostate Affymetrix microarray studies (Section 3.2.4) and compared the gene expression profile of these whole tissue studies with my LCM epithelium only microarray data. I used multiple datasets (Welsh *et al.*, 2001a, Singh *et al.*, 2002, Yu *et al.*, 2004) from studies (utilizing Affymetrix GeneChip HG-U95Av2) that provided a 'normal' component, using RNA isolated from whole tissue, for their gene expression studies. CEL or .txt files from whole tissue and LCM studies were normalized using standard protocols and homology tables were constructed to standardize different generations of Affymetrix gene chips (*e.g.* HGU95A and HGU133 Plus 2.0).

Unsupervised and supervised analyses (Appendix) were used, for both types of analysis statistical tests and false detection rate (FDR) corrections were applied to create starting gene lists (details given in figure legends) for various comparisons. Unsupervised analysis confirms the differences in gene expression patterns (Figure 3.21) between the Singh (Singh *et al.*, 2002) whole normal prostate dataset and LCM normal epithelium only.

Gene clusters appear distinct for the whole tissue and LCM epithelial microarray datasets and hierarchical tree branches were significantly different (analytical details are given in figure 3.21 legend). Condition tree clustering of the initial datasets revealed 2 distinct clusters within the Singh normal dataset (Figure 29). Nineteen whole normal samples seemed to cluster with the epithelial only samples. Supervised clustering of normal samples grouped as LCM (n=8) and 'normal' whole tissue

samples of samples (n=40, from Welsh dataset), with FDR, shows major and significant differences in gene expression in whole tissue compared to LCM epithelium only gene expression (Figure 30, and gene list supplemental file 2). A similar analysis was performed for whole tissue gene expression dataset of Yu *et al.*, (Yu *et al.*, 2004) (Supplemental Figure 1 – Appendix), with similar results as that observed for the Welsh study, revealing distinct gene clusters for the whole tissue and LCM epithelial microarray and significantly different hierarchical tree branches. These results indicate major discrepancies between epithelial and whole tissue human prostate transcriptome.

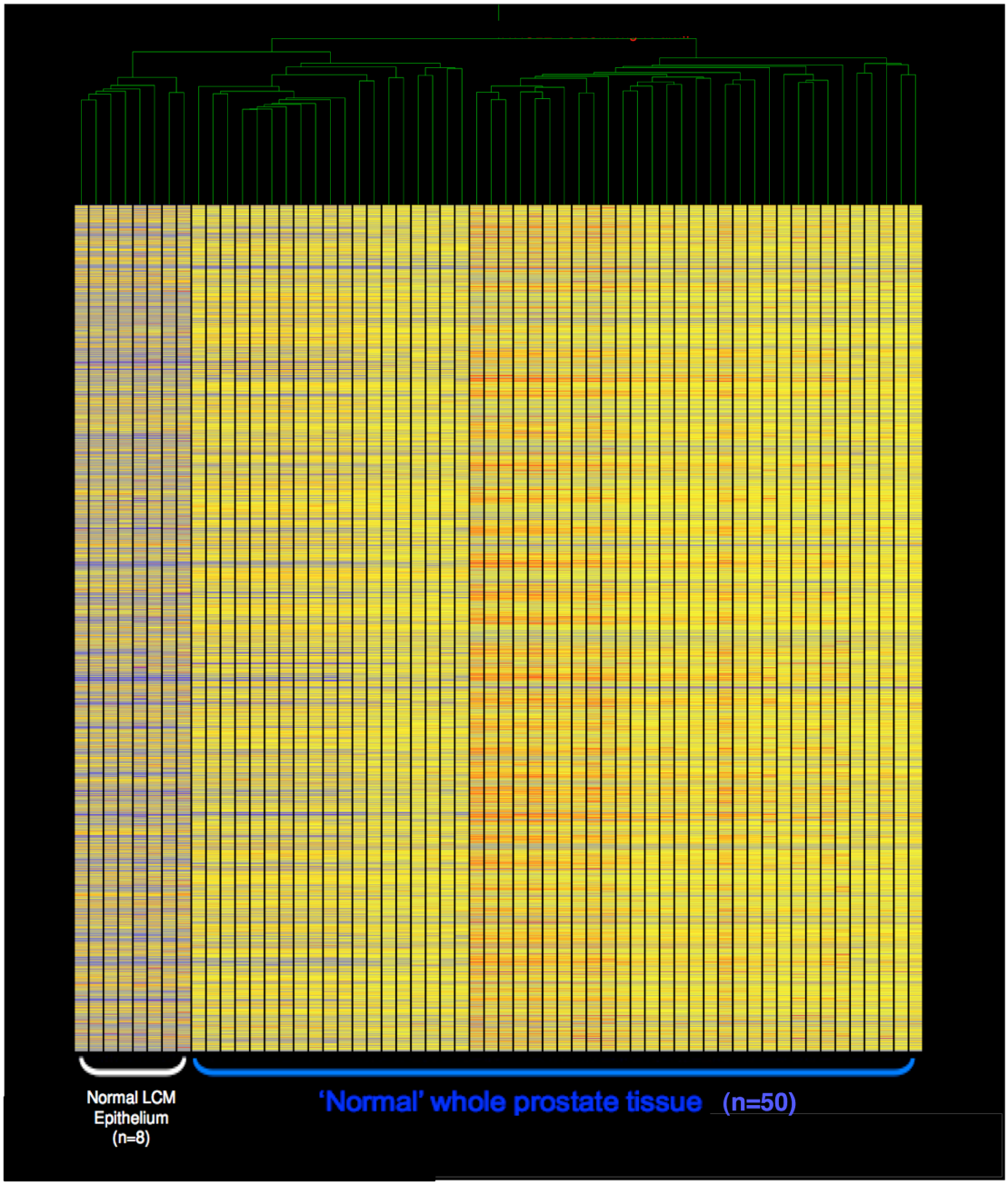


Figure 29 Gene expression profile of normal LCM epithelium vs. 'normal' whole prostate show major differences. Unsupervised hierarchical clustering of samples (all genes) group distinctly into LCM and whole prostate ("average-linkage" method). Whole prostate data on Affymetrix chip HGU95A from *Singh et al.* (Singh *et al.*, 2002). Concurrent normalization of LCM and Singh prostate dataset performed using homology tables (GeneSpring) for direct comparative analysis between the two datasets.

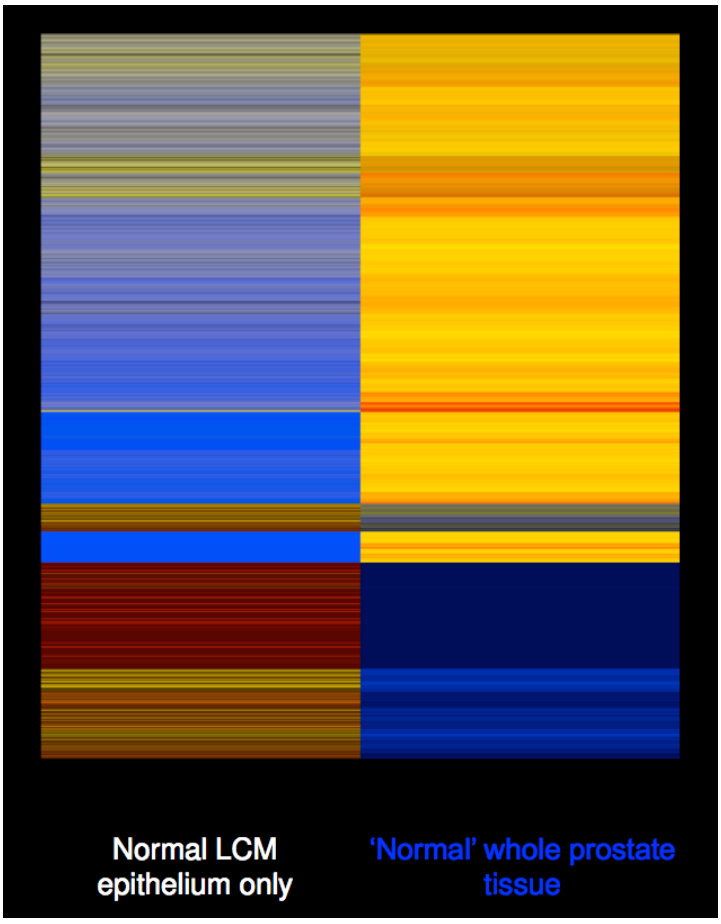


Figure 30 Supervised clustering of normal LCM epithelium vs. 'normal' whole prostate. Clustering of all samples, grouped as LCM vs. whole prostate (from Welsh *et al*, 2001, on HGU95A), under stringent conditions (FDR and t-test, $p < 0.01$), shows major (5498 genes) differences in gene expression (gene list supplemental file 2).

3.3.4 Identification of novel prostate adenocarcinoma specific genes

I next identified genes that are likely to be dysregulated in prostate adenocarcinoma. Numerous studies (Dhanasekaran *et al.*, 2001, Singh *et al.*, 2002, Welsh *et al.*, 2001a, Yu *et al.*, 2004) have used whole tissue gene expression profiling to identify targets differentially expressed in prostate cancer. In the light of my results comparing LCM epithelium only vs. whole tissue profiling I wished to extend the normal vs. cancer comparison to human prostate epithelium. Prostate cancer is an adenocarcinoma, and a comparison of normal epithelium only vs. cancer epithelium only gene expression profile is more likely to yield genes that are dysregulated in prostate cancer. To achieve this I used the dataset from a seminal report from the Febbo group (Febbo *et al.*, 2006). This paper used the same approach as employed in this study, by isolating

epithelium using LCM, but from the cancerous prostate and performing a microarray study using Affymetrix HGU 133A gene expression array. However, there was no normal LCM control dataset for this study. I obtained the raw data files of the LCM prostate cancer dataset (13 samples, as described in section 3.2.5) and by using the homology tables protocol in GeneSpring GX 7.2, performed a comparative analysis with my normal epithelium LCM dataset to obtain a novel prostate adenocarcinoma gene profile. An initial gene list (10168 genes) was constructed by clustering statistically significant ($p<0.01$) expressed genes, grouped by cancer (Febbo *et al.*, 2006) vs. normal (our dataset) with FDR correction (Benjamini and Hochberg, 1995). The normal and cancer epithelium only datasets were then analyzed using filtering on fold change. To limit the size of the dataset a 2-fold change was applied to the starting gene list of statistically significant genes (10168) of which yielded 9318 gene that showed differential expression between the two datasets (Figure 31); of these 7844 (Supplemental Figure 2A – Appendix, supplemental file 4A) were found to have increased expression whereas 1474 (Supplemental Figure 2B – Appendix, supplemental file 4B) genes showed decreased expression in cancer epithelium compared to normal epithelium. Numerous prostate oncogenes that have been previously identified were over-expressed in cancer vs. normal, including PTEN, Hepsin, AMACR, KLK3, Caveolin, KLK2, CDKN1B, and ERBB2. An excellent review of these prostate oncogenes is provided by Kumar-Sinha et al (Kumar-Sinha and Chinnaiyan, 2003). In addition, two known prostate tumour suppressor genes (Kumar-Sinha and Chinnaiyan, 2003) TP53 and TGFB2 were down-regulated in the prostate cancer dataset. This analysis also identified several genes not previously known to be involved in prostate carcinogenesis including YWHAE, GLO1 and ZNF143 (section 3.4.2).

Further analysis was performed to establish the validity of normal vs. cancer gene expression analysis in the epithelium only and whole tissue studies (Venn diagram - Figure 32) using three different microarray studies (Singh *et al.*, 2002, Welsh *et al.*, 2001a, Yu *et al.*, 2004). Only 308 genes in whole tissue normal vs. cancer match up with the prostate adenocarcinoma genes obtained from LCM normal vs. cancer epithelium only data set.

Thus this data indicates that whole tissue prostate cancer microarray studies give a distorted picture of the prostate epithelium gene expression profile. Consequently

using whole tissue for identification of differentially expressed genes in prostate cancer is also invalid because of 1) the mixed cellular contingent (stroma and epithelium), and 2) 'whole normal' prostate is not truly normal as it may contain field effect changes or represent other disease processes such as BPH.

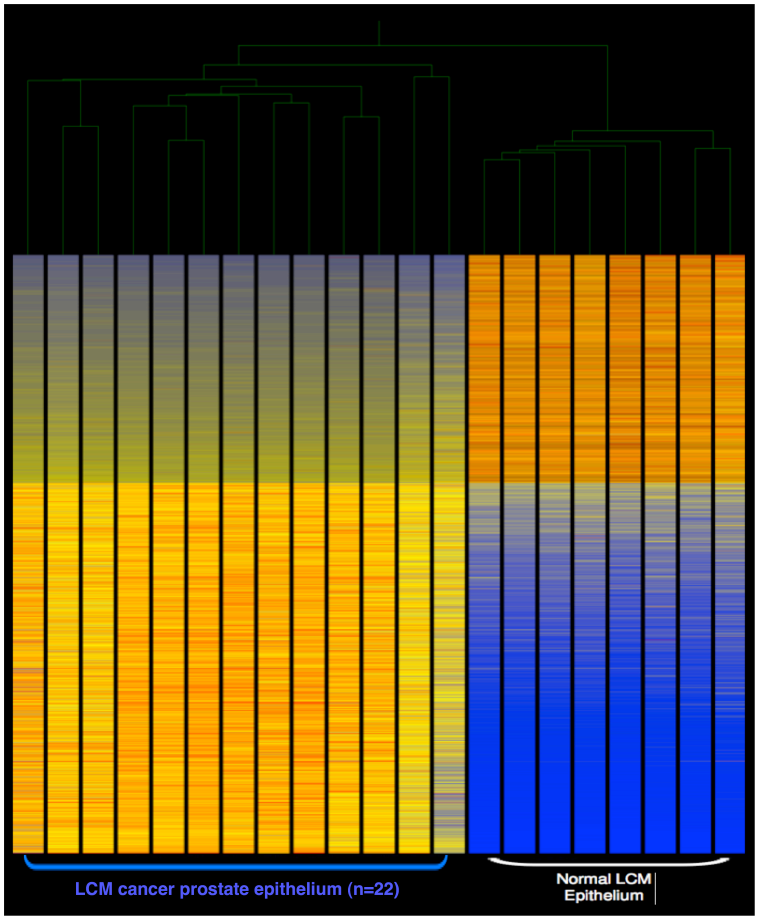


Figure 31 Prostate adenocarcinoma specific genes. Clustering of statistically significant ($p<0.01$) and differentially expressed genes (up- or down-regulated at 2-fold), grouped by cancer vs. normal with FDR (Benjamini and Hochberg). Cancer prostate data on Affymetrix chip HGU133A from Febbo *et al.* (Febbo *et al.*, 2006). Concurrent normalization of normal and cancer prostate datasets, using homology table protocol (GeneSpring) for a direct comparative analysis between the two datasets.

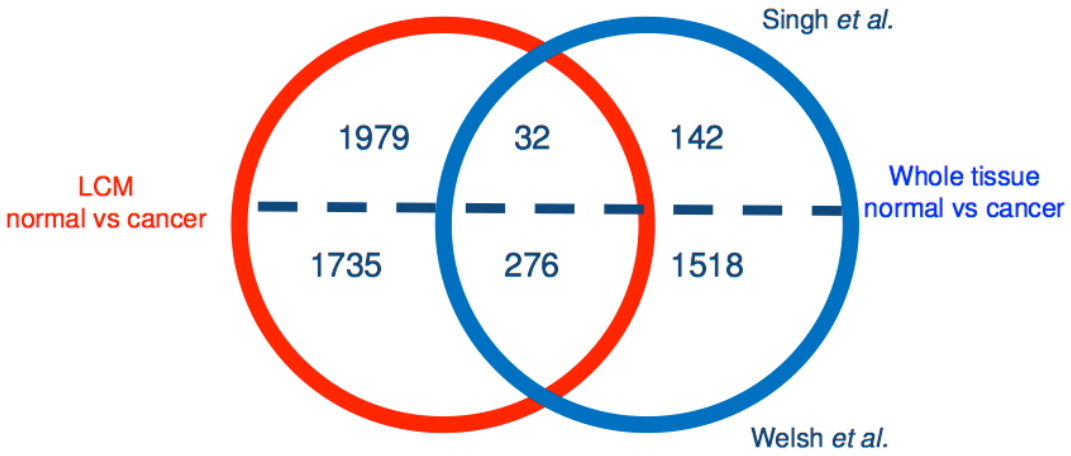


Figure 32 Venn diagram of gene expression analysis in LCM vs. whole and normal vs. cancer samples. Differentially expressed genes in whole tissue compared to LCM epithelium (red circle). Using the homology table analysis, necessary for subsequent comparison with epithelium only data, 174 genes were identified to be differentially expressed (at 5 fold change and FDR of $p<0.1$) in the Singh study and 1794 in Welsh study. 630 genes were significantly differentially expressed between the pure epithelial dataset and the whole normal Yu dataset at a 5-fold change. Further analysis indicated, *e.g.* at 5-fold change, that 171 were significantly down-regulated in the whole normal sample set with 459 up-regulated.

3.3.5 Identification of epithelium specific protein markers from genes over-expressed in prostate cancer

I selected 3 genes (MCM2, ABCA1 and NR1D1), identified from gene expression analysis to be over-expressed in prostate cancer epithelium, for further characterization using immunohistochemistry on a prostate tissue array (Wang *et al.*, 2010). The tissue array was constructed using archived, paraffin embedded, benign or adjacent tissue and prostate cancer tissue identified by a histopathologist. Protein expression of ABCA1 and NR1D1 has not been investigated in prostate tissue previously and hence represented putative novel targets for prostate cancer. MCM2 has been previously shown to be over-expressed in prostate cancer (Meng *et al.*, 2001), and was used as a positive control for verification.

DAB label representing the expression of NR1D1, ABCA1 and MCM2 was increased in prostate cancer cores compared to benign or normal cores (Figure 33). The label was quantified, in an unbiased manner, using a reproducible, semi-automated particle

analysis (Analyze Particles) protocol (Wang *et al.*, 2010) using ImageJ software (Rasband, 1997) using grayscale images (Figure 33) from 600 individual prostate tissue cores (chapter 3.2.6). Calculated parameters of count, total area, average size and area fraction are given in Table 6. Integration of area under the curve revealed a 22-60% increase in the expression (total area and area fraction (total area /total pixels), representing the extent and intensity of staining) of NR1D1, ABCA1 and MCM2 in malignant cores compared to benign cores ($p<0.0001$, Student t-test) (Figure 34). These results identify NR1D1 and ABCA1 as novel gene and protein markers in prostate cancer.

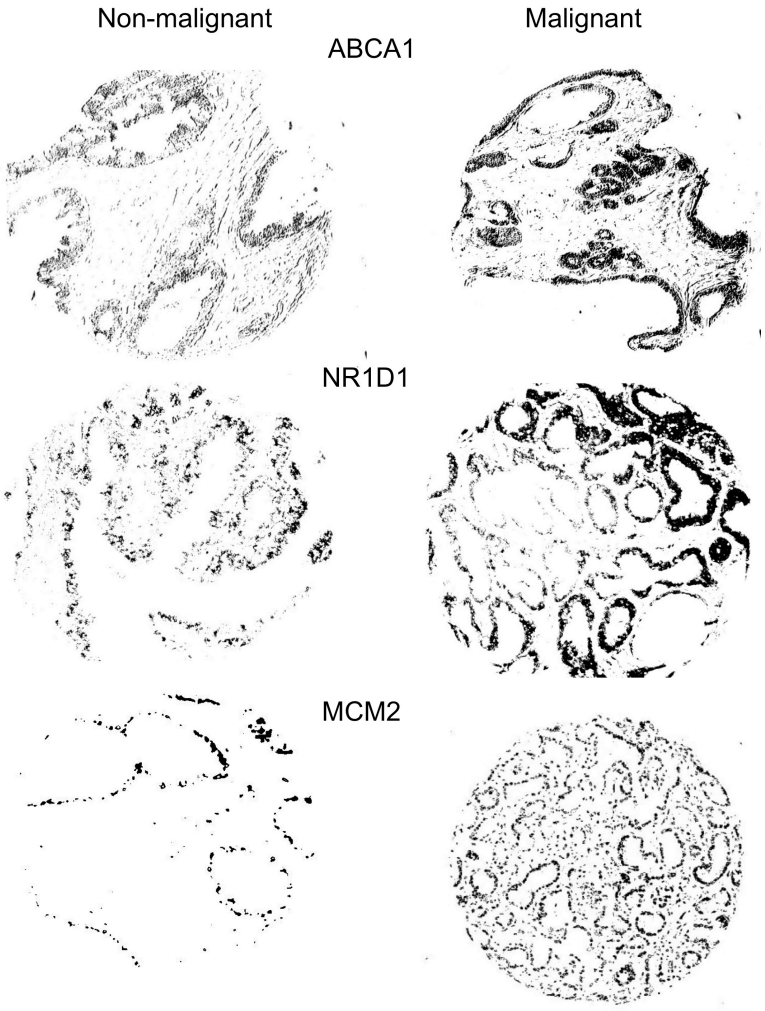


Figure 33 Protein expression of ABCA1, NR1D1 and MCM2 in non-malignant and malignant tissue cores. Representative malignant and non-malignant DAB label (converted to 16 bit grayscale) micrographs for ABCA1, NR1D1 and MCM2 used for signal quantitation (Figure 34).

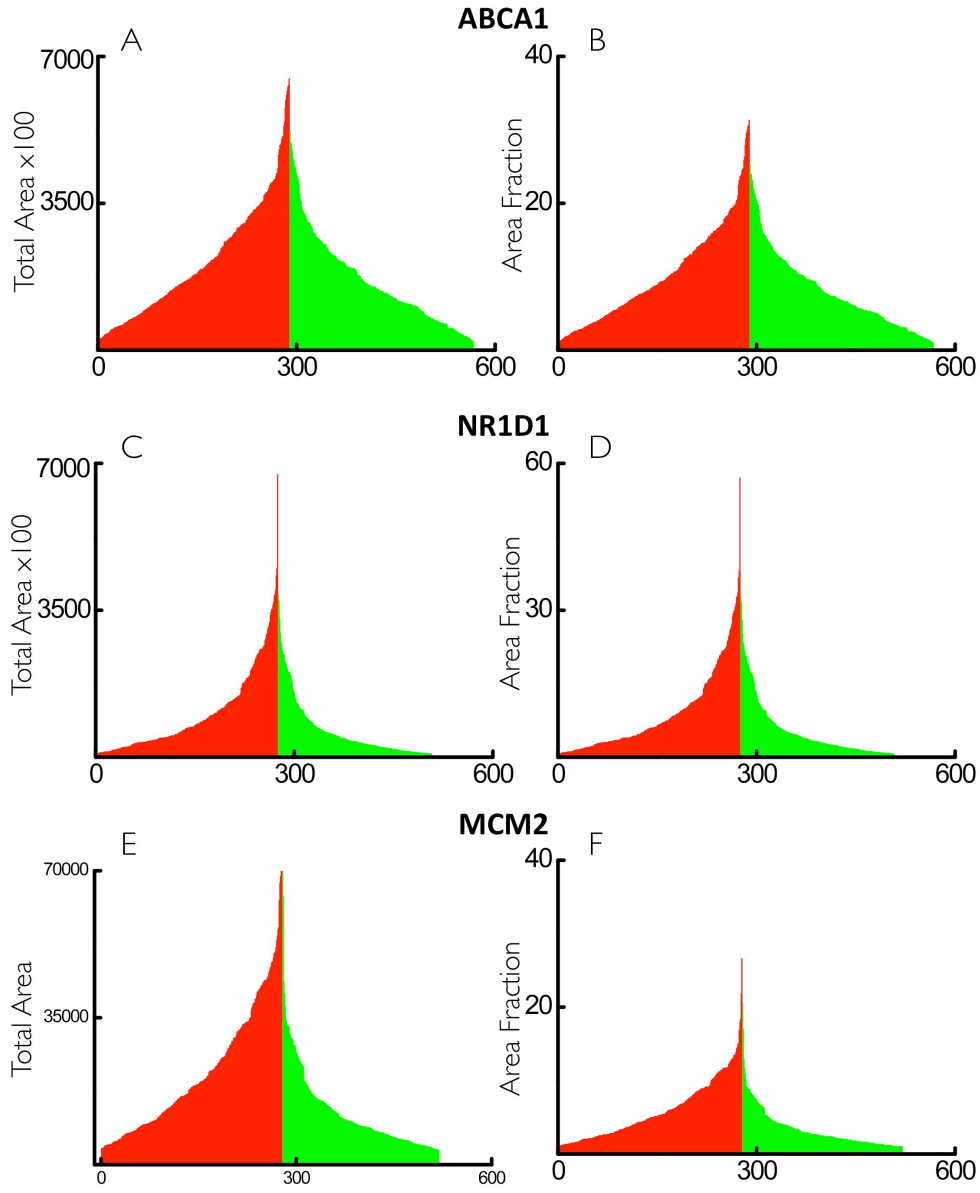


Figure 34 Quantitation of ABCA1, NR1D1 and MCM2 staining in non-malignant and malignant human prostate tissue. DAB signal for ABCA1, NR1D1 and MCM2 was quantified from tissue cores (Figure 33) using Analyze Particle protocol in ImageJ software to obtain Total Area stained (A,C,E) and Area Fraction (B,D,F, total area divided by the total pixels in the image). ABCA1, NR1D1 and MCM2 expression was increased in malignant v non-malignant cores ($p<0.0001$). Each bin is data for an individual, malignant (red) or benign (green) n=278, 232 and 242 non-malignant and 289,275 and 278 malignant, usable tissue cores for ABCA1, NR1D1 and MCM2, respectively.

Protein	Condition	Count	Total area	Average size	Area fraction
NR1D1	Benign	14098 ± 480	627967 ± 47974	41 ± 3	5.3 ± 0.4
	Malignant	17517 ± 443	1024909 ± 60059	54.9 ± 3	8.7 ± 1
ABCA1	Benign	28953 ± 824	938498 ± 36529	32.6 ± 0.9	7.9 ± 0
	Malignant	37314 ± 1022	1186946 ± 46157	31.0 ± 0.8	10.1 ± 0.4
MCM2	Benign	729 ± 40	12523 ± 728	27 ± 7	3.3 ± 0.2
	Malignant	944 ± 32	21792 ± 913	30 ± 2	6 ± 0.2

Table 6 Quantitation of protein expression in malignant and non-malignant human prostate tissue arrays using ImageJ software. DAB label, representing NR1D1, ABCA1 and MCM2 expression was quantified, in an unbiased manner, by using a reproducible, semi-automated particle analysis (Analyze Particles) protocol with ImageJ software (see methods for details). Over 500 individual prostate tissue cores RGB images were converted into 16-bit grayscale (*e.g.* from images shown in Figure 3.35). The results are means ± SE for the calculated parameters of count, total area, average size and area fraction.

3.4 Discussion

The combination of LCM and microarrays to investigate differential gene expression between normal and cancerous epithelium has been published for most human cancers including mesothelioma (Mohr *et al.*, 2004), head and neck squamous cell carcinoma (Leethanakul *et al.*, 2000), endometrial carcinoma (Wen-Xin and Xi-Shan, 2007), gastric carcinoma (Wu *et al.*, 2005), pancreatic carcinoma (Crnogorac-Jurcevic *et al.*, 2002), breast carcinoma (Seth *et al.*, 2006) and urothelial carcinoma (Wallard *et al.*, 2006) to name a few. Whilst platforms and commercial kits may differ (e.g. Affymetrix vs. Agilent, Arcturus vs. PALM), the same broad principles apply. Low quantities (nanogram) of RNA are isolated after LCM. This requires an additional amplification process prior to microarray hybridisation, with the most common being T7 linear amplification (as used in all the above studies). After subsequent hybridization to a microarray platform, differential gene expression and subsequent exploratory analysis are performed using statistical methods and software (e.g. GeneSpring). Targets of interest identified are then confirmed at either / both the transcript and protein level. This study is no different and uses these well-described, validated techniques. As in other cancers it has been possible to discover genes that are up / down regulated in normal vs. cancer epithelium and to validate some of these.

Knowledge of the expression profile of normal, disease free, epithelial cells from human prostate is a prerequisite for the identification of dysregulated genes in prostate adenocarcinoma. I used a combination of laser capture microdissection of human prostate epithelium, gene expression and tissue microarrays to provide the first transcriptome map of normal human prostate epithelium. By comparing my normal epithelial data set with an existing cancer epithelium dataset (Febbo *et al.*, 2006), I have been able to identify novel, differentially expressed genes in prostate adenocarcinoma. Also, by using a tissue microarray and a quantitative analytical approach I have identified two new epithelial specific protein targets for prostate cancer.

Furthermore, by analyzing normal epithelium only gene expression profile, I have identified several targets that have otherwise not previously been described in prostate cancer microarray studies (YWHAE, GLO1, and ZNF143), as well as confirming many that have been previously reported (e.g. Hepsin, AMACR, PTEN, CDKN1B). These results are discussed below.

3.4.1 Whole tissue vs. LCM epithelium gene expression profiles

Prostate cancer is a disease of the epithelium with the epithelial cells the most likely site of initiation and maintenance of the disease. The majority of prostate cancer microarray studies have used whole tissue from normal adjacent or diseased tissue (BPH) as reference or control tissue (Dhanasekaran *et al.*, 2001, Singh *et al.*, 2002, Welsh *et al.*, 2001a, Yu *et al.*, 2004). RNA isolated from whole tissue contains representative genes expressed in different cell type of that tissue. This rational led to the development of various cell dissection techniques including laser capture microdissection (Emmert-Buck *et al.*, 1996). Thus a major problem with using whole prostate tissue is that it contains mixed populations of epithelium and mesenchyme and stroma, reflecting several cell types such as glandular epithelial cells and stromal cells. Total RNA for a given gene from epithelium will therefore be diluted, if it is expressed in other cell types in the tissue. As most microarray studies aim to provide an understanding how gene expression changes in prostate cancer, this becomes an even more critical issue because prostate cancer is a carcinoma and is un-regulated proliferation of the epithelium, predominantly. Analysis of gene expression from whole tissue will be clouded by these differing cellular components and in case of

gene expression profiling of prostate cancer will mask the true changes that are occurring in the epithelium. I reasoned that to assess true gene expression alterations in prostate epithelium, pure populations of 'normal' and cancer epithelium should be compared. By comparing LCM epithelium and representative whole tissue gene expression arrays I have demonstrated that, as expected, whole tissue gene expression arrays distort the gene expression profile of epithelial cells.

Two studies have previously addressed the question of whether whole tissue affects the overall gene profile generated by microarray profiling of cancers, when compared with a pure epithelial cancer subset isolated by LCM (Harrell *et al.*, 2008, El-Serag *et al.*, 2009). Harrel *et al.* (Harrell *et al.*, 2008) used expression profiling to define genes that contribute to breast cancer spread into and/or growth within draining lymph nodes (LN). Whole tumor xenografts and their matched whole LN metastases were compared to LCM captured cancer cells from the same tumors and matched LN metastases. Whilst similar numbers of differentially expressed genes were identified (1930 vs. 1281 genes), less than 1% (30 genes) were common to both methods. Thus distinctly different lists of metastasis-promoting genes were generated. This mirrors the findings from my study.

El-Serag *et al.* (El-Serag *et al.*, 2009) examined whether comparative microarray analysis of LCM diseased vs. normal epithelium in Barrett's oesophagus, showed similar changes to that of when whole tissue normal vs. diseased microarray studies were performed. They found that the LCM experiment gave a more detailed picture of the Barrett's oesophagus phenotype (3443 differentially expressed genes) than obtained from the whole tissue study (1797 differentially expressed genes). In addition not only did the LCM samples have a larger number of differentially expressed genes, but they showed more genes that had a high magnitude of differential expression. They felt that the extent of overlap between LCM and whole biopsy samples was high, with 74% of the genes differentially expressed in whole tissue being confirmed as differentially expressed in LCM samples.

3.4.2 Identification of novel biomarkers for prostate cancer

By using an LCM dataset for prostate cancer (Febbo *et al.*, 2006), and homology table protocol in GeneSpring software, I have been able to identify novel genes under- and over-expressed in prostate cancer (Supplemental Figure 2, genelist supplemental file

4A and B). Prostate cancer is a disease with origins in the prostate epithelium. It is therefore imperative that a cell type specific gene and protein profiling is available for a better understanding of its origin and mechanisms. I believe this report has begun to address this issue. In addition to numerous genes already documented as playing a role in prostate carcinogenesis (Section 3.3.4), this study has also discovered several genes not previously identified by experiment. I believe that these genes may have been masked in previous microarray studies by the use of whole tissue, or by the nature of normal control tissue used.

The YWHAE gene encodes for the protein 14-3-3e, which binds to phosphoserine-containing proteins and mediates signal transduction (Roy *et al.*, 1998). 14-3-3 binding is required for the stabilization of active RAF-152 and CDC25-mediated cell cycle control (Roy *et al.*, 1998), whereas its interaction with BAD and BAX prevents their pro-apoptotic release to mitochondrial membrane (Won *et al.*, 2003, Nomura *et al.*, 2003). Up-regulation of YWHAE in breast cancer is associated with a poorer overall survival and worsening disease progression (Cimino *et al.*, 2008). Nothing is known about YWHAE in the prostate and it is therefore a novel adenocarcinoma gene that was significantly over-expressed in this study. However down-regulation in the expression of pro-apoptotic BAX has been associated with reduced responsiveness to radiotherapy in prostate cancer (Mackey *et al.*, 1998), and this provides a potential pathway / mechanism by which it may act.

Glyoxalase 1 (GLO1) is involved in the glycolytic pathway by detoxifying the reactive methylglyoxal (MGO) into D-lactate in a two-step reaction using glutathione (GSH) as cofactor (Santel *et al.*, 2008). Inhibitors of glyoxalases are considered as anti-inflammatory and anti-carcinogenic agents (Santel *et al.*, 2008). Recently the screening of a set of 618 human cancer cell lines using real time qPCR identified GLO1 as the most frequently amplified gene, with 8.4% of the informative samples having two-fold or greater amplification (Santarius *et al.*, 2010). In addition RNAi knockdown of GLO1 had the greatest and most consistent impact on cell accumulation and apoptosis. No prostate cancer samples were included however. The expression of GLO1 was recently examined in human melanoma tissue at a gene and protein level (Bair *et al.*, 2010). Significant up-regulation was seen at both levels, and subsequent siRNA interference targeting GLO1 expression, sensitized 2 human

metastatic melanoma cell lines towards becoming antiproliferative and apoptogenic. In a recent study 27 transcripts were investigated as potential novel markers for prostate cancer, including GLO1 (Romanuik *et al.*, 2009). Although not significantly differentially expressed between laser microdissected malignant and benign samples of human prostate tissue, GLO1 was increased in patients with primary prostate cancer compared to those who later had biochemical failure. Expression of GLO1 was significantly decreased in metastatic castration-recurrent disease compared with androgen-dependent primary prostate cancer (Romanuik *et al.*, 2009). My study represents the first time that GLO1 has been shown to be over-expressed in prostate cancer epithelium compared with normal prostate epithelium. It is a novel potential biomarker that may therefore be useful in the diagnosis / prognosis of prostate cancer.

Staf, originally identified in Xenopus laevis, plays a pivotal role in transcriptional activation not only of snRNA and snRNA-type promoters by RNA Pol II and Pol III, but also of mRNA promoters (Myslinski *et al.*, 1998). One human equivalent is Zinc finger protein 143 (ZNF143) which is 84% equivalent to its Xenopus equivalent, and which was significantly down-regulated in the prostate cancer dataset in this study. To date, several protein-coding genes have been described as regulated by ZNF143: the cytosolic chaperonin containing t-complex polypeptide 1 (TCP1), the interferon regulatory factor (IRF3), the neuronal nitric-oxide synthase (NOS1), the transaldolase (TALDO1), the aldehyde reductase (AKR1A1), the mitochondrial ribosomal protein S11 (MRPS11), the synaptobrevin-like 1 (SYBL1), the human cell cycle regulated BUB1B gene (Myslinski *et al.*, 2007), and the mitochondrial transcription factor (TFAM) (Gerard *et al.*, 2007). ZNF143 is induced by cisplatin treatment and binds preferentially to cisplatin-modified DNA (Ishiguchi *et al.*, 2004), suggesting that it plays an important role in cisplatin resistance. More recently it has been shown to interact with p73, being involved in cisplatin sensitivity through the regulation of DNA repair gene expression (Wakasugi *et al.*, 2007). Thus its levels are increased in cancer cells resistant to anticancer drugs such as cisplatin, but aside from this its role in carcinogenesis is unknown. It has not been previously described in relation to normal or diseased human prostate, and its role as a potential tumour suppressor gene requires further evaluation.

Thus at least three potential novel biomarkers, described here in detail, have been identified; these were not previously shown by global gene expression analyses using whole tissue (Dhanasekaran *et al.*, 2001, Luo *et al.*, 2001, Magee *et al.*, 2001, Welsh *et al.*, 2001a). It is likely that their discovery has been enabled by the design of this experiment, with the use of truly 'normal' prostate epithelium free from the corruption of diseased 'normal' tissue and free from stroma as a control tissue.

3.4.3 Comparison of gene expression results with previously published data

Thus study has confirmed many targets previously identified by microarray studies (Welsh *et al.*, 2001a, Luo *et al.*, 2001, Singh *et al.*, 2002, Yu *et al.*, 2004) to be dysregulated in prostate cancer including over-expressed: Hepsin, AMACR, KLK3, KLK2, Caveolin, PTEN and CDKN1B, and under-expressed genes: TP53 and TGFB2 (Gene list supplemental files 4A and B).

Whilst the previous microarray studies of Welsh (Welsh *et al.*, 2001a), Singh (Singh *et al.*, 2002) and Yu (Yu *et al.*, 2004), that were used in this study were designed with other objectives than pure differential gene expression discovery, there is concordance in this study with some of their findings, as well as discordance with others. Welsh *et al.* (Welsh *et al.*, 2001a) analyzed gene expression in 25 prostate cancer tissues, 9 non-malignant prostate tissues, and 21 cell line samples. They were able to show a precise distinction between normal and tumour samples, with roughly 400 genes that were significantly over-expressed in prostate cancer. Of their top 20 over-expressed genes in prostate cancer, 15 were also identified in this study as over-expressed in prostate cancer including Hepsin, AMACR, LIM, TSPAN-1, CAMKK2, and KLK3. Two targets they identified and subsequently validated MIC-1 (RT-PCR) and fatty acid synthase (immunohistochemistry), were not significantly differently expressed in this study.

Singh et *al.* (Singh *et al.*, 2002) analyzed global gene expression in 52 prostate cancers and 50 'non-tumour' prostate samples. These were whole tissue samples and their analysis indicated that 317 genes had higher expression in the tumor samples whereas 139 genes were more highly expressed in normal prostate samples. Of their top 50 over-expressed targets in cancer there was concordance with 37 over-expressed genes in cancer this study. Interestingly again fatty acid synthase was up regulated in the Singh dataset, but not mine. There was however poor concordance with under-expressed targets in cancer with only 8 of their top 50 targets mirrored in this study.

Yu *et al.* (Yu *et al.*, 2004) performed a comprehensive gene expression analyses on RNA isolated from 152 whole tissue human prostate samples, including prostate cancer (66), prostate tissues adjacent to cancer (60), and donor prostate tissue free of disease (23). They identified a set of 671 genes whose expression levels were significantly altered in cancer compared with normal tissues. Interestingly, the expression patterns of histological benign prostate tissues (adjacent) were significantly overlapped with those of cancer, and were distinctly different than donor prostate tissue, suggesting a 'field effect'. This was validated at the protein level with increased expression of AMACR both in cancer and normal adjacent tissue. A full gene list is not provided in the paper, however in a table they provide 20 known prostate cancer related genes that were over-expressed in their paper. Of these 50% were over expressed in my study including again AMACR, HEPSIN, PSMA, p27 Kip1 and Caveolin.

This study has also revealed new targets not previously identified (YWHAE, GLO1, Wnt4, ABCA1 and NR1D1) and there are clearly large differences in differential gene expression between this study and those already mentioned (Welsh *et al.*, 2001a, Singh *et al.*, 2002, Yu *et al.*, 2004). There are numerous potential explanations for these discrepancies including as discussed, that these previous studies are disadvantaged by the heterogenous nature of their tissue (stroma included), and by the lack of a truly normal control tissue (normal adjacent - field effect, or diseased – BPH).

There are also many additional sources of variation in a microarray experiment, which can be attributed to biological and technical causes (Churchill, 2002). Biological variation results from tissue heterogeneity, genetic polymorphism, and changes in mRNA levels within cells and among individuals due to sex, age, race, genotype-environment interactions and other factors (Leung and Cavalieri, 2003). This variation is what investigators are trying to establish, however, preparation of samples, labeling, hybridization, and other steps of a microarray experiment can contribute to technical variation, which can significantly impact the quality of array data (Bakay *et al.*, 2002). Thus some differences might be explained by technical variation alone introduced by the use of different GeneChip generations (HG-U95AV2 vs. HG-U133 vs. HG-U133 Plus 2.0), differing downstream analytical techniques (normalization, filtering, etc) and laboratory practices, tissue handling etc.

In addition using the latest Affymetrix GeneChips (HG-U133 Plus 2.0), which were produced following completion of the human genome and thus contain many more transcripts, allows for the discovery of many more potential targets.

An alternative approach to the one used in this study, that could be employed, would be based on meta-analysis (Rhodes *et al.*, 2002, Rhodes *et al.*, 2004, Rhodes *et al.*, 2007a). This approach specifically aims to combine independent and heterogeneous microarray studies by combining the summary statistics from each individual study, where the commonly used summary statistics are significance levels (p values) and effect sizes. These summary statistics are combined across different studies to estimate the overall summary statistic. When differing microarray platforms are analysed together, then this is the only realistic approach. In this study although different GeneChip generations were used and analysed together, they were all from a single source (Affymetrix) and this allowed cross comparison using the homology table function of GeneSpring. This function automates the process of building homology tables for different generations of chips, even across species. Within GeneSpring experiments or genelists can be translated, and in the case of the former these can be normalized together, allowing for statistical analyses to be performed.

3.4.4 Microarray validation

Not only does this study provide an analysis of differential gene expression (Figure 31) in normal and cancer prostate epithelium, it has also used some of the genes that showed higher expression in cancer and investigated their expression at the protein level in an unbiased, image based analysis. This method was developed and used previously to quantify the expression of Wnt5A protein in prostate cancer (Wang *et al.*, 2010). Unlike the conventional, non-parametric methods that rely on scoring of the observed expression in tissue arrays, this approach allows quantitation of the DAB signal (a common tool in tissue array studies) to which parametric statistical analysis could be applied. This analysis provides a proof of principle that this normal dataset could be used to identify novel prostate cancer specific genes.

MCM2 is over-expressed in prostate cancer and was proposed as a marker for proliferation in the prostate (Meng *et al.*, 2001). MCM2 was therefore used as a positive control for the tissue array experiment. MCM2 is a member of the mini chromosome maintenance nuclear protein family, which consists of six major

isoforms (MCM2-7) that are essential in the initiation and regulation of DNA replication (Tanaka *et al.*, 1997). MCM2 is considered to be a putative marker for tumorigenicity in many tissues, such as lung (Ramnath *et al.*, 2001), colon (Giaginis *et al.*, 2009, Hanna-Morris *et al.*, 2009) and also as a predictor of recurrence for some bladder cancers (Burger *et al.*, 2007). Although MCM2 is thought to be a marker of proliferation in the prostate (Meng *et al.*, 2001, Ananthanarayanan *et al.*, 2006) no quantitative information exists on its expression in prostate cancer. I analyzed the expression of MCM2 in our prostate tissue array and found that expression of MCM2 was increased by ~2.3 fold in malignant prostate cores compared to non-malignant cores ($p<0.001$). These result support previous findings of increased MCM2 expression in prostate cancer.

Nuclear receptor subfamily 1, group D, member 1 (NR1D1 or Rev-erb) (Lazar *et al.*, 1990) belongs to a superfamily of nuclear receptors (O'Malley and Conneely, 1992) and is known to play an integral role in regulating circadian rhythm (Preitner *et al.*, 2002, Casey *et al.*, 2009). NR1D1 has been also been implicated in lipid metabolism, metal ion and heme binding (Yin *et al.*, 2007, Burris, 2008). Very recently, NR1D1 and peroxisome proliferator activated receptor γ binding protein (PBP) have been identified as survival factors for breast cancer cells with ERBB2 signature (an adverse prognostic marker for breast cancer) (Kourtidis *et al.*, 2010), where it regulates enzymes (malate dehydrogenase 1 and malic enzyme 1) that link glycolysis and lipid synthesis. No information exists on its expression or role in prostate cancer. These results demonstrate, for the first time, that the gene and protein expression of NR1D1 is increased in prostate cancer (Figures 33, 34).

ABCA1 is a member of the superfamily of ATP-binding cassette (ABC) transporter proteins (Luciani *et al.*, 1994). This superfamily is one of the largest protein families conserved through evolution. ABCA1 belongs to the ABC1 subfamily and like many proteins belonging to this subfamily is involved in lipid metabolism, particularly high density lipid formation (van Meer *et al.*, 2006). To my knowledge, the expression of ABCA1 has not been investigated in prostate tissue, previously. However, because ABCA1 are a key regulator of lipid homeostasis and androgens regulate the growth of prostate tumor (Huggins, 1967) the role of this protein has been investigated in prostate cancer cell line LnCaP (Fukuchi *et al.*, 2004, Chuu *et al.*, 2006). Using these

model cell lines, it has been postulated that ABCA1 protein may be involved in prostate cancer progression (Fukuchi *et al.*, 2004). This study shows, for the first time, that the expression of ABCA1, at the gene (Supplemental Figure 2, file 4A) and protein level (Figure 33, 34) is increased in prostate cancer. Numerous other members of ABC1 transporter family were also up-regulated at the gene expression level in my dataset (Supplemental Figure 3A, file 4A), perhaps indicating a key role for ABCA1 in particular and ABC1 family in general in prostate cancer.

3.4.5 Drawbacks to techniques used in this study

Tissue procurement and handling: human prostates were retrieved at the time of donor organ harvest, and inevitably there was a delay from time of aortic clamping to snap freezing in liquid nitrogen prior to storage at -80^0C. Indeed these patients were brain dead prior to this and the effects of this on gene expression in tissues is not known. As discussed (section 2.1.3) this delay has not been shown to significantly impact on subsequent gene expression analyses but some RNA degradation must occur as a result. Huang et al (Huang *et al.*, 2001) demonstrated elegantly that despite apparent good quality RNA as shown by 18S / 28S bands on ethidium bromide stained gels, warm ischaemia did have a significant effect on human colon cancer tissue when characterized by microarray study. They felt that warm ischaemia of greater than 20 minutes should be avoided. Ideally tissue in this study would have been snap frozen immediately at the time of harvest.

Sample size: the limited number of samples in this study reflects the difficulty in obtaining tissue for research purposes from organ donors. Formal sample size and power calculations were therefore not performed as it was accepted that tissue availability would govern the number of experiments undertaken.

RNA degradation and amplification: the reliability of microarrays to detect transcriptional differences representative of original samples is affected by several factors such as array production, RNA extraction, probe labeling, hybridization conditions and image analysis (Schuchhardt *et al.*, 2000). One of the limiting factors for obtaining meaningful gene expression data is the quality of the initial RNA preparation. Processing steps of tissue in this study included cryosectioning, staining and LCM, prior to RNA isolation. There was a balance that needed to be found

between microdissection times, and number of cells isolated. Initially results were poor as dissection times were too long, with RNA degradation. With shorter times, came better RNA quality, but this was at the expense of smaller quantities of RNA. Twenty-minute sessions were eventually settled upon. All these steps provide potential for RNA degradation. I used the Agilent Bioanalyser to quantify and qualify RNA prior to and following amplification as recommended by the 'Tumor Analysis Best Practices Working Group' (Group, 2004). They recommend clear 18S and 28S ribosomal bands (Figure 17) with ratios of 1.8 – 2.1 (Table 2). As seen there is subsequent degradation, however when tested for RIN number (section 2.3.4) (Imbeaud *et al.*, 2005), the majority are suitable for microarray analysis.

The 'Tumor Analysis Best Practices Working Group' (Group, 2004) also recommends a two-round amplification protocol when RNA amount is limiting, as was the case with LCM. They recommend that aRNA should be between 500 – 3,000 bp. Figures 20 and 22 demonstrate that the amplified RNA in this study was indeed of the appropriate size. There is concern that bias maybe introduced during the amplification process, resulting in significant changes in gene expression profiles with microarray studies. T7 linear amplification (section 2.1.5) using random primers results in a 3' bias to the aRNA. This problem may become even more prominent when RNA amplification is applied to clinical samples because of poorer RNA quality resulting from reduced mRNA content and losses associated with sample handling and processing. Several authors have tried to address this dilemma in prostate and other tissues (Luzzi *et al.*, 2003, Kube *et al.*, 2007, Ding *et al.*, 2006, Li *et al.*, 2005, Diboun *et al.*, 2006, Stoyanova *et al.*, 2004). An optimal mRNA amplification method should provide reproducible results, maintain fidelity of gene expression profile compared with non-amplified controls, and retain capacity to discriminate differences in gene expression profiles between two different samples. All these studies were able to successfully demonstrate two rounds of linear amplification, with successful microarray hybridization, and fulfill the above criteria, especially being reproducible. The use of Affymetrix 3' expression arrays in this study helps to eliminate this bias as the targets are designed from the 3' end (Luzzi *et al.*, 2003), unlike their exon arrays (designed to match exons).

Li *et al.* (Li *et al.*, 2005) sounds a note of caution however about amplification. In their study they used four differing protocols for amplification: Baugh's modified protocol, a modified Affymetrix protocol, a standard Affymetrix protocol, and an

Arcturus RNA amplification Kit. They all performed differently when applied to the amplification of small RNA samples from clinical specimens. Small differences in methodology and materials introduced considerable variability in gene expression profiling results. They felt that too strong a focus on a very small number of genes picked from an array analysis could be influenced by choice of kit and laboratory practice.
Thus both RNA quality at the outset and following amplification is crucial for reproducible microarray experiments. In this study RNA quality was adequate but given the above it highlights the importance of downstream validation.

Microarray chips: The quality control performed on the microarray data revealed a problem with one of the chips – E transition zone (Figure 23). The Dat images revealed a defect likely to be a design flaw (admitted by Affymetrix, with another offered free of charge). However the aRNA for this sample had already been used with no sample left. The effect that this may have had on the overall analysis is unclear, and whilst it did not affect any other quality control parameters (Figures 24, 25), it is certain that some targets will be missing from this chip.

After considering the drawbacks, it is important to iterate one point. Usefulness of any disease biomarker is enhanced, if it is differentially regulated at both gene and protein level. My results show that despite the drawbacks (e.g. those associated with RNA amplification and degradation or tissue handling), these have not resulted in a major adverse impact on the overall study. For example, various genes identified in the gene expression analysis showed similar levels of change when tested using real time PCR. Furthermore, the ultimate validation of a dysregulated gene marker is whether the protein expression is altered. At least for 3 proteins, tested on a large-scale prostate tissue array, this appears to be the case. Since the conclusion of my work described in this thesis, my supervisor's group has conducted further investigations using the same tissue array for 5 other proteins, the genes of which genes were identified in my study. These have also shown increased protein expression in cancer compared to non-malignant tissue. These observations suggest that the results presented using the techniques described here may prove useful, despite the drawbacks discussed in this section.

3.5 Conclusions

The results described in this chapter provide the first global gene expression profile of normal human prostate epithelial cells. Comparative analysis of this normal, prostate epithelial only data with previously published whole tissue normal or non-malignant prostate datasets, demonstrates major distortions in epithelial specific gene expression profile. By further comparative analysis of my normal prostate epithelial dataset with an epithelial only prostate cancer dataset, I have been able to identify, numerous, prostate adenocarcinoma specific genes. Some of these genes were investigated for protein expression using a prostate tissue array using an unbiased, quantitative protein expression analysis method. These investigations yield two novel protein markers, ABCA1 and NR1D1, of prostate cancer. I believe that this normal prostate epithelial gene expression profile could be used as an important comparative tool to identify and characterize new gene and protein markers of prostatic diseases particularly adenocarcinoma.

3.6 Summary points

1. This analysis provides the first normal epithelial transcriptome of human prostate, which was compared with whole normal prostate to reveal differences in gene expression
2. Comparison with a LCM prostate cancer dataset confirms many genes previously delineated by microarray studies, but also reveals new targets including: YWHAE, GLO1, ZNF143, ABCA1, and NR1D1
3. Three genes (NR1D1, ABCA1, and MCM2) have been validated on a tissue array at the protein level, with quantification performed using a semi-automated particle analysis (Analyze Particles) protocol with ImageJ software
4. Validation at mRNA level was performed using a low density real time PCR method

Chapter 4 Zonal variation in epithelial gene expression in the human prostate

4.1 Introduction

The human prostate is divided into different regions or tightly fused zones (McNeal, 1981), macroscopically, namely peripheral, central and transition zones (section 1.1). This morphology is of clinical usefulness in the development of age-associated conditions such as benign prostatic hypertrophy (BPH) and prostate cancer (Wein *et al.*, 2006). BPH is a non-malignant overgrowth that appears to arise exclusively within the transition zone (McNeal, 1988). Prostate cancer is a multi-focal entity and is mainly confined to the peripheral zone (Brossner *et al.*, 2003).

Approximately 70% of the glandular tissue of the prostate is found in the peripheral zone and some 5% in the transition zone; the remaining 25% occurs in a third zone, the central zone (McNeal, 1981). Prostate cancer rarely originates in the central zone but the reasons for this remain unclear. Approximately 20% of prostate cancers arise in the transition zone, and historically these have had a better prognosis (Noguchi *et al.*, 2000).

Prostate cancer is the most common cancer in men in the UK – it accounts for a quarter (24%) of all new male cancer diagnoses (www.statistics.gov.uk/downloads/theme_health/MB1-38/MB1_No38_2007.pdf.). In 2007, there were 36,101 new cases of prostate cancer diagnosed in the UK, and the lifetime risk of being diagnosed with prostate cancer is 1 in 10 for men in the UK. Histologically distinguishable BPH is present in about 8% of men aged 31 to 40 years, and this prevalence increases markedly with age to about 90% by the ninth decade of life (Berry *et al.*, 1984), establishing BPH as a chronic disease that spans decades. Not all men with BPH will go on to develop lower urinary tract symptoms (LUTS) requiring treatment. Results from the Olmsted County Study (Chute *et al.*, 1993) showed a progressive increase in the prevalence of moderate-to-severe LUTS, rising to nearly 50% by the eighth decade of life. The presence of moderate-to-severe LUTS was associated with the development of acute urinary retention as a symptom of BPH progression, increasing from an incidence of 6.8 episodes per 1000 patient-years of follow-up in the overall study population, to a high of 34.7 episodes in persons 70 years or older with moderate-to-severe LUTS. Whist BPH is not a life-

threatening condition, the impact of BPH on quality of life (QOL) can be significant and along with prostate cancer represent two of the largest health problems facing men in the UK today.

Three studies to date have examined global gene expression in the human prostate. Stamey *et al.* (Stamey *et al.*, 2003) attempted to address which prostatic zone was the best control for gene expression analysis of prostate cancer. Gleason grade 4/5 prostate cancer was used to make comparisons for over expressed and under expressed genes, with peripheral, central, and transition zones. This study was prompted by concerns that prior investigations had used BPH, in itself a disease, and normal adjacent prostate tissue, often atrophied or dysplastic, as non-cancerous control tissue. Affymetrix GeneChip HuGeneFL (Affymetrix, Inc.) arrays (6,800 genes) were used, with some targets validated by real-time PCR. The authors concluded that central zone was the preferred control tissue of choice, with BPH second. Peripheral zone was a poor control tissue that appeared too similar genetically to cancer to demonstrate much difference in gene expression. No inter-zonal comparisons were made however.

Van der Heul-Nieuwenhuijsen *et al.* 2006, used custom cDNA microarrays to compare the gene expression profiles of peripheral and transition zone normal whole prostate taken from 5 radical prostatectomy specimens (van der Heul-Nieuwenhuijsen *et al.*, 2006). 346 differentially expressed genes were identified with 199 more highly expressed in the peripheral zone and 147 more highly expressed in the transition zone. They also compared these gene lists with gene lists generated from prostate cancer microarray studies (Singh *et al.*, 2002, Dhanasekaran *et al.*, 2001, Lapointe *et al.*, 2004). They found a prominent overlap between genes expressed in their 'normal' prostate PZ and genes over expressed in prostate cancer. In addition they were able to extract a list of zonal-specific genes from the study of Stamey *et al.* (Stamey *et al.*, 2003). After accounting for differing gene chip design, they identified 18 similar zonal specific genes, of which 12 showed similar expression between the 2 studies. To try and address whether their differential gene expression reflected differing cell type distributions, they report that they quantified the amount of stromal and epithelial cells in frozen sections of both prostate zones. These authors (van der Heul-Nieuwenhuijsen *et al.*, 2006) also suggested that the stromal / epithelial ratio was similar for each zone, however no method or results are given in the paper. They also used RT-PCR with markers for most common cell types (two stromal markers -

fibronectin, myosin, and two epithelial markers - keratin 5, keratin 8) and found no differences in the gene expression between the zones.

Noel *et al.* 2008, used Affymetrix HG-U133 Plus 2 GeneChips to compare differential gene expression between normal peripheral and transition zone tissues obtained from radical prostatectomy specimens for prostate cancer (Noel *et al.*, 2008). They hoped to identify genes whose zonal-specific preferential expression might be associated with susceptibility or resistance to prostate cancer. Whole tissue specimens were used for RNA extraction from 3 prostates. Forty-three genes were identified as differentially regulated in the peripheral zone compared with the transition zone, 33 under expressed and 12 over expressed. Genes associated with neurogenesis development (e.g. GREM1), signal transduction (e.g. SFRP4), embryo implantation and cell adhesion (e.g. CHL1) were expressed at a higher level in the peripheral zone. Those over expressed in the transition zone were associated with neurogenesis development (e.g. ZFHX1B), signal transduction (e.g. NELL2), cell motility (e.g. the S100 calcium-binding protein of S100A4), and development (e.g. BMP5). The authors discussed their findings but were unable to correlate their findings with disease predisposition.

There is a need to elucidate the normal zonal epithelial gene expression profile of human prostate. Achievement of this aim requires a thorough comparison of normal and diseased tissue from respective zones. The rationale for experiments described in this chapter was to provide a foundation upon which molecular predisposition of diseases to different zones could be investigated. It is hoped that these preliminary investigations will lead to insights into the mechanisms of BPH and prostate cancer, which may allow design of better biomarkers or elucidation of potential novel therapeutic targets, ultimately. This chapter provides the first normal prostatic zonal human epithelial transcriptome, and explores some genes that are differentially expressed between the zones. In an attempt to elucidate potential functional differences between zones, DAVID (Huang da *et al.*, 2007) (section 2.4.5) was used to map differences in Gene Ontologies, functional clusters and pathways.

4.2 Materials and methods

Specimens were obtained, stained, sectioned, and the epithelial compartment laser microdissected as previously described (Chapter 2). Samples were hybridized to Affymetrix HG-U133 Plus 2.0 microarrays, with subsequent normalization and

quality control performed on generated CEL files in GeneSpring 7.2 (Chapter 2). Hierarchical clustering was performed, either by condition or by gene, using the centroid clustering method (Chapter 2). Differential gene expression was calculated between prostatic zones using a parametric Welch t-test, p-value cut-off 0.05 (GeneSpring 7.2). Benjamini and Hochberg false discovery rate correction was applied with cut-off 0.1. Filtering on fold change was applied to reduce the size of datasets when required. Gene lists with up and down-regulated targets were produced, and analyzed for biological meaning using a web-based software, DAVID 2008 (Database for Annotation, Visualization and Integrated Discovery) (Dennis *et al.*, 2003), using Fischer exact test. Validation of the microarray data by real-time PCR was described in Chapter 3.3.2.

4.2.1 Immunohistochemistry

Sections were cut 4-8 um thick on a cryostat and mounted on superfrost plus slides and stored at - 80 º C. Before staining, slides were warmed at room temperature for 30 minutes, and then fixed in ice cold acetone for 5 minutes. They were allowed to dry in air dry for 30 minutes before being washed in phosphate buffered saline (PBS). Tissue staining was as follows:

1. Sections were rinsed in washing buffer for 2x2 min.
2. Serum blocking: sections were incubated in rabbit blocking serum.
3. Primary Antibody: sections were incubated in TGM4 primary antibody (Stratech) at 1:125 dilution in primary antibody dilution buffer for 1 hour at room temperature.
4. Rinsing was performed in washing buffer for 3x2 min.
5. Secondary Antibody: sections were incubated in biotinylated secondary antibody (1:125, goat anti rabbit FITC labeled) in secondary antibody dilution buffer for 30 minutes at room temperature.
6. Rinsing was performed in washing buffer for 3x2 min.
7. Detection: sections were incubated in FITC-Avidin D (1:500, Vector Labs) in PBS for 30 minutes at room temperature. Slides were covered with aluminum foil to protect them from light.
8. Rinsing was performed in washing buffer for 3x2 min.
9. Counterstaining with DAPI was done for 30 minutes at room temperature.
10. Slides were rinsed in washing buffer for 3x2 min.

11. A coverslip was mounted and sealed with nail polish.
12. Slides were stored in the dark at 4 ºC.

(Blocking solution, primary and secondary antibody dilution buffer, and wash buffer constituents can be found in Appendix)

To view slides they were examined using a Nikon Diaphot 200 microscope (Nikon, UK) with epifluorescence filter-set for DAPI and FITC (Omega Optical, USA). To record immunofluorescence, the nuclear staining was photographed separately from the FITC-staining for the identical field and a composite picture generated using Photoshop software (Adobe, USA). FITC-detection for comparative fields were photographed with identical parameters (exposure, brightness etc) using a Nikon DXM1200 digital camera and ACT1 software (Nikon, UK). Representative pictures from 3-5 individual experiments are shown.

4.3 Results

4.3.1 Exploratory unsupervised analysis

As a first approach to data analysis, unsupervised learning techniques have been widely applied to find groups of either samples or co-regulated genes on microarray data. Hierarchical cluster analysis (Appendix) is a statistical method to group samples (or genes) unsupervised in different clusters or branches of the hierarchical tree. In this way, the relationships between the different groups are shown (Eisen *et al.*, 1998). It identifies sets of correlated genes or samples, with similar behavior across the experiments, but may yield thousands of clusters in a tree-like structure (Eisen *et al.*, 1998). Principal components analysis (PCA) is an unsupervised decomposition method to reduce multidimensional data into three dimensions. Each dimension represents a principal component with a certain percentage of variance. PCA recognizes patterns and clusters in data sets in multiple dimensions, such as gene expression data, and is therefore a valuable visualisation and summarizing tool (Holter *et al.*, 2000).

The normalised microarray data (section 2.4.1) were initially subjected to a principal components analysis. The PCA algorithm in GeneSpring GX 7.2 was applied to all 8 samples, using the ‘all genes’ (54,675) gene list for the Affymetrix HG U133 Plus 2 genome. Using three principal components (x, y and z), there were no zonal similarities seen, but prostate 301104 did group differently from the others (Figure 35).

Subsequently, hierarchical clustering (all samples) was applied to the data files in GeneSpring. The condition tree was displayed as a heat map (Figure 36), based on the measured intensities (expression levels) of the probe sets. Hierarchical clustering analysis (on samples) also did not show zonal clustering.

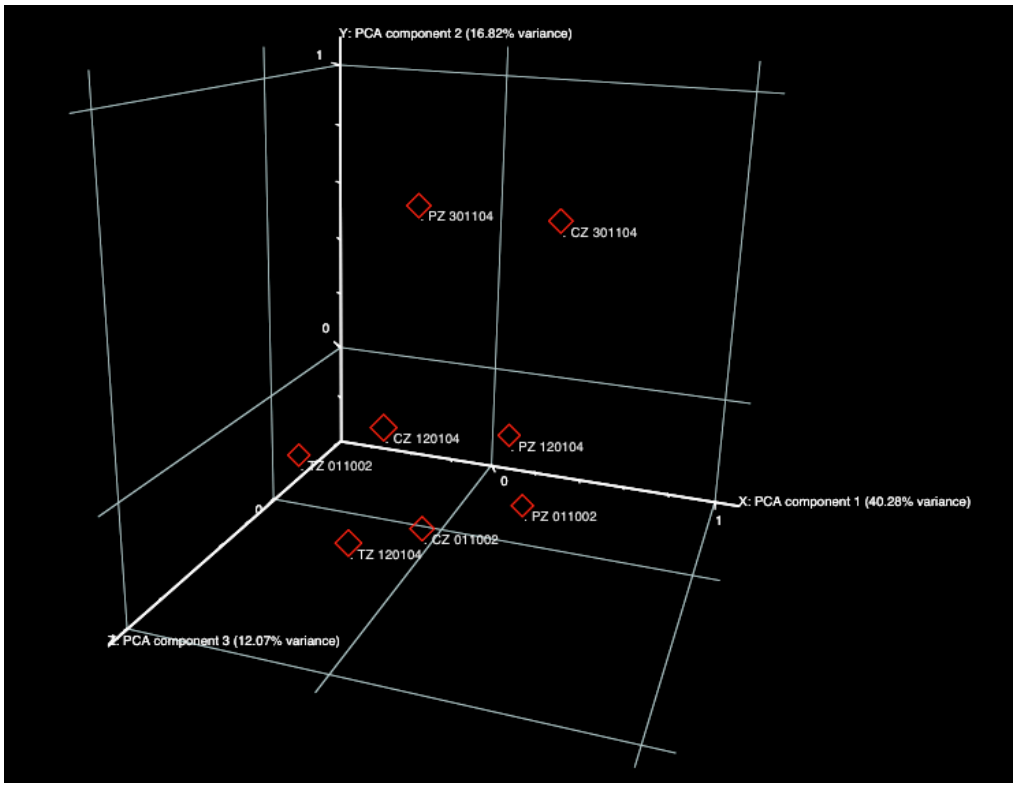


Figure 35 PCA analysis from all 8 samples. Each diamond represents a sample, characterized by the gene expression of all probe sets (54,675) on the Affymetrix HG U133 Plus 2 array. The first, second and third principal components are displayed on the X, Y and Z-axis, respectively. These three components represent the largest fraction of the overall variability.

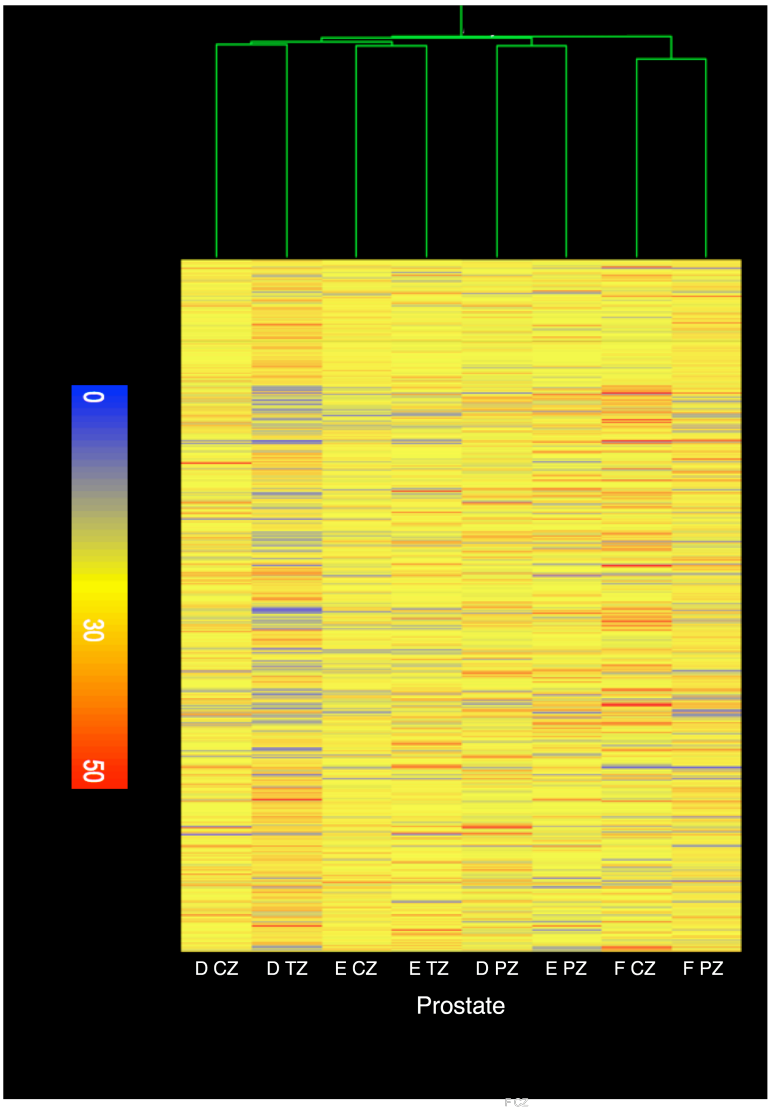


Figure 36 Hierarchical clustering by conditions (54,675 probe sets) does not show samples clustering according to prostatic zones. The condition tree is displayed as a heat map, based on the measured intensities of the probe sets.

4.3.2 Supervised analysis – epithelial zonal gene expression

As shown in section 4.3.1, samples did not cluster into zones with unsupervised techniques, and this can occur with this method especially when sample sizes are small, as is the case in this study. In order to find out which genes might be involved in classifying the zones, a supervised learning method was employed. Differential gene expression between the different zones (assigned) was calculated according to the protocol described earlier (Chapter 4.2). It was not possible to perform false discovery rate testing with the Benjamini-Hochberg method, as this revealed no

significant genes, because of the small sample size. Zonal comparisons were performed individually yielding 3 gene lists (PZ vs. CZ, PZ vs. TZ, CZ vs. TZ), containing statistically significant differential gene expression (gene lists supplemental files 5A, B and C).

4.3.2.1 Peripheral zone versus Central zone

130 differentially expressed genes were identified in CZ vs. PZ including 89 up regulated in the CZ and 41 down-regulated in the CZ (Figure 37, gene list supplemental file 5A). The 20 genes showing the greatest differential expression are shown in Table 7. Genes preferentially expressed in the PZ included lipoprotein lipase (LPL, 12.3 fold difference) and Calmodulin1 (CALM 1, 3.1 fold difference), with those preferentially expressed in the CZ including Transglutaminase 4 (TGM4, 172.1 fold difference) and Angiotensin II receptor, type 1 (AGTR1, 6.5 fold difference).

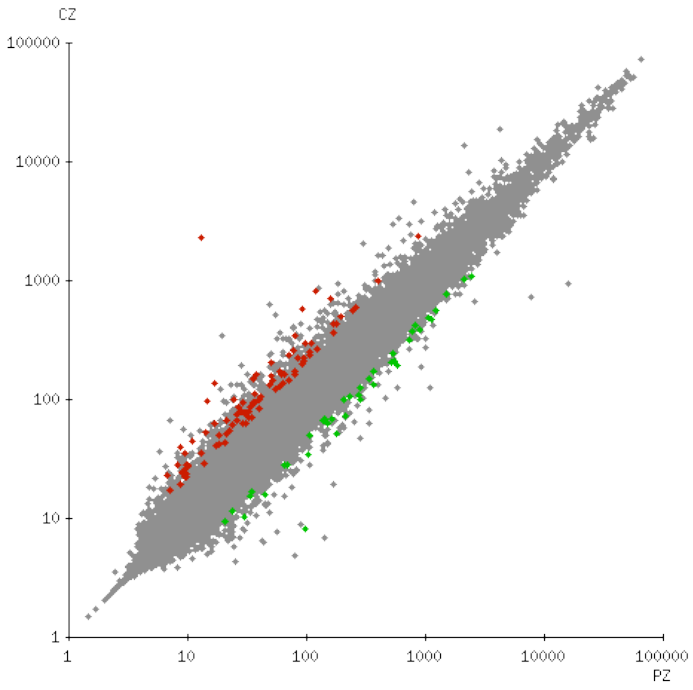


Figure 37 Two-dimensional scatter plot of differential gene expression between central zone and peripheral zone of normal prostate epithelium. Red and green show up and down-regulated targets respectively, which were statistically differently expressed (parametric Welch t-test, p-value cut-off 0.05). The X and Y-axis correspond to normalised mean intensity.

Gene Name	Gene ID	Fold change	Direction (CZ vs. PZ)	Other ID
Transglutaminase 4 (prostate)	TGM4	172.2	Up (CZ)	206260_at
Lipoprotein lipase	LPL	12.3	Down (CZ)	203549_s_at
Nipped-B homolog (Drosophila)	NIPBL	7.8	Up	242352_at
Angiotensin II receptor, type 1	AGTR1	6.6	Up	205357_s_at
Interleukin enhancer binding factor 3, 90kDa	ILF3	6.5	Up	208930_s_at
Transcribed locus	-	6.1	Up	239671_at
Family with sequence similarity 115, member A	FAM115A	4.4	Up	212979_s_at
Tumor necrosis factor receptor superfamily, member 1A	TNFRSF1A	4.4	Up	207643_s_at
Alpha thalassemia/mental retardation syndrome X-linked (RAD54 homolog, S. cerevisiae)	ATRX	4.3	Up	208859_s_at
Zinc finger protein 294	ZNF294	4.2	Up	233819_s_at
Hypothetical protein LOC9728	KIAA0256	4.1	Up	212451_at
E1A binding protein p400	EP400	4.1	Up	230629_s_at
F-box protein 32	FBXO32	4.0	Up	241762_at
Arginine/serine-rich coiled-coil 1	RSRC1	3.9	Up	219507_at
Acyl-Coenzyme A binding domain containing 5	ACBD5	3.7	Up	1568877_a_at
Protein phosphatase 1, regulatory (inhibitor) subunit 10	PPP1R10	3.7	Up	201702_s_at
GTF2I repeat domain containing 2	GTF2IRD2	3.6	Up	1557289_s_at
KIAA0101	KIAA0101	3.6	Down	202503_s_at
Family with sequence similarity 44, member A	FAM44A	3.4	Up	235009_at
Thrombospondin 1	THBS1	3.3	Up	239336_at
Transcribed locus	-	3.3	Up	229150_at

Table 7 The 20 genes showing the greatest differential expression between central and peripheral zone normal human prostatic epithelium.

4.3.2.2 Peripheral zone versus Transition zone

1185 genes were identified as differentially expressed in the TZ compared with the PZ, including 261 that were up-regulated and 924 that were down-regulated (Figure 38, supplemental file 5B). The 20 genes showing the greatest differential expression are shown in Table 8. Genes preferentially expressed in the PZ included Protocadherin 8 (PCDH8, 73.1 fold difference) and Transferrin (TF, 38.3 fold difference) with those preferentially expressed in the TZ including Collagen, type IX, alpha 1 (COL9A1, 27.6 fold difference) and Transglutaminase 4 (TGM4, 27.1 fold difference).

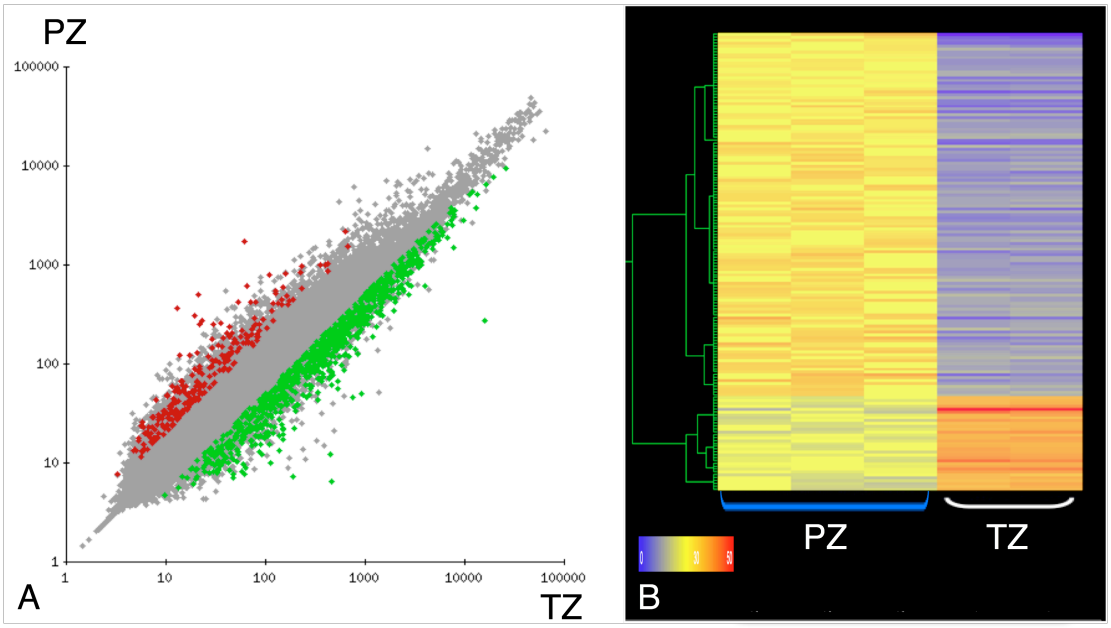


Figure 38 A Two-dimensional scatter plot of differential gene expression between transition zone and peripheral zone normal epithelium reveals 1185 differentially expressed genes (parametric Welch t-test, p-value cut-off 0.05). B Hierarchical clustering of differential gene expression between LCM peripheral and transition zone epithelium. 160 genes showed statistically significant differential expression (Welch t-test, p 0.05) at 2-fold change.

Gene Name	Gene ID	Fold change	Direction (TZ vs. PZ)	Other ID
Protocadherin 8	PCDH8	73.1	Down (TZ)	206935_at
Olfactomedin	OLFM4	59.0	Down	212768_s_at
Transferrin	TF	38.3	Down	203400_s_at
Collagen, type IX, alpha 1	COL9A1	27.6	Up (TZ)	222008_at
Transglutaminase 4 (prostate)	TGM4	27.1	Up	206260_at
Cell adhesion molecule with homology to L1CAM (close homolog of L1)	CHL1	26.3	Down	204591_at
Transcribed locus, strongly similar to XP_936141.1 PREDICTED: hypothetical protein XP_936141	-	23.1	Up	228919_at
Solute carrier family 26, member 4	SLC26A4	18.6	Down	206529_x_at
Prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)	PTGS2	16.7	Down	204748_at
Transglutaminase 4 (prostate)	TGM4	15.3	Up	217566_s_at
CDNA clone IMAGE:6025865	-	11.7	Down	212444_at
Nuclear mitotic apparatus protein 1	NUMA1	11.5	Up	214250_at
Kinesin family member 5C	KIF5C	11.2	Up	1557089_at
Plastin 1 (I isoform)	PLS1	10.5	Down	205190_at
GDP-mannose 4,6-dehydratase	GMDS	9.8	Down	204875_s_at
Odz, odd Oz/ten-m homolog 2 (Drosophila)	ODZ2	9.3	Up	231867_at
Synaptojanin 2	SYNJ2	8.9	Down	212828_at
Tumor suppressor candidate 3	TUSC3	8.9	Down	209228_x_at
Transcribed locus	-	8.7	Up	236617_at
Gastrin-releasing peptide	GRP	8.2	Down	206326_at
Proteasome (prosome, macropain) subunit, alpha type, 1	PSMA1	8.1	Down	210759_s_at

Table 8 The 20 genes showing the greatest differential expression between transition and peripheral zone normal human prostatic epithelium.

4.3.2.3 Central zone versus Transition zone

449 genes were identified as differentially expressed in the TZ compared with the CZ, including 168 up-regulated and 281 down-regulated genes (Figure 39, gene list supplemental file 5C). The 20 genes showing the greatest differential expression are shown in Table 9. Genes preferentially expressed in the CZ included Transmembrane protein 178 (TMEM178, 31.8 fold difference) and Cytochrome P450, family 1, subfamily B, polypeptide 1 (CYP1B1, 21.5 fold difference), with those preferentially expressed in the TZ including Homeobox C4 (HOXC4, 28.5 fold difference) and Cytochrome P450, family 3, subfamily A, polypeptide 5 (CYP3A5, 6.6 fold difference).

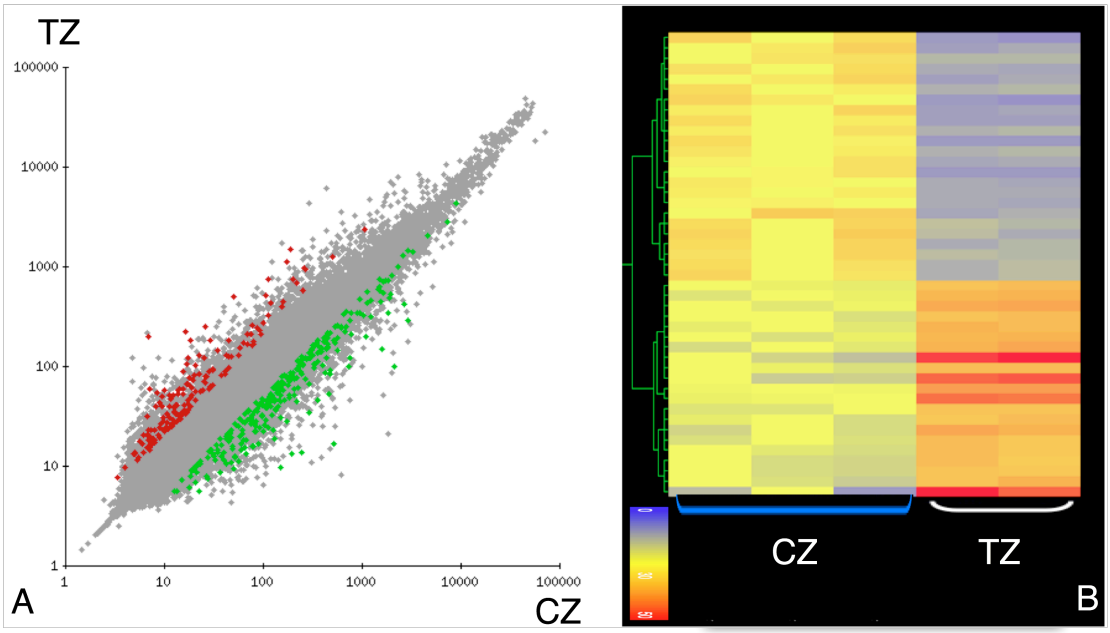


Figure 39 A Two-dimensional scatter plot of differential gene expression between transition zone and central zone normal epithelium reveals 449 differentially expressed genes (parametric Welch t-test, p-value cut-off 0.05). B Hierarchical clustering of differential gene expression between LCM central and transition zone epithelium. 45 genes showed statistically significant differential expression (Welch t-test, p 0.05) at 2-fold change.

Gene Name	Gene ID	Fold change	Direction (TZ vs. CZ)	Other ID
Transmembrane protein 178	TMEM178	31.8	Down (TZ)	229302_at
Homeobox C4	HOXC4	28.5	Up (TZ)	206858_s_at
Cytochrome P450, family 1, subfamily B, polypeptide 1	CYP1B1	21.5	Down	202437_s_at
G protein-coupled receptor, family C, group 5, member A	GPRC5A	18.3	Down	203108_at
Ceruloplasmin (ferroxidase)	CP	15.7	Down	1558034_s_at
Solute carrier family 5, member 1	SLC5A1	15.1	Down	242773_at
Transcribed locus	LP2209	13.2	Up	239860_at
Cytochrome P450, family 1, subfamily B, polypeptide 1	CYP1B1	11.0	Down	202435_s_at
Cytochrome P450, family 1, subfamily B, polypeptide 1	CYP1B1	10.3	Down	202436_s_at
Transcribed locus	-	9.8	Up	228919_at
ESTs	-	9.7	Up	239017_at
CDNA FLJ37098 fis	-	9.4	Up	222368_at
Hexokinase 2	HK2	8.9	Down	202934_at
Klotho	KL	8.6	Down	205978_at
MRNA full length insert cDNA clone EUROIMAGE 85905	-	8.1	Up	1559910_at
V-fos FBJ murine osteosarcoma viral oncogene homolog	FOS	7.9	Down	209189_at
CDNA: FLJ21228 fis, clone COL00739	-	7.9	Up	234723_x_at
Secretory leukocyte peptidase inhibitor	SLPI	7.6	Down	203021_at
Ceruloplasmin (ferroxidase)	CP	7.0	Down	227253_at
CDNA FLJ35490 fis	FLJ45482	6.8	Up	1565786_x_at
Cytochrome P450, family 3, subfamily A, polypeptide 5	CYP3A5	6.6	Up	214235_at

Table 9 The 20 genes showing the greatest differential expression between transition and central zone normal human prostatic epithelium.

4.3.3 Functional analysis – Gene Ontology and Functional Annotation Clustering

Gene lists of statistically significantly genes differing in expression between zones were uploaded into DAVID 2008 software (Dennis *et al.*, 2003). Affymetrix probe ID's were used and significant gene ontologies represented were identified using the EASE score (modified Fischer's exact test). DAVID adopts the GO vocabulary consisting of 5 different levels with level 1 being a general description while level 5 is a more detailed description for a given gene. To maximize the outcome of the analysis, all the levels of the GO vocabulary were used for the analyses. The default parameters for DAVID analyses were used: maximum EASE score / P-Value 0.1, and minimum threshold gene count 2 in this study. The threshold of EASE Score (modified Fisher Exact P-Value) for gene-enrichment analysis ranges from 0 to 1. Fisher Exact P-Value = 0 represents perfect enrichment. The recommended default setting in DAVID is 0.1, however P-Value equal or smaller than 0.05 can be considered strongly enriched in the annotation categories (Huang da *et al.*, 2009). The threshold of minimum gene counts has to be equal or greater than 0. The default is 2, i.e. you do not trust the term only having one gene involved.

Functional Annotation Clustering groups and displays similar annotations / genes together. The grouping algorithm is based on the hypothesis that similar annotations should have similar gene members. DAVID uses an agglomeration method, which groups related genes or terms into functional groups (biological modules) based on the similarity distance measure. A gene or term may participate in more than one functional group, unlike other clustering techniques such as Hierarchical, K-means, or self-organizing maps. Clusters (functional gene groups) are produced which may be viewed as a heatmap and to determine which are more significant an 'enrichment score' is allocated (Huang da *et al.*, 2007). The enrichment score is the geometric mean of the EASE Scores (Fischers exact test) associated with each enriched annotation term that belonging to the cluster. The geometric mean is a relative score instead of an absolute p value; therefore minus log transformation is applied on the geometric mean. The group enrichment scores are intended to order the relative importance of the gene groups instead of as absolute decision values. A higher score for a group indicates that the group members are involved in more important (enriched) roles. However, all gene groups are potentially interesting despite lower rankings. An enrichment score of less than 0.05 translates to 1.3 on the minus log scale.

Using the DAVID Functional Annotation Clustering Tool (Chapter 2.4.5) functional clusters (biological modules) were identified between the zones. An enrichment score of greater than 1.3 was used as the cut off (EASE significance score < 0.05 = 1.3 on minus log scale).

4.3.3.1 Peripheral zone versus central zone

Analysis reveled that 1 GO biological process and 5 GO molecular functions were significantly over-represented in peripheral zone (Table 10), but this did not correspond to significant clusters of genes or annotations when Functional Annotation Clustering was performed in DAVID (Huang da *et al.*, 2007) (section 2.4.5).

Biological process	P Value
DAVID analysis parameter = GO term level:ALL	
GO:0009056~catabolic process	0.06
Molecular function	
DAVID analysis parameter = GO term level:ALL	
GO:0016881~acid-amino acid ligase activity	0.05
GO:0008639~small protein conjugating enzyme activity	0.042
GO:0004842~ubiquitin-protein ligase activity	0.04
GO:0016879~ligase activity, forming carbon-nitrogen bonds	0.072
GO:0019787~small conjugating protein ligase activity	0.04

Table 10 Gene ontologies that were significantly over-represented in the peripheral zone compared to the central zone (Ease score / P-Value 0.1, minimum threshold gene count 2).

Twenty-nine GO biological processes, 6 GO cellular components, and 15 GO molecular functions were significantly under-represented in the peripheral zone compared with the central zone (Table 11, supplemental file 6) with two annotation clusters shown, with enrichment scores greater that 1.3 for the same comparison (Figure 40, Table 12). Both the GO ontology and annotation clusters reflect that mRNA transcription, along with increased cellular metabolic processes, were over-represented in the central zone (under-represented peripheral zone), perhaps suggesting towards differing functions.

Biological process	P Value
DAVID analysis parameter = GO term level:ALL	
GO:0045935~positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	0.07
GO:0019219~regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	0.01
GO:0043170~macromolecule metabolic process	0.01
GO:0006139~nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	5.79E-04
Cellular component	PValue
DAVID analysis parameter = GO term level:ALL	
GO:0015629~actin cytoskeleton	0.03
GO:0005622~intracellular	0.07
Molecular function	
DAVID analysis parameter = GO term level:ALL	
GO:0003779~actin binding	0.04
GO:0016301~kinase activity	0.10
GO:0051020~GTPase binding	0.05
GO:0003677~DNA binding	0.04
GO:0003676~nucleic acid binding	0.01
GO:0030528~transcription regulator activity	0.10

Table 11 Examples of gene ontologies that were significantly over-represented in the central zone compared to the peripheral zone (Ease score / P-Value 0.1, minimum threshold gene count 2). Full table may be found in supplemental file 6.

Annotation cluster	Representative annotation terms	Enrichment score
1	mRNA transcription (Figure 4.8)	1.82
2	Regulation of mRNA transcription (Figure 4.9)	1.73

Table 12 Two Functional Annotation Clusters were identified in DAVID as being significantly over-represented in human LCM central zone prostate when compared with peripheral zone.

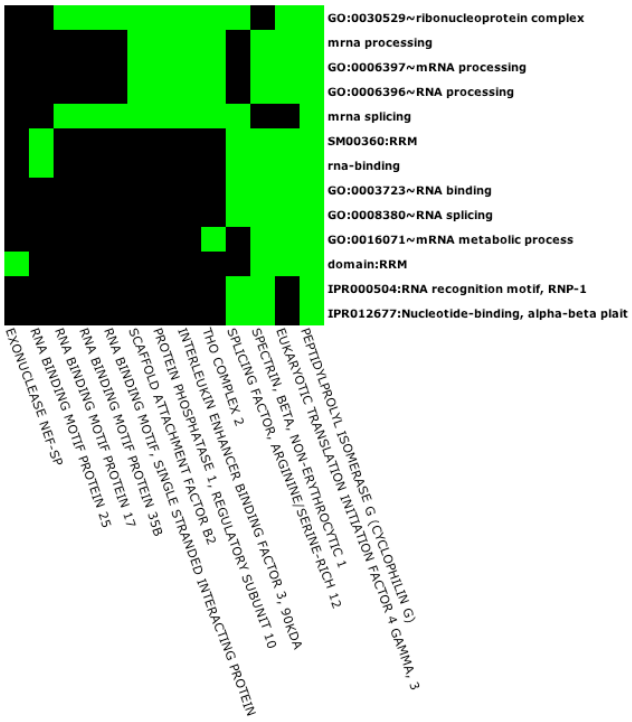


Figure 40 Functional Annotation Cluster of mRNA transcription significantly over-represented in the central zone (compared to peripheral zone). (All the related 13 mRNA transcription genes and their associated annotation terms are displayed in a 2-D heat map. Green represents the positive association between the gene-term; conversely, black represents an unknown relationship. The annotation terms are ordered based on their enrichment scores associated with the group)

4.3.3.2 Peripheral zone versus transition zone

There were numerous significantly over-represented GO ontologies in PZ compared to TZ, including 145 biological processes, such as steroid synthesis and signaling pathways, 69 cellular components such as protein and proteasome complex, and 67 molecular functions, such as enzymatic activity related to transcription and translation (supplemental file 7). Twenty-nine functional annotation clusters were significantly over represented in the peripheral zone and had enrichment scores of greater than 1.3, several are shown in Table 13 (supplemental file 8). The clusters and ontologies reveal a common theme of increased cellular activity from transcription to translation to post translational modification to protein transport and eventual protein breakdown. In young men with normal prostates the transition zone is small, and these results suggest that, when compared with the peripheral zone, it may be less active.

Annotation cluster	Representative annotation terms	Enrichment score
1	Intracellular organelle	11.73
2	Protein transport	8.61
3	Endoplasmic reticulum	8.27
4	Golgi vesicle transport / secretion	5.28
5	Proteolysis	4.02
6	Golgi apparatus	3.34
7	mRNA processing / splicing	3.28
8	Mitochondrial membrane	2.99

Table 13 Examples of Functional Annotation Clusters identified in DAVID as being significantly over-represented in human LCM peripheral zone prostate when compared with transition zone (supplemental file 8).

Over-represented GO ontologies in TZ compared with PZ included, 36 biological processes, such as regulation of progression through the cell cycle and differentiation, 10 cellular components such as spindle and microtubule, and 18 molecular functions such as actin and growth factor binding (supplemental file 9). Four annotation clusters were significantly over-represented in the transition zone including actin binding (Figure 41) and regulation of cell cycle (Table 14). These clusters and ontologies are related to cell development, cell division and cellular activity. Perhaps increased cell regulation is present in the transition zone of prostates from young men with altered cellular division as a result.

Annotation cluster	Representative annotation terms	Enrichment score
1	Actin binding	1.71
2	Nuclear hormone receptor DNA binding	1.48
3	Signal transduction	1.45
4	Regulation of cell cycle - inhibition	1.44

Table 14 Four Functional Annotation Clusters were identified in DAVID as being significantly over-represented in human LCM transition zone prostate when compared with peripheral zone.

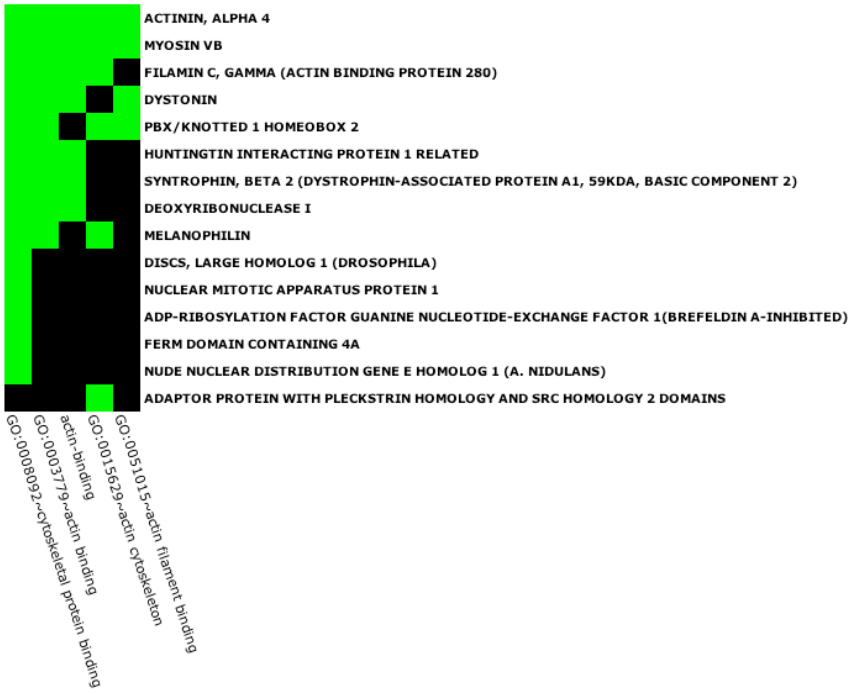


Figure 41 Functional Annotation Cluster of Actin binding significantly over-represented in the transition zone (compared to peripheral zone).

4.3.3.3 Central zone versus transition zone

Analysis revealed 30 GO biological processes that were significantly over-expressed in the CZ including translation and protein transport, 28 GO cellular components including peroxisome and Golgi apparatus, and 20 GO molecular functions including symporter and translation factor activity (supplemental file 10). Six functional clusters were significantly over-represented in the central zone (Table 15). Although less in number there seemed to be a similarity in groups under represented in the transition zone compared with both the peripheral zone and the central zone. These included intracellular organelles, protein transport and proteolysis. Again this suggest that normal human transition zone is less 'biologically active' than both its counterparts, perhaps as it has less of a secretory role than them.

GO ontologies that were significantly over-represented in the TZ included: 8 biological processes such as angiogenesis and vasculature development, 1 cellular component, the cell surface, and 7 molecular functions including transcription repressor activity and enzyme regulator activity (supplemental file 11). No clusters were identified in the transition zone containing enrichment scores of greater than 1.3.

Annotation cluster	Representative annotation terms	Enrichment score
1	Intracellular organelle	2.52
2	Organelle membrane	2.4
3	Protein transport	2.37
4	Fatty acid metabolism (peroxisome)	1.71
5	Golgi apparatus	1.33
6	Ubiquitin mediated proteolysis	1.32

Table 15 Six Functional Annotation Clusters were identified in DAVID as being significantly over-represented in human LCM central zone prostate when compared with transition zone.

4.3.4 Functional analysis – assignment of genes to pathways

Pathway assignment was carried out in DAVID using KEGG pathways (Dennis *et al.*, 2003, Kanehisa and Goto, 2000, Kanehisa *et al.*, 2006). Differentially expressed genes between zones were uploaded into DAVID for pathway inclusion and annotation to known KEGG pathways. Only annotations (genes) that were over-represented compared with random chance, as identified using EASE analysis (significance at the $P < 0.1$ level), were selected.

4.3.4.1 Peripheral versus central zone

The TGF-beta signaling pathway was the only pathway significantly differently represented between the PZ and the CZ (over-represented CZ), corresponding to 3 genes (Figure 42).

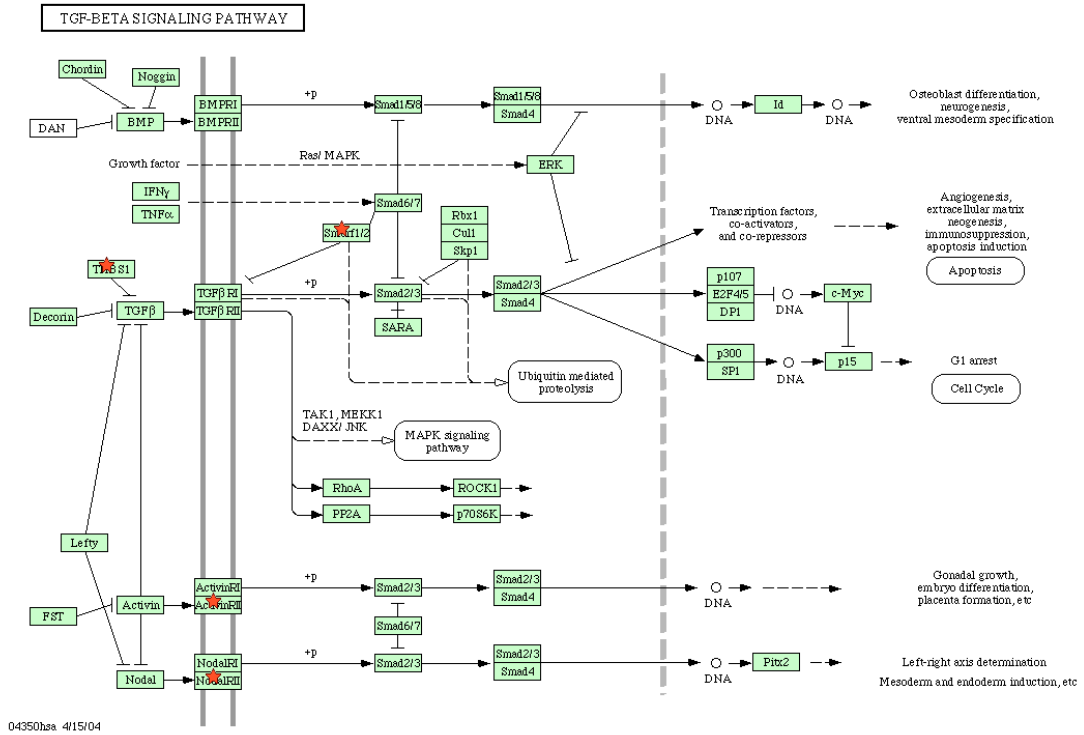


Figure 42 The TGF-beta signaling pathway was significantly over-represented in the central zone compared with the peripheral zone (figure generated by DAVID software). The three genes are highlighted with red stars, and are Activin A receptor, Type IIA, Thrombospondin 1, and SMAD specific E3 ubiquitin protein 2.

4.3.4.2 Peripheral versus transition zone

Thirteen pathways were significantly over-represented in the peripheral zone compared with the transition zone, and 8 were statistically over represented in the transition zone (Table 16).

Over-represented in PZ			
KEGG Pathway	Count	%	P-Value
hsa03060: Protein export	4	0.48%	0.02
hsa04910: Insulin signaling pathway	13	1.55%	0.06
hsa03050: Proteasome	7	0.84%	8.98E-04
hsa00512: O-Glycan biosynthesis	7	0.84%	0.01
hsa00020: Citrate cycle (TCA cycle)	5	0.60%	0.06
hsa05120: Epithelial cell signaling in Helicobacter pylori infection	9	1.07%	0.03
hsa05110: Cholera - Infection	6	0.72%	0.07
hsa01030: Glycan structures - biosynthesis 1	12	1.43%	0.05
hsa04120: Ubiquitin mediated proteolysis	13	1.55%	0.05
hsa00190: Oxidative phosphorylation	12	1.43%	0.07
hsa05010: Alzheimer's disease	7	0.84%	0.01
hsa00361: gamma-Hexachlorocyclohexane degradation	5	0.60%	0.03
hsa00100: Biosynthesis of steroids	6	0.72%	0.01
Over-represented in TZ			
hsa04510: Focal adhesion	8	3.65%	0.01
hsa04010: MAPK signaling pathway	8	3.65%	0.03
hsa04730: Long-term depression	4	1.83%	0.04
hsa04670: Leukocyte transendothelial migration	5	2.28%	0.04
hsa04520: Adherens junction	4	1.83%	0.06
hsa04360: Axon guidance	5	2.28%	0.06
hsa04120: Ubiquitin mediated proteolysis	5	2.28%	0.07
hsa04960: Aldosterone-regulated sodium reabsorption	3	1.37%	0.08

Table 16 Thirteen pathways were significantly over-represented in the peripheral zone, with 8 over-represented in the transition zone. It is surprising that an infectious pathway, Cholera, was over-represented in the peripheral zone.

4.3.4.3 Central zone versus transition zone

Four pathways were over-represented in the central zone compared with the transition zone (Table 17), whilst no pathways were over represented in the transition zone.

KEGG Pathway	Count	%	P-Value
hsa05212: Pancreatic cancer	5	1.89%	0.01
hsa05222: Small cell lung cancer	5	1.88%	0.02
hsa04120: Ubiquitin mediated proteolysis	6	2.26%	0.03
hsa05200: Pathways in cancer	9	3.38%	0.06

Table 17 Four pathways were over-represented in the central zone compared with the transition zone

4.3.5 Validation at a protein level

The expression level of one gene, TGM4, was examined at the protein level using immunohistochemistry (Figure 43). TGM4 was seen to be expressed at a higher level in prostate central zone compared with peripheral zone, and this mirrors the gene expression analysis. Differential protein expression between peripheral and transition zone was not obviously seen however, despite the significant differences seen at gene level (up transition zone 27-fold).

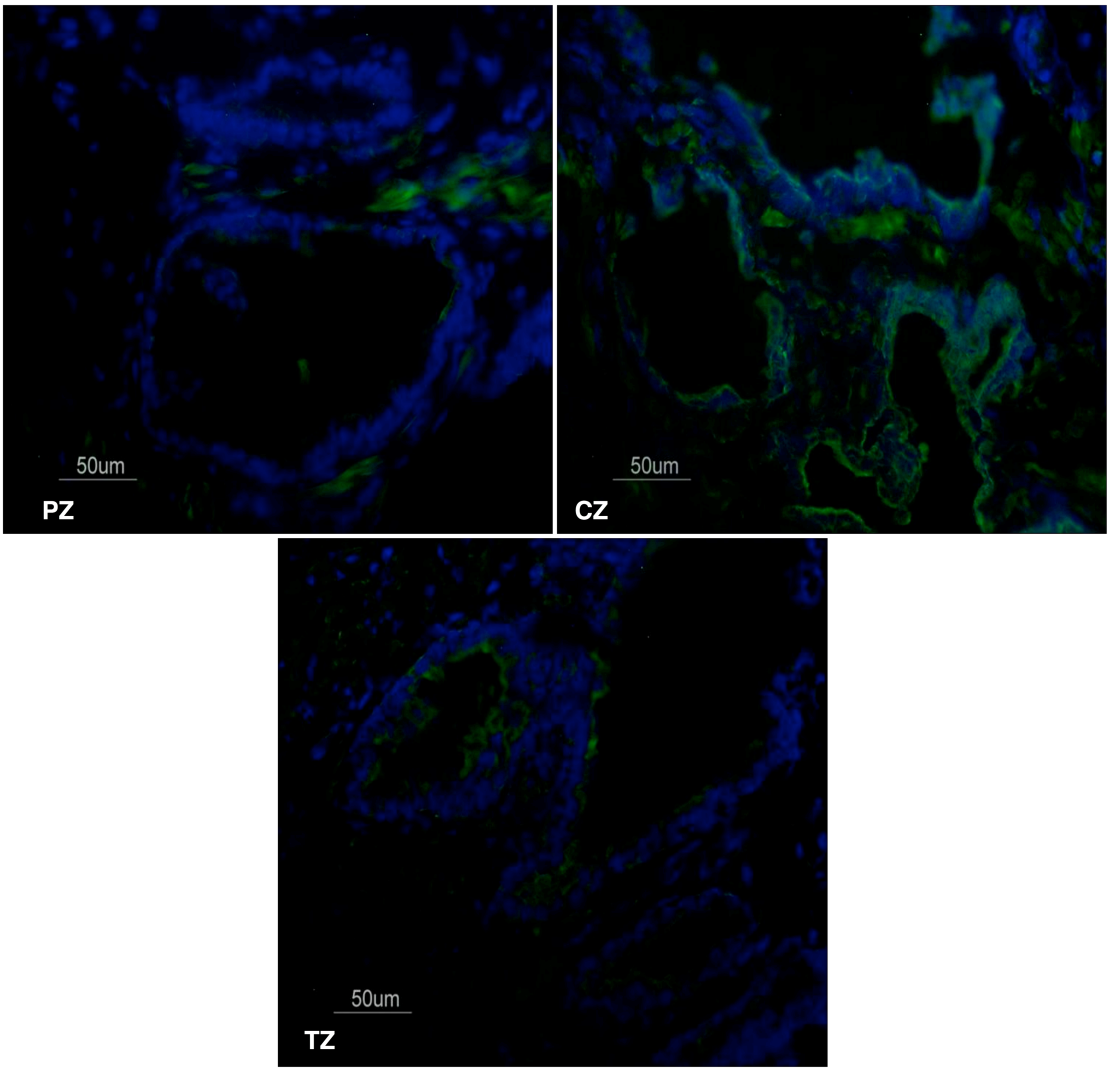


Figure 43 Immunoflourescence with anti-TGM4 antibody (green) reveals higher staining in prostate central zone than peripheral zone (DAPI – blue, FITC – green). No obvious difference was seen between peripheral and transition zone.

4.4 Discussion

Several studies have attempted to delineate global gene expression profiles of individual zones of the human prostate (Stamey *et al.*, 2003, van der Heul-Nieuwenhuijsen *et al.*, 2006, Noel *et al.*, 2008). These studies have been limited as they have used (1) whole tissue containing both epithelial and mesenchymal cell lineages and (2) prostate tissue from donors at the time of radical prostatectomy. On this second point, this tissue is unlikely to be normal, either as it reflects BPH or as a result of field effect changes in 'normal adjacent' prostate (Chandran *et al.*, 2005, Ananthanarayanan *et al.*, 2005, Ananthanarayanan *et al.*, 2006, Mehrotra *et al.*, 2008).

This is the first investigation combining laser capture microdissection on normal prostates from young men before the advent of disease with global gene expression analysis with Affymetrix microarrays to provide a zonal transcriptome map of normal human prostate.

Laser capture microdissection was used to successfully capture a pure population of epithelial cells from each of the three zones of the prostates. It was not possible to obtain any epithelium from the transition zone of prostate F (301104) as it did not contain any acini, despite sectioning of the whole sample. Pure populations of mesenchymal cells were also microdissected, with RNA subsequently extracted and stored at -80 C. Epithelial RNA was of sufficient quality to allow downstream analysis. As Affymetrix chips require large quantities of RNA two rounds of amplification were performed successfully with the resultant product hybridized to chips. Amplification issues are discussed in Chapter 2.

When performing unsupervised learning analysis on both the prostates and the zones, it was apparent both with principal component analysis and hierarchical clustering that the zones did not cluster together. If anything prostates clustered tighter reflecting the inherent biological variability. This lack of differences between the three zones can be explained by several factors. Firstly there may be more differences between individuals than between zones, and unsupervised techniques will merely show this. Secondly the zones were dissected out macroscopically by visual inspection from areas that correspond to known zonal anatomy. This is potentially an imprecise methodology, however there is not currently a better technique. Indeed it corresponds to techniques used by others when trying to assess zonal differences (Stamey *et al.*, 2003, van der Heul-Nieuwenhuijsen *et al.*, 2006, Noel *et al.*, 2008). Potentially mixed zonal tissues types could be analysed together thus skewing the results. Thirdly the exact functions of the zones are unknown, and it maybe that there are in reality few differences between the zones both functionally and genetically. Lastly unsupervised techniques are less powerful with small sample size.

To overcome this a supervised learning analysis was applied, with assignment of samples to zones, showing it is possible with differential gene expression to detect zonal differences. Differential gene expression was calculated using a parametric Welch t-test ($p<0.05$), accompanied with a false discovery testing with the Benjamini-Hochberg FDR (Benjamini and Hochberg, 1995). The significance analysis of micro-arrays (SAM) was also applied (Tusher *et al.*, 2001), however no significant

difference in gene expression was observed with either false discovery method (results not shown). Microarrays present a special challenge statistically, when genes are compared across two conditions. Given the large number of genes tested, more than 50,000 in this study, with conventional p values of for example of less than 0.05, one can expect 5% to appear significant by chance alone. This equates to over 2000 genes. False discovery testing is a way to control for this and is discussed in the Appendix. Unfortunately the design of this study, with a small sample size and number of replicates precludes false discovery testing. This reinforces the need for downstream work and validation of targets. This was performed for one gene TGM4 at a protein level with immunohistochemistry, but more work is needed if the microarray findings are to be confirmed.

Zonal comparisons were performed individually revealing more similarity between the peripheral and central zones. There were more differential gene expression differences when analyzing these zones with the transition zone. Unfortunately it was not possible to isolate RNA from one of the transition zones, as no epithelial component was seen despite sectioning of the entire sample. This may have impacted on the increased differences seen between transition and other zones. The differential gene expression reported above represents the first epithelial zonal transcriptome of normal human prostate. This is unique, as unlike other studies, it has used prostates free form disease and field effect changes, as well as having isolated pure epithelial populations.

In an attempt to mine meaningful biological information from the gene lists generated for zonal differences, GENE Ontology, functional cluster and pathway analysis were performed using a free web based software, DAVID (Dennis *et al.*, 2003, Huang da *et al.*, 2007). This method has been extensively used by researchers with more than 2000 current DAVID citations (Google Scholar – Jan 2010). Whist the exact significance of many these is unclear they provide a potential reference or avenues for future work. Several of these are now discussed along with individual genes that differed significance between prostatic zones.

4.4.1 Differences between peripheral and central zone normal prostate epithelium

TGM4 was significantly under expressed at the gene level in the peripheral zone compared with the central zone and to a lesser extent the transition zone (fold change 172.16 / 27.13). Furthermore TGM4 was expressed less at protein level in peripheral

zone than central zone, mirroring the microarray findings. TGM4 is a member of a family of enzymes that all have the ability to catalyze cross-linking of glutamine residues with lysine is all primary amines (Dubbink *et al.*, 1998). Transglutaminases all have specific tissue distribution patterns, with TGM4 found mainly in the prostate. Little is known about the function of this enzyme with most of the work having been performed on mouse and rat prostates. Some work has shown region specific expression of TGM4 both in the human and mouse prostate (Thielen *et al.*, 2007).

That TGM4 is found mainly in prostate was confirmed using ONCOMINE™ (Rhodes *et al.*, 2004), showing statistically higher expression ($p = 1.6E-9$) in human prostate compared with other tissues in 3 separate microarray studies (Figure 44) (Shyamsundar *et al.*, 2005). ONCOMINE™ is a cancer microarray database and web-based data-mining platform aimed at facilitating discovery from genome-wide expression analyses. Differential expression analyses comparing most major types of cancer with respective normal tissues as well as a variety of cancer subtypes and clinical-based and pathology-based analyses are available for exploration. Data can be queried and visualized for a selected gene across all analyses or for multiple genes in a selected analysis.

Dubbink *et al.* showed that TGM4 was expressed in the epithelial cells of a subset of human prostatic ducts (Dubbink *et al.*, 1998). They did not however comment on whether these regions corresponded to particular prostatic zones. In situ hybridisation has revealed region specific expression of TGM4 in radical prostatectomy specimens from humans, but zones were not commented on. A positive finding was that TGM4 was never expressed in prostate cancer.

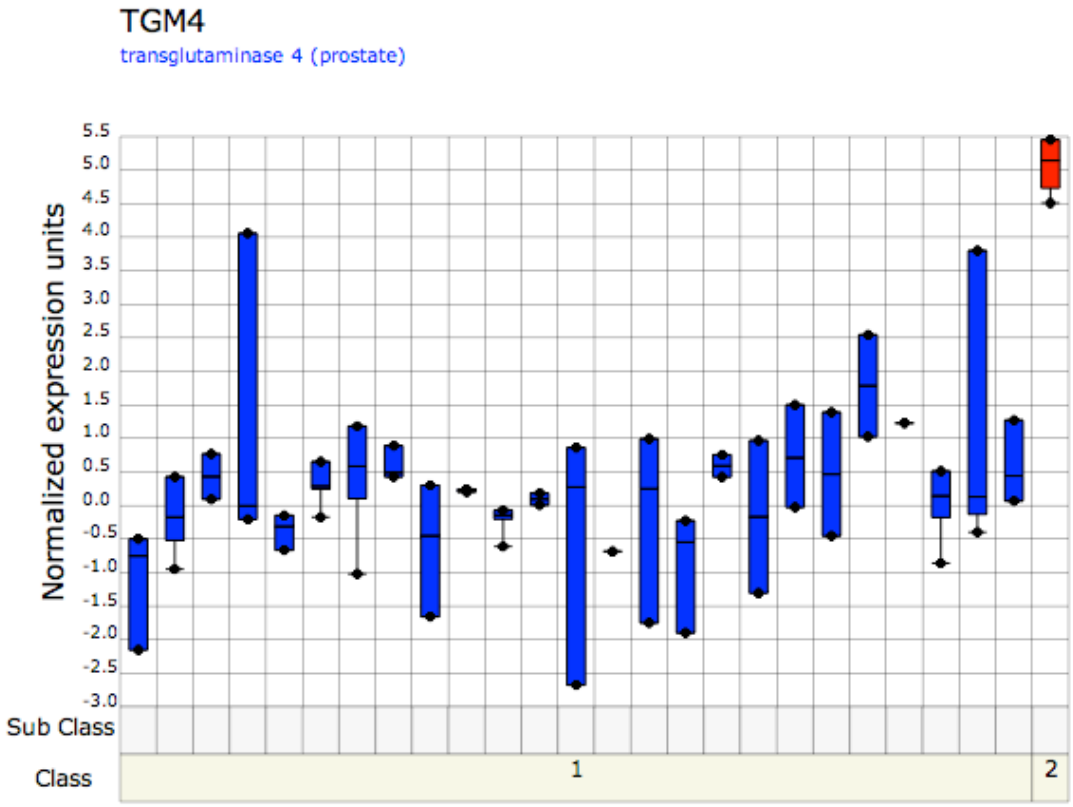


Figure 44 Representative normal human tissues (bone marrow, brain, heart, kidney, liver, lung, pancreas, prostate, skeletal muscle, spinal cord, spleen, and thymus), each pooled from 10-25 individuals, were analyzed on 5 Affymetrix Gene Chips (U95A-E) in ONCOMINE™. Column 2 represents prostate with column 1 the remaining tissues. TGM4 is significantly higher expressed in the prostate than in other human tissues.

Thelium *et al.* (Thielen *et al.*, 2007) revealed expression of TGM4 in the mouse anterior prostate, and in some areas of the mouse dorsolateral prostate, with no expression in the ventral prostate, using computer serial reconstruction and in-situ hybridization. TGM4 did not correspond identically however to the mouse lobes, but did show region specific epithelial identity. The rodent prostate is composed of four head loads situated circumferentially around the base of the bladder: the anterior prostate (AP), ventral prostate (VP), dorsal prostate (DP) and lateral prostate (LP). Each lobe has a distinct histology and secretory protein production (Hayward *et al.*, 1996). The AP is sometimes referred to as the coagulating gland, as it produces transglutaminases that assist in the formation of the copulatory plug (Hayward *et al.*, 1996). The dorsolateral prostate of the mouse has been reported as being most similar

to the peripheral zone of the human prostate based on descriptive studies (Price, 1963). More recently, inter species comparisons based on the mRNA expression signatures of the human peripheral zone and mouse anterior, dorsolateral, and ventral lobes have supported the view that the mouse dorsolateral lobe has greater similarity to the human peripheral zone than the other mouse prostatic lobes (Berquin *et al.*, 2005). Based on similar descriptive studies the mouse anterior prostate lobes (coagulating gland), which run alongside the seminal vesicles, have been postulated to be equivalent to the human central zone (Price, 1963). The high expression of TGM4 in the central zone in this study may mirror the fact that although TGM4 did not appear to match the exact anatomic boundaries of the lobes of the mouse (Thielen *et al.*, 2007), it did show a higher affinity for the anterior prostate. One might extrapolate from this that this provides further evidence of a similarity between the anterior prostate of the mouse and the central zone of the human, both in terms of function and origin.

When TGM4 is analysed in ONCOMINE™ there are significant differences in expression between benign prostate, prostate carcinoma, and hormone refractory metastatic prostate cancer in a study by Varambally looking at signatures of metastatic progression (Varambally *et al.*, 2005) (Figure 45). In a cDNA microarray study by Fujita *et al.* (Fujita *et al.*, 2008), TGM4 was in the top seven most discriminative genes, distinguishing between normal and tumoral conditions of prostate cancer. Stamey et al also found that TGM4 was in their gene list of highly significant genes that were down-regulated in Gleason 4/5 prostate cancer (Stamey *et al.*, 2003). Whether low levels of TGM4 seen in the peripheral zone in this study correspond to any disease processes is unknown, but it is a potentially novel area to be researched.

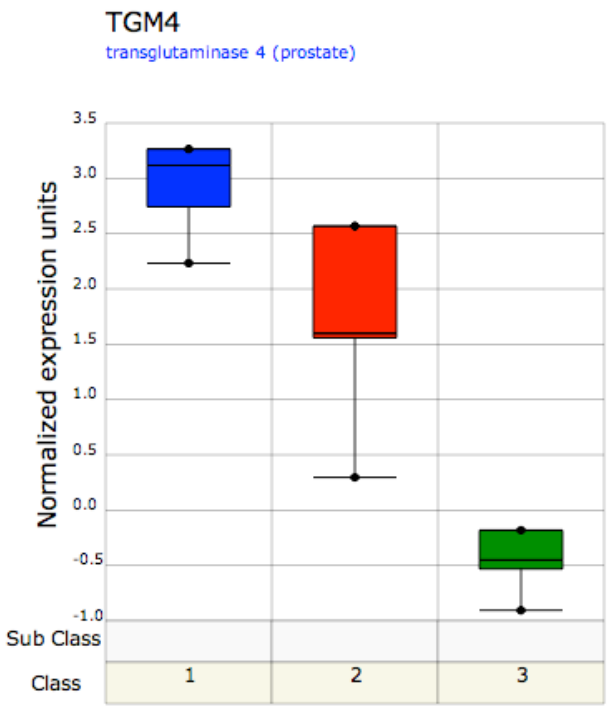


Figure 45 ONCOMINE™ data from Varambally *et al.* (Varambally *et al.*, 2005) show significant changes in expression (p = 1.6E-9) of TGM4 between benign prostatic tissue (1), primary prostate adenocarcinoma (2), and castrate-resistant metastatic prostate cancer (3).

Lipoprotein lipase (LPL) has a central role in lipid metabolism and transport, as well as having other additional non-catalytic functions (Mead *et al.*, 2002). LPL determines how dietary lipids are partitioned towards storage or utilization and therefore plays a central role in obesity and weight loss (Mead *et al.*, 2002). Abnormalities in LPL function are associated with a number of pathophysiological conditions including atherosclerosis, obesity, insulin resistance and metabolic syndrome (Mead *et al.*, 2002). High fat intake may contribute to the development and progression of prostate cancer, although this is controversial (Narita *et al.*, 2004). Metabolic syndrome (Figure 46) has been suggested as a risk factor in the development of prostate cancer, with epidemiological evidence to support this (Laukkanen *et al.*, 2004). Polymorphisms of the LPL have in addition been associated with the onset and possible progression of prostate cancer (Narita *et al.*, 2004). The LPL loci is situated on chromosome 8p22, loss of which is a frequently known chromosomal alteration seen in prostate cancer (Bova *et al.*, 1993).

IDF	NCEP	WHO	AACE
Diagnosed if glycemia is abnormal and 2 further criteria are present	Diagnosed if 3 out of 5 criteria are present	Diagnosed if glycemia is abnormal and 2 further criteria are present	Indicates risk factors
Fasting glycemia 100-125 mg/dL or DM2	Glycemia 110-125 mg/dL	Glucose intolerance, DM2 or insulin-resistance due to HOMA-IR	Fasting glycemia 110-125 mg/dL or > 140 mg/dL 2 hours after oral GTT
WC ≥ 94 cm MWC ≥ 80 cm W	WC > 102 cm MWC > 88 cm W	BMI > 30 and HWR > 0.9 M and > 0.85 W	BMI ≥ 25 and WC >102 cm M and WC > 88 cm W
Tg ≥ 150 mg/dL or HDL< 40 M and < 50 W	Tg ≥ 150 mg/dL or HDL< 40 M and < 50 W	Tg ≥ 150 mg/dL or HDL< 35 M and < 39 W	Tg ≥ 150 mg/dL or HDL< 40 M and < 50 W
On treatment for SAH or BP ≥ 130x85 mmHg	BP ≥ 130x85 mmHg	On treatment for SAH or BP ≥ 160x90 mmHg	BP ≥ 130x85 mmHg
		Microalbuminuria ≥ 20 mcg/min	

AACE = American College of Endocrinology/American Association of Clinical Endocrinologists; BMI = Body mass index; BP = arterial blood pressure; DM2 = diabetes mellitus type 2; GTT = oral glucose tolerance test; HOMA = homeostasis model assessment; HWR = hip:waist ratio; IDF = International Diabetes Federation; M = men; NCEP = US National Cholesterol Education Program; SAH = systemic arterial hypertension; Tg = triglycerides; W = women; WC = waist circumference; WHO = World Health Organization.

Figure 46 Definition of metabolic syndrome

LPL was significantly over expressed in this study in the peripheral zone of normal human prostate (fold change 12.3). Loss, mutation or epigenetic inactivation of LPL may play a role in human prostate cancer. Kim *et al.* (Kim *et al.*, 2009) found that the LPL gene is commonly methylated in prostate tumors, and that biallelic inactivation of LPL by chromosomal deletion and promoter hypermethylation may play a role in human prostate cancer. This may be one potential mechanism by which loss of LPL occurs in the normal human peripheral zone, which could be a contributory factor in carcinogenesis.

The TGF-beta pathway was significantly over-represented in the central zone in this study. TGF-beta is a potent inhibitor of cell proliferation, and in cancer cells, mutations in the TGF-beta pathway have been described that confer resistance to growth inhibition, thus allowing uncontrolled proliferation of cells (Blobe *et al.*, 2000). Thus TGF-beta pathway acts in a tumour suppressor like fashion, which may be significant in terms of risk of prostate carcinogenesis. Whether this is a potential avenue for research to explain why the central zone has a lower risk than the peripheral zone for developing prostate cancer remains to be seen.

4.4.2 Differences between peripheral and transition zone normal prostate epithelium

Olfactomedin 4 ($OLFM4^{GW112/hGC-1}$) is a glycoprotein originally described as human granulocyte colony stimulating factor-stimulated clone-1 (hGC-1) (Zhang *et al.*, 2002). It was subsequently identified as the novel anti-apoptotic molecule GW112 by

Zhang *et al.* in 2004 (Zhang *et al.*, 2004). In the most recent NCBI database, these molecules were classified as *OLFM4* (GenBank accession no. NM-006418), as they possess an olfactomedin domain. OLFM4 is localized in the nucleus and mitochondria, and inhibits activation of the caspase cascade at the level of cytochrome *c* release (Zhang *et al.*, 2004). This protein also binds to GRIM-19, a potent mediator of apoptosis, (Zhang *et al.*, 2003) to inhibit its function. Thus, OLFM4 can prevent apoptosis. However, the role of constitutively expressed OLFM4 in cancer cells is still unknown. OLFM4 is expressed at higher levels in colon, breast, and lung cancers, and has been suggested as a diagnostic marker for these cancers (Koshida *et al.*, 2007). In addition OLFM4 has been shown to promote proliferation of a human pancreatic cancer cell line (PANC-1) by favoring transition from the S to G_2/M phase. When analysed in ONCOMINE™ loss of OLFM4 expression is associated with metastatic prostate cancer (Figure 47). There was a significant difference in the expression of OLFM4 in this study favoring peripheral zone epithelium, suggesting it may act as a marker of region specific identity.

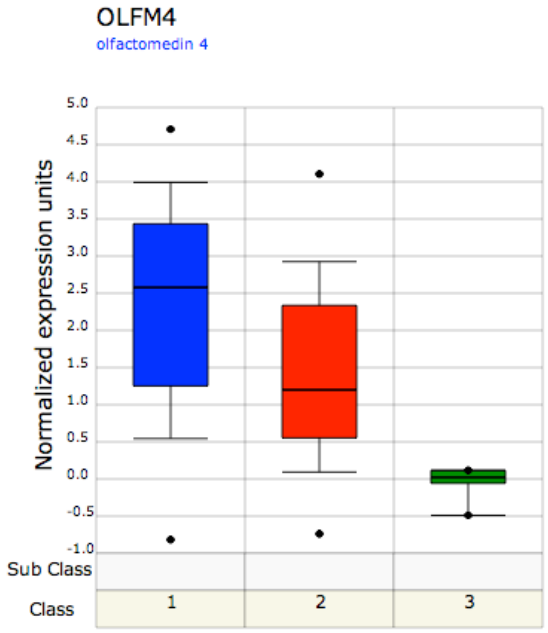


Figure 47 OLFM4 expression in Oncomine shows significant down-regulation in metastatic prostate cancer (3), compared with prostate cancer (2), and matched normal prostate (1) (p = 1.6E-9).

Other genes showing significant differences in expression between the peripheral and transition zone were Protocadherin 8 (PCDH8), Transferrin (TF), Transglutaminase 4 (TGM4), and cell adhesion molecule with homology to L1CAM (CHL1). Little is known about PCDH8, especially in the prostate, but a recent paper demonstrated that PCDH8 is mutated and epigenetically silenced in a large proportion of breast tumors, and that PCDH8 functions to suppress breast epithelial migration and proliferation (Yu *et al.*, 2008). There are significant differences in PCDH8 expression seen between the TZ and PZ, although how this translates to differing functions is unclear. Transferrin (TF) is a potent mitogen for prostate cancer (Kaighn *et al.*, 1981, Rossi and Zetter, 1992), where its levels are four times higher than in patients with benign prostatic hyperplasia and six times higher than in men with clinically normal prostates (Grayhack *et al.*, 1980). Up-regulated levels of transferrin often correlate with increased adhesion, invasion and metastasis (Bhatti *et al.*, 1997). Herman *et al.* (Herman and Meadows, 2007) examined in vitro, invasion and adhesion of stably semaphorin (sema) 3E-transfected PC-3 prostate cancer cells, in the presence and absence of TF. TF was able to reverse the decreased invasion and adhesion seen in cells transfected with sema3E, when compared to untransfected cells. TF was over-expressed in the peripheral zone is this study, although whether this contributes to disease susceptibility needs further work.

CHL1 located on human chromosome 3p26.1, is expressed in neurons and glia of both the central and peripheral nervous system, and promotes neurite outgrowth and neuronal survival; its role may be in integrin-dependent cell migration (Buhusi *et al.*, 2003). Whilst nothing is know about its function in the prostate, it was one of the genes that was significantly differentially expressed in the global zonal gene expression analysis described in Chapter 4.1 (Noel *et al.*, 2008). This expression of this target was validated by Noel *et al.* with RT-PCR and the differences are similar to my study, showing greater expression in the peripheral zone that the transition zone. The significance of this is unknown.

Several functional clusters and pathways were significantly differently represented between these zones, many to do with protein modification, transport, and breakdown, in particular involving intracellular organelles (Table 13, 14). In this study functional cluster analysis (Ubiquitin-mediated proteolysis) and pathway analysis (proteasome) showed significant differences between in protein regulatory mechanisms between the peripheral and transition zone (Table 16). Figure 48 demonstrates genes that were up-

regulated in the peripheral zone in the proteasome pathway. To identify potential genes modified in this fashion requires elucidation but as an example the androgen receptor has previously been shown to be modified in this way (Lee and Chang, 2003). The androgen receptor (AR) plays a central role in the development, progression, and treatment of prostate cancer (Richter *et al.*, 2007). Degradation of the AR plays an integral role in the regulation of AR function and is regulated by systemic protein degradation pathways, specifically the MDM2–ubiquitin–proteosome pathway (Lee and Chang, 2003).

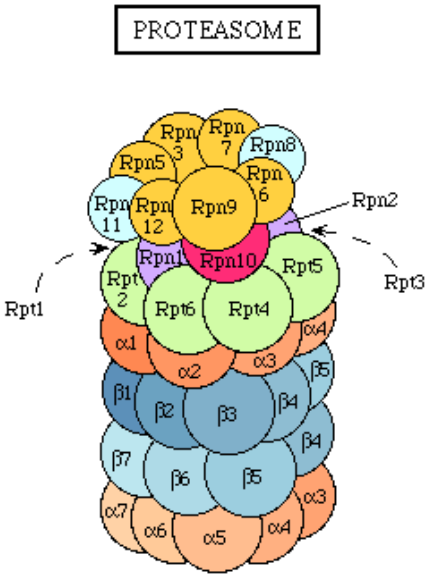


Figure 48 The KEGG Proteasome pathway (image from DAVID). This was significantly over-represented in the peripheral zone compared with the transition zone. Genes involved are marked with a red star.

Noel *et al.* (Noel *et al.*, 2008) used Gene Ontology analysis in their comparison of gene expression between peripheral and transition zone. None of their peripheral zone ontologies overlap with my findings, however there was significant correlation between transition zone findings. They found genes preferentially expressed in the

transition zone related to neurogenesis, signal transduction and development, which mirror my gene ontology and functional cluster analyses. Their study of course did not use pure cellular populations and is therefore contaminated by stroma. A recent microarray study assessed global gene differences between 'normal' and 'reactive' (histologically abnormal adjacent to prostate cancer) stroma using LCM and Affymetrix microarrays (Dakhova *et al.*, 2009). A total of 544 genes were significantly higher in the reactive stroma and 606 genes were lower. They also performed Gene ontology analysis, which revealed significant alterations in a number of processes including amongst others neurogenesis and signal transduction pathways. Why these mesenchymal processes should be over-represented in transition zone epithelium, as shown by my study, is unclear.

4.4.3 Differences between central and transition zone normal prostate epithelium

Cytochrome P450s are a multi-gene family of constitutively expressed and inducible enzymes involved in the oxidative metabolic activation and deactivation of carcinogens and cancer therapeutics (McFadyen *et al.*, 2001, Murray *et al.*, 2001). Cytochrome P450 1B1 (CYP1B1) is a member of the CYP1 gene family and one of the major enzymes involved in the hydroxylation of estrogens and activation of potential carcinogens. Differences in the capacity to metabolically activate endogenous / exogenous agents may underlie the in situ occurrence of chromosomal damage, eventually resulting in disease. The metabolites of CYP1B1 are known to induce prostate cancer in an experimental animal model (Williams *et al.*, 2000, Cavalieri *et al.*, 2002). One or more genetic variants of this gene have also been shown to be associated with an increased risk of prostate cancer (Chang *et al.*, 2003). Stamey *et al.* however found CYP1B1 to be significantly down-regulated in prostate cancer in a microarray study (Stamey *et al.*, 2003). CYP1B1 has been observed at 2 to 6-fold higher levels in human peripheral zone as compared to transition zone (Ragavan *et al.*, 2004, John *et al.*, 2009). In this study CYP1B1 was differentially expressed between the transition and central zone in this study, with higher levels in the central zone compared to the transition zone. The finding of low expression in the transition zone is mirrored by that of Van der Heul-Nieuwenhuijsen *et al.* (van der Heul-Nieuwenhuijsen *et al.*, 2006), who found low levels of CYP1B1 in the transition zone but when compared with the peripheral zone. Whether low levels of CYP1B1 in

the transition zone are significant in terms of function or disease predisposition remains to be elucidated.

In humans, HOX genes comprise the largest of several families of genes containing a sequence motif termed the homeobox. The homeobox was first identified as a motif shared among the *Drosophila* homeotic genes. HOX genes represent their human counterpart. Very little has been published on HOX gene expression in human prostate. One homebox gene, NKX3.1, is expressed in normal human prostate and has been implicated as a tumor suppressor gene (Dong, 2001). HOXC4 has been shown to be over-expressed in prostate cancer malignant cell lines, lymph node metastases, and laser microdissected prostate cancer epithelial cells (Miller *et al.*, 2003). The significance of up-regulation of this gene in the central zone in this study is unclear, but provides more information about the true epithelial transcriptome of human prostatic epithelium.

4.5 Conclusions

This analysis provides first zonal map of gene expression in normal human prostate epithelium. Several genes have been identified, including TGM4 (CZ), LPL (PZ), and CYP1B1 (TZ), as showing potential zonal region specific identity. TGM4 has also been shown to demonstrate region specific identity in mouse anterior prostate, which corresponds to human central zone. In addition some of these genes were identified by other zonal microarray studies as being region specific. This analysis also provides information about potential differing biological functions, using Gene Ontology, and pathways (KEGG) between the zones. It is however provisional work that needs further investigation and clarification, before any firm conclusions can be made. This will form the part of future work.

4.6 Summary points

1. Epithelium was isolated from prostatic zones (peripheral, central and transition) using techniques described in Chapter 2, including laser capture microdissection
2. Affymetrix microarrays were performed on RNA isolated from these as described in Chapter 2.
3. Unsupervised learning analysis did not show zonal specific clustering
4. Supervised analysis revealed potential zonal specific genes.
5. Further analysis identified potential differing functions using Gene Ontology, Functional Annotation Clusters, and pathway analysis.
6. TGM4 showed increased expression in the central zone at the gene and protein level

Chapter 5 Conclusions

This study provides the first normal gene expression profile of human prostatic epithelium free from contamination with stroma and potential prostatic disease. This has been established using a combination of laser capture microdissection and Affymetrix microarrays. Knowledge of this normal, disease free, expression profile is essential if dysregulated genes are to be identified in prostate cancer. Although numerous prostate cancer microarray studies exist describing global gene expression profiles (Dhanasekaran *et al.*, 2001, Singh *et al.*, 2002, Welsh *et al.*, 2001a, Luo *et al.*, 2001, Stuart *et al.*, 2004, Yu *et al.*, 2004), they have all used 'normal' tissue obtained from the same cancerous prostates, as control tissue. This may have a dual effect of diluting the expression of genes that are inevitably expressed in both the epithelium and the mesenchyme and exaggerating the effect of those that may be mesencyhyme specific. In addition whilst macroscopically this tissue may appear free from disease, several studies have clearly demonstrated field effect changes in normal prostate (Ananthanarayanan *et al.*, 2005, Yu *et al.*, 2004, Chandran *et al.*, 2005), as well as other tissues (Franklin *et al.*, 1997, Prevo *et al.*, 1999, Chu *et al.*, 1999, Takahashi *et al.*, 1998).

This study has demonstrated, using comparative microarray analysis with whole 'normal' prostate tissue transcriptomes (Singh *et al.*, 2002, Welsh *et al.*, 2001a, Yu *et al.*, 2004), major discrepancies between epithelial and whole tissue human prostate transcriptomes. As most microarray studies aim to provide an understanding of how gene expression changes in prostate cancer, this becomes a critical issue because prostate cancer is a carcinoma, predominantly.

In an attempt to identify some of the genetic changes in prostate cancer I compared my dataset with a prostate cancer epithelium dataset from Febbo et al (Febbo *et al.*, 2006). This was obtained using similar techniques to mine: laser capture microdissection on fresh frozen tissue, RNA amplification and Affymetrix microarray GeneChip hybridization, but without a reference, normal epithelial transcriptome for comparison. Many genes previously identified as dysregulated in prostate cancer, were confirmed using this approach (e.g. Hepsin, AMACR, PTEN, CDKN1B), as

well as several novel genes (YWHAE, GLO1, and ZNF143), some of which were subsequently confirmed at the protein level (NR1D1, ABCA1).

I also attempted to provide the first molecular map of normal epithelium based upon the macroscopic, anatomical description of McNeal (McNeal, 1981) that divides the prostate into 3 zones. Prostatic tissue was initially harvested, by visual discrimination, corresponding to the anatomical location of individual zones. Unsupervised learning techniques were used to see whether these zones grouped together. No zonal clustering was seen, rather that individual prostates tended to cluster together. Potential reasons for this include inherent biological variability between subjects, small sample number, lack of large differences in zones both at a gene and functional level, or perhaps problems with the visual technique used to ascribe zonal locality. Using supervised learning techniques, differential gene expression with zones could ascribed, gene lists were generated demonstrating potential zonal specific genes. The numbers of genes involved were small, and a false discovery rate analysis was not possible due to small sample size. In an attempt to discover biological meaning from the zones, further analysis was performed in DAVID software (Dennis *et al.*, 2003, Huang da *et al.*, 2007) using Gene Ontology, Functional Cluster Analysis, and pathway analysis. These are discussed further in Chapter 4, and it is currently unclear whether they relate to zonal predisposition to disease.

To achieve the aims of this thesis a number of technical challenges had to be overcome. Laser capture microdissection formed an integral part of obtaining pure epithelial populations free from surrounding mesenchyme. A key part of this involved sectioning and staining of prostatic tissue, ensuring that RNA degradation was kept to a minimum. Microarray studies rely on the quality of RNA at the outset, and as amplification (with potential bias) was subsequently used to generate sufficient quantities of starting material for chip hybridization, these steps were critical at the outset. Numerous cryosectioning and staining techniques were tested on both animal and human tissue, with subsequent RNA extraction and assessment, before suitable methods were settled upon. The LCM technique was also challenging as it was performed ‘off site’ requiring the transportation of frozen tissue across London. Unlike the Arcturus method (Chapter 2.1.4), the PALM microbeam laser catapults tissue into an upended eppendorf cap. Initial results were poor due to RNA degradation, until an appropriate collection medium (RLT buffer - RNeasy® Mini Kit

[Qiagen, Basel, Switzerland]) was found. As discussed in Chapter 2.3.3, a compromise also had to be made between microdissection times and the amount of tissue obtained.

There are several issues that could have been improved with this study in retrospect. Most revolve around the RNA quality and quantity. Firstly there were long delays from organ harvesting to tissue freezing at appropriate levels to prevent RNA degradation. The pitfalls with this are discussed further in Chapter 2.1.3. Secondly, some RNA degradation was apparent when samples were analysed on an Agilent bioanalyser (Chapter 2.3.4). It is likely that some gene expression has been lost / altered as a result with microarray analysis. RNA amplification is a well-established method of generating large enough quantities of starting material for microarray hybridisation, but it is also likely to generate a degree of bias, which must be considered (discussed in Chapter 2.1.5). Thirdly the comparative analyses performed in this study are likely to in themselves be subject to bias, owing to the differing chips used, differing laboratories and methods employed, all of which are known to generate potential bias. This highlights the need for further validation at both gene and protein level, something that this study has begun to address, but which needs further work.

Although not forming the original aims of this thesis, I also conducted two subsidiary projects that do not form part of the present discussion, but are provided here (attached CD), as adjunct information with regards to future work, directly arising from the work I discussed in this thesis.

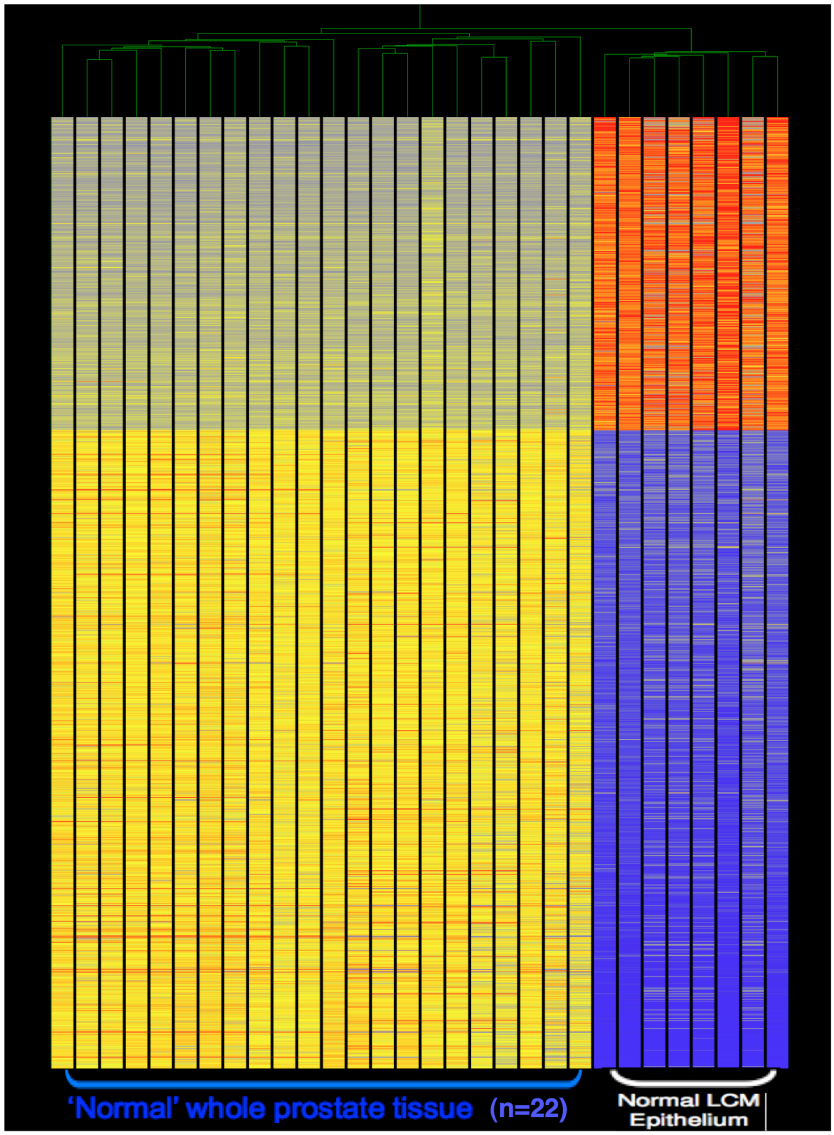
Gene expression profiling of primary cell cultures derived from prostatic zones: when this project initially started, primary cell cultures were grown from individual prostatic zones from 4 patients. RNA was subsequently extracted of sufficient quality and quantity (without amplification), with Affymetrix microarrays performed. This data has not been included in this thesis, as the results demonstrated few genetic differences between zones, when analyzed using unsupervised or supervised analysis. The reasons for this are likely to be that primary cell cultures do not reflect fresh frozen tissue. By subjecting cells to certain growth factors to encourage growth, it is likely that the underlying transcriptome is altered as a result, with any subsequent underlying differences either lost or masked. The microarray data files (CEL files) are

included with this project but due to constraints of space this data is not presented. The data however appears useful in informing the limitations of use of model systems such as cultured cells.

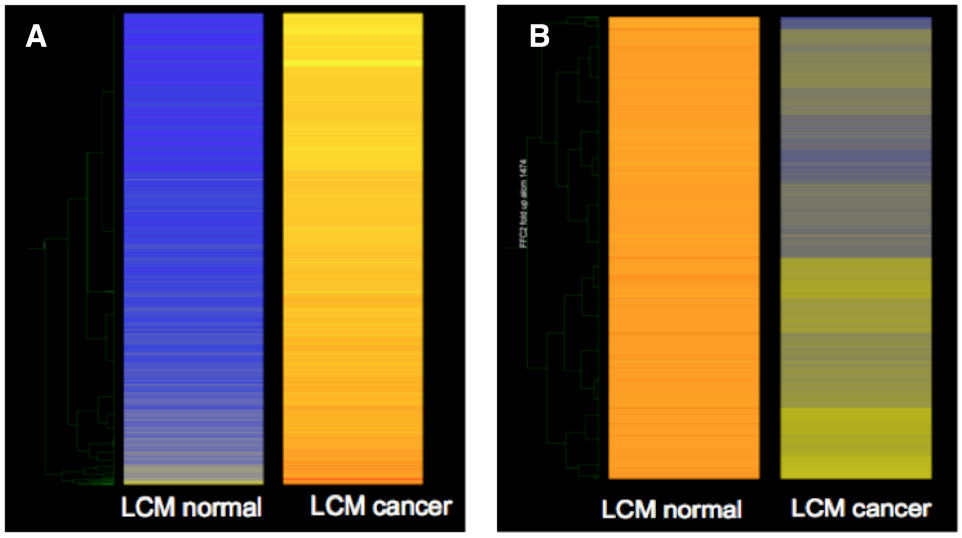
Gene expression profiling of laser microdissected normal human prostate stroma: as well as normal prostatic epithelium, stroma was also microdissected from the three prostates and three zones used in this study. RNA was isolated and amplified at the same time as epithelium, and has been stored prior to any further work. This work has not been included in this thesis but will form the basis for future microarray analysis in the same vain as this project, with subsequent comparison with a stromal prostate cancer microarray dataset (Dakhova *et al.*, 2009). Currently work is ongoing with St Georges hospital in developing a more comprehensive tissue array (as the one used in this study is exhausted), using the normal prostatic tissue obtained from this project with prostate cancer (at various stages of carcinogenesis), so meaningful prognostic and diagnostic markers for prostate cancer genes can be identified.

I believe the description of gene expression of normal prostate epithelium provided here, might facilitate the process by which prostate cancer can both be diagnosed and tackled.

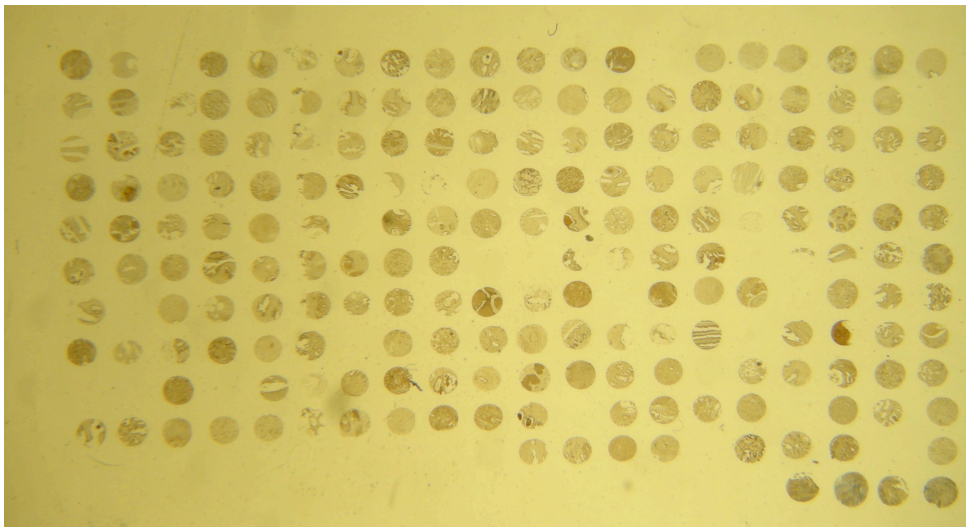
Appendix



Supplemental Gene expression profile of normal LCM epithelium vs. 'normal' whole prostate show major differences (1742 genes) with dataset from Yu et al (Yu *et al.*, 2004). Genes from all genes with statistically significant differences when grouped by 'type'; parametric test, variances not assumed equal (Welch t-test). p-value cut-off 0.05, multiple testing correction: Benjamini and Hochberg False Discovery Rate, >5-fold change). This restriction tested 12,651 genes; 1,493 genes had insufficient data for a comparison.



Supplemental Figure 2 Genes filtered on fold change (>2-fold). The figures show genes selected from condition, normal vs. cancer, that have normalized Data values that are greater or less than those in condition(s) by a factor of 2-fold. A total of 7844 genes showed increased expression (A) and 1474 genes showed decreased expression (B) in prostate cancer epithelium compared to normal epithelium. Gene lists can be found in supplemental files 4A and B.



Supplemental Figure 3 Prostate tissue microarray. A representative 6µm section is shown, cut onto coated slides and dried overnight at 60°C. (Tissue blocks were constructed using archival formalin-fixed, paraffin-embedded radical prostatectomy specimens from the 82 patients with pathological stage pT3a or b and pre-operative PSA stage of >3)

Section 4.2.1 Immunohistochemistry solutions

Normal Rabbit Serum Blocking Solution:

- 2% Rabbit serum (blocking)
- 1%BSA (stabilizer)
- 0.1% cold fish skin gelatin (blocking)
- 0.1% Triton X-100 (penetration enhancer)
- 0.05% Tween 20 (detergent and surface tension reducer)
- 0.05% sodium azide (preservative)
- 0.01M PBS, pH 7.2

Primary Antibody Dilution Buffer:

- 1%BSA (stabilizer and blocking)
- 0.1% cold fish skin gelatin (blocking)
- 0.05% sodium azide (preservative)
- 0.01M PBS pH7.2

Secondary Antibody Dilution Buffer:

- 0.01M PBS, pH 7.2
- 0.05% sodium azide (preservative)

Wash buffer

- 10X PBS-Tween 20 (0.1M PBS, 0.5% Tween 20, pH 7.2):
- Na2HPO4 (anhydrous) 10.9 g
- NaH2PO4 (anhydrous) 3.2 g
- NaCl 90 g
- Distilled water 1000 ml
- Mix to dissolve and adjust pH to 7.2 and then add 5 ml of Tween 20
- Diluted 1:10 with distilled water before use and pH adjusted as necessary.

Statistical Analysis – Overview

Background

After preprocessing, Affymetrix HGU133 Plus 2.0 GeneChips yield data about the transcriptional activity of approximately 47,000 transcripts, including 38,500 well-characterized human genes. This is comprised of more than 54,000 probe sets and 1,3000,000 distinct oligonucleotide features. These gene expression measurements are subject to random variation, largely because of the following 4 sources:

Undesired biological variability: When comparing the gene expression between two conditions e.g. tumours and normal tissue, the expression profile could be different if another patient had been analysed, or even if another portion of tissue from the same patient had been used.

Microarray manufacturing: This is truer for spotted cDNA arrays than Affymetrix arrays, where differences in cDNA clones as well as artifacts in their placement can cause discrepancies

mRNA preparation: Variation is caused by the extraction of the mRNA, its subsequent amplification and purification. Additionally the fluorescent dye binds to the genes with different efficiency

Hybridisation: An important error source is cross hybridization – mRNA that wrongly binds to non-corresponding probes on the array. Temperature, exposure time and characteristics of the mRNA solution can affect binding ability.

As a result each individual gene or target cannot be regarded as an exactly measured quantity, and data analysis has to rely on statistical methods.

Data analysis and interpretation

Sample Size: The probability of successfully identifying a differentially expressed gene (statistical power) for a fixed confidence level depends on three factors: (i) the magnitude of the true differential expression, (ii) the magnitude of random fluctuations (random noise) in the experimental system, and (iii) the number of times the experiment is replicated. The statistical power generally increases with the increase of the magnitude of the true differential expression and with the increase in the number of experimental replicates. Sample size calculations are complex in microarray analysis. To precisely determine the number of replicates needed, an estimate of total variability in the system is needed. Ideally a pilot experiment,

consisting of 2 replicates from 2 independent samples, should be performed. This allows estimation of variability in the experiment and allows sample size calculations to be performed. When there is homogeneous variability, there is a substantial benefit in using 3 rather than 2 microarrays, but not much benefit can be seen in increasing the sample size from 3 to 10 microarrays (Lee *et al.*, 2000). A minimum of three experimental replicates is therefore accepted (Lee *et al.*, 2000). When the variability is not homogenous however, a substantial improvement can be seen in each additional experimental replicate (Lee *et al.*, 2000). Ultimately, real world factors, such as availability of human samples and cost of GeneChips, may also determine the number of replicates used.

Quality control: The starting points with any microarray data analysis are quality control, normalization and filtering. The aim is to eliminate unreliable or uninformative data, and render the remainder comparable (Quackenbush, 2001). Quality is assessed at the level of RNA and chip data, with quality of RNA, efficiency of cDNA synthesis, signal to noise ratio, and hybridisation efficiency all being monitored, with cut offs enforced to exclude poorly performing chips (Hubank, 2004).

Normalisation: Since microarrays are subject to variation both at a biological, manufacturing and hybridisation level they require normalisation to reduce this variation. There are multiple differing strategies for normalisation depending on the array platform and the particular question being addressed (Bolstad *et al.*, 2003). Among the most commonly used methods for Affymetrix oligoarrays are the Affymetrix Microarray suite 5 method (MAS 5.0) (Affymetrix, 2002), the perfect match only model of Li and Wong (Li and Wong, 2001), and the Robust Multi-array average with (GCRMA) and without (RMA) correction for GC content of the oligo (Wu *et al.*, 2003, Irizarry *et al.*, 2003b). No gold standard exists, and this field is the area of great debate (Lim *et al.*, 2007). Some studies have concluded that the GCRMA / RMA method is superior in terms of sensitivity and specificity to MAS 5.0 and the method of Li and Wong (Irizarry *et al.*, 2003a). Other studies have found the Li-Wong method better at identifying networks of co-expressed genes (if large numbers of chips are used), with the GCRMA / RMA method better at detecting differentially expressed genes (Harr and Schlotterer, 2006). Ultimately normalisation is user defined but care should be taken at the outset when analyzing data.

Filtering: The Affymetrix HG U133 Plus 2.0 Array contains about 54,000 probe sets, however not all these genes will be expressed at a biological meaningful level. This large number of genes expressed at a low level adds noise to the analysis and will reduce the sensitivity of detecting differentially expressed genes. Eliminating these genes in advance is called filtering. There are a number of techniques for filtering Affymetrix data including: filtering on call (present / absent), filtering on expression levels, filtering on confidence, and filtering on variance. Once filtered this leaves a smaller set of genes for analysis greatly improving the sensitivity and reducing the false discovery rate (FDR) (see unsupervised statistical analysis).

Unsupervised Statistical Analysis: Microarrays produce an enormous amount of data, which is difficult to conceptualize. Human brains are not capable of processing numbers efficiently, and the purpose of unsupervised statistical analysis is therefore to structure the data into a form that is readable by the human eye. Using this rationale, Eisen *et al.* (Eisen *et al.*, 1998) introduced the concept of heatmaps in 1998 for displaying the results of large scale gene expression analysis. In a heatmap, each row corresponds to a gene and each column to an array (or vice versa). The representative expression value is depicted as a coloured rectangle. The colour range is traditionally depicted from green for under expressed genes, over black to red, indicating genes with higher expression (although in advance analysis software, such as GeneSpring, these are user selected), although this is not universally accepted. Both the genes and samples are often rearranged in a way that facilitates the detection of structures. It is important to remember that the heatmap is mainly for illustrative purposes.

Clustering: Clustering algorithms organize objects into a small number of groups, where the objects are more alike within one group than the objects across. With microarray data, these objects can either be genes or samples. Clustering techniques have become a frequently used standard in the genomic literature and represent an explorative form of data analysis (Slonim, 2002). They should be viewed as the start, rather than the end of a microarray study. The scope of clustering is the visualization of relationship and distance among the objects. Clustering relies on the principle that objects with unknown characteristics are assumed to share the properties of known objects that they are grouped with.

Hierarchical clustering: Clustering techniques can be divided into hierarchical and

non-hierarchical ones. The hierarchical clustering algorithm (Eisen *et al.*, 1998) is started bottom-up, with all objects as individual clusters, which are then merged. In each step, the algorithm joins the two closest clusters according to a distance matrix and a linkage method. This produces a series of clusters, which is visualized by a dendrogram. The branch length between two objects represents their degree of similarity.

K-means (non-hierarchical) clustering: The most widely used non-hierarchical clustering techniques for microarray data are k-means clustering (Hartigan and Wong, 1979), and self-organising maps (Kohonen, 1982). Both require the number of clusters to be pre-specified by the user. The k-means algorithm tries to achieve maximum purity within and minimal similarity across its clusters. It is started by assigning the samples to random clusters and iteratively moving them around until the ratio of across cluster variation to within cluster variance can no longer be improved.

Self-organising maps: Self-organising maps are a technique based upon neural networks. They usually begin with a dimension reduction step and detect similarities among objects by iteratively mapping similar input to similar regions of the output space.

Dimension reduction: Dimension reduction tools try to summarise the massive amount of data from thousands of genes into a few representative variables. While this often greatly improves the overview and handling of the data, it usually still retains most of the information. The most widely used tool is principal components analysis (PCA) (Yeung and Ruzzo, 2001). Its goal is to find a sequence of linear combinations that are uncorrelated and capture as much of the variability in the full dataset as possible. While PCA can be successful in excluding redundant information, the biological interpretation of linear combinations obtained from thousands of genes is often difficult.

Supervised Statistical Analysis: Supervised methods rely on a known, preliminary grouping of the sample, which guides through the statistical learning process. The two most important tasks are the identification of differentially expressed genes between the two conditions or phenotypes, as well as class prediction, i.e. the assignment of samples based on their gene expression patterns into known categories.

Many microarray studies are interested in the identification of genes that change their expression in different populations, under different conditions or for different

phenotypes. The purpose of these studies is usually to identify targets for further experimentation, with the hope of discovering biomarkers or therapeutic targets. When genes are compared across two conditions then there are already well-developed statistical machinery for this task, most commonly Student's or Welsh's t-test, and its non-parametric companion, the Wilcoxon test (Forrester and Ury, 1969). However microarrays present a special challenge since as thousands of tests are performed simultaneously, the non-adjusted p values are too low, the so-called multiplicity problem. For example if 50000 targets are differentially tested, using a p-value of less than 0.05, then 2500 targets will be identified as significant by chance alone. There are 'ways out', including the concept of the false discovery rate (Benjamini and Hochberg, 1995). It is an estimate of the fraction of truly altered genes among a set that is declared as significantly differential. Significance Analysis of Microarrays (SAM) (Tusher *et al.*, 2001), is one of many other methods of controlling for false positives and negatives, the scope of which lies outside of this thesis.

Class prediction: Class prediction is an important application of microarray technology. Classifiers may be built that can reliably indicate say tumour subtype, invasiveness potential, expected progression, and best treatment. Groups are defined on the basis of expression profiles or clinical distinction (or both) and then profiled using a learning algorithm such as Support Vector Machine or Nearest Neighbour analysis. This identifies the smallest set of genes that can reliably differentiate between the categories (Slonim, 2002). This type of analysis has not been undertaken in this study.

Data Mining: Finally biological meaning has to be gained from the dataset for the study to have relevance. Whilst this might be at the single gene level, diseases are often the result in imbalance in pathways and functions.

Comparative studies: Comparative studies across microarray platforms have inherent problems. Attempts to combine results across platforms assume that spot intensities or signal values for a given gene can be directly compared even though they represent different segments of the gene. That is, a spot for a given gene on a cDNA array represents the entire gene, while each spot for the same gene on an Affymetrix array represents a specific small section of the gene. Thus, combining results across

technologies using only spot intensities is problematic from a biological perspective because the measurements represent different physical quantities. Even if the average spot intensity on an Affymetrix array is used, it is not certain that this average spot intensity value is at all comparable to the spot intensity value of the gene on a cDNA array. It is not therefore recommended to directly compare different platforms, but they can be compared using a statistical approach, as employed by ONCOMINE™ (Rhodes *et al.*, 2004). ONCOMINE™ is a web-based database aimed at collecting, standardizing, analyzing, and delivering cancer transcriptome data to the biomedical research community. Raw data is imported into the software, and after analysis by grouping (e.g. cancer vs. normal), each gene is analyzed for differential expression with either Student's t-test (2 class analyses), or Pearson's correlation (multiclass). To account for multiple hypothesis testing, *Q* values (estimated false discovery rates) are calculated. Further analyses possible include co expression analysis, and molecular concepts analysis.

When the same chip type is analyzed between differing laboratories / experiments, then comparative analysis become easier, but it is important to have the raw data available so that previous downstream modification of the data (filtering, normalization etc) does not skew the results. This technique has been used by others to perform meta-analysis with publicly available microarray data from the Gene Expression Omnibus (GEO – a database repository of gene expression data and hybridization arrays) (Park and Stegall, 2007, Greco *et al.*, 2008). CEL files from different experiments are imported directly into analysis software (e.g. GeneSpring) allowing subsequent analysis. The CEL file stores the results of the intensity calculations on the pixel values of the DAT file from the Affymetrix GeneChip. A single representative intensity value is stored per cell (feature) of the image.

List of presentations and publications associated with this work

Symes A, Eilertsen M, Brouilette KT, Feneley MR, Nariculam J, Patel HRH, Masters JRW, Ahmed A. Gene expression profile of normal prostate epithelium and the discovery of novel prostate adenocarcinoma specific proteins. Submitted BMC Genomics 2010

Wang Q, Symes AJ, Kane CA *et al*. A novel role for Wnt/Ca2+ signaling in actin cytoskeleton remodeling and cell motility in prostate cancer. PLoS One. 2010;5:e10456.

Symes A, Eilertsen M, Brouilette KT, Feneley MR, Nariculam J, Patel HRH, Masters JRW, Ahmed A. Identification of novel epithelial markers of prostate cancer. Pearce Gould Visiting Professor Lectures. UCL London 2010

Symes A, Feneley M, Masters J, Ahmed A. Gene expression profile of normal human prostate epithelium. BAUS Section of Academic Urology, Royal College of Surgeons London 2008

Symes A, Ahmed A, Masters J. Zonal variation in Gene Expression in the Human Prostate. BAUS Annual meeting 2006

References

Abruzzo LV, Lee KY, Fuller A *et al*. Validation of oligonucleotide microarray data using microfluidic low-density arrays: a new statistical method to normalize real-time RT-PCR data. Biotechniques. 2005;38:785-792.

Affymetrix I. Statistical Algorithms Description Document. 2002

Ahram M, Best CJ, Flaig MJ *et al*. Proteomic analysis of human prostate cancer. Mol Carcinog. 2002;33:9-15.

Alaiya AA, Oppermann M, Langridge J *et al*. Identification of proteins in human prostate tumor material by two-dimensional gel electrophoresis and mass spectrometry. Cell Mol Life Sci. 2001;58:307-311.

Alberts B. Molecular biology of the cell. New York: Garland Science; 2008:1 v. (various pagings).

Alevizos I, Mahadevappa M, Zhang X *et al*. Oral cancer in vivo gene expression profiling assisted by laser capture microdissection and microarray analysis. Oncogene. 2001;20:6196-6204.

Amin M, Boccon-Gibod L, Egevad L *et al*. Prognostic and predictive factors and reporting of prostate carcinoma in prostate needle biopsy specimens. Scandinavian Journal of Urology and Nephrology. 2005;39:20-33.

Ananthanarayanan V, Deaton RJ, Yang XJ, Pins MR, Gann PH. Alpha-methylacyl-CoA racemase (AMACR) expression in normal prostatic glands and high-grade prostatic intraepithelial neoplasia (HGPIN): association with diagnosis of prostate cancer. Prostate. 2005;63:341-346.

Ananthanarayanan V, Deaton RJ, Yang XJ, Pins MR, Gann PH. Alteration of proliferation and apoptotic markers in normal and premalignant tissue associated with prostate cancer. BMC Cancer. 2006;6:73.

Auer H, Newsom DL, Kornacker K. Expression Profiling Using Affymetrix GeneChip Microarrays. Methods Mol Biol. 2009;509:35-46.

Bair WBr, Cabello CM, Uchida K, Bause AS, Wondrak GT. GLO1 overexpression in human malignant melanoma. Melanoma Res. 2010;20:85-96.

Bakay M, Chen YW, Borup R, Zhao P, Nagaraju K, Hoffman EP. Sources of variability and effect of experimental approach on expression profiling data interpretation. BMC Bioinformatics. 2002;3:4.

Bartsch G, Keen F, Daxenbichler G *et al*. Correlation of biochemical (receptors, endogenous tissue hormones) and quantitative morphologic (stereologic) findings in normal and hyperplastic human prostates. J Urol. 1987;137:559-564.

Bartsch G, Muller HR, Oberholzer M, Rohr HP. Light microscopic stereological analysis of the normal human prostate and of benign prostatic hyperplasia. J Urol. 1979;122:487-491.

Bassett DEJ, Eisen MB, Boguski MS. Gene expression informatics--it's all in your mine. Nat Genet. 1999;21:51-55.

Belbin TJ, Singh B, Barber I *et al*. Molecular classification of head and neck squamous cell carcinoma using cDNA microarrays. Cancer Res. 2002;62:1184-1190.

,Benjamini Hochberg. Controlling the False Discovery Rate: A Practical Approach. JRSSB. 1995;57:289-300.

Berquin IM, Min Y, Wu R, Wu H, Chen YQ. Expression signature of the mouse prostate. J Biol Chem. 2005;280:36442-36451.

Berry SJ, Coffey DS, Walsh PC, Ewing LL. The development of human benign prostatic hyperplasia with age. J Urol. 1984;132:474-479.

Bhatia-Gaur R, Donjacour AA, Sciavolino PJ *et al*. Roles for Nkx3.1 in prostate development and cancer. Genes Dev. 1999;13:966-977.

Bhattacharjee A, Richards WG, Staunton J *et al*. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc Natl Acad Sci U S A. 2001;98:13790-13795.

Bhatti RA, Gadarowski JJ, Ray PS. Metastatic behavior of prostatic tumor as influenced by the hematopoietic and hematogenous factors. Tumour Biol. 1997;18:1-5.

Bittner M, Meltzer P, Chen Y *et al*. Molecular classification of cutaneous malignant melanoma by gene expression profiling. Nature. 2000;406:536-540.

Blacklock NJ. The anatomy of the prostate: relationship with prostatic infection. Infection. 1991;19 Suppl 3:S111-4.

Blobe GC, Schiemann WP, Lodish HF. Role of transforming growth factor beta in human disease. N Engl J Med. 2000;342:1350-1358.

Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics. 2003;19:185-193.

Bookstein R, Shew JY, Chen PL, Scully P, Lee WH. Suppression of tumorigenicity of human prostate carcinoma cells by replacing a mutated RB gene. Science. 1990;247:712-715.

Bostwick DG. Prostate-specific antigen. Current role in diagnostic pathology of prostate cancer. Am J Clin Pathol. 1994;102:S31-7.

Bova GS, Carter BS, Bussemakers MJ *et al*. Homozygous deletion and frequent allelic loss of chromosome 8p22 loci in human prostate cancer. Cancer Res. 1993;53:3869-3873.

Bradford TJ, Tomlins SA, Wang X, Chinnaiyan AM. Molecular markers of prostate cancer. Urol Oncol. 2006;24:538-551.

Brossner C, Winterholer A, Roehlich M *et al*. Distribution of prostate carcinoma foci within the peripheral zone: analysis of 8,062 prostate biopsy cores. World J Urol. 2003;21:163-166.

Brown PO, Botstein D. Exploring the new world of the genome with DNA microarrays. Nat Genet. 1999;21:33-37.

Buhusi M, Midkiff BR, Gates AM, Richter M, Schachner M, Maness PF. Close homolog of L1 is an enhancer of integrin-mediated cell migration. J Biol Chem. 2003;278:25024-25031.

Burger M, Denzinger S, Hartmann A, Wieland WF, Stoehr R, Obermann EC. Mcm2 predicts recurrence hazard in stage Ta/T1 bladder cancer more accurately than CK20, Ki67 and histological grade. Br J Cancer. 2007;96:1711-1715.

Burris TP. Nuclear hormone receptors for heme: REV-ERBalpha and REV-ERBbeta are ligand-regulated components of the mammalian clock. Mol Endocrinol. 2008;22:1509-1520.

Canales RD, Luo Y, Willey JC *et al*. Evaluation of DNA microarray results with quantitative gene expression platforms. Nat Biotechnol. 2006;24:1115-1122.

Carter BS, Bova GS, Beaty TH *et al*. Hereditary prostate cancer: epidemiologic and clinical features. J Urol. 1993;150:797-802.

Casey T, Patel O, Dykema K, Dover H, Furge K, Plaut K. Molecular signatures reveal circadian clocks may orchestrate the homeorhetic response to lactation. PLoS One. 2009;4:e7395.

Casneuf T, Van de Peer Y, Huber W. In situ analysis of cross-hybridisation on microarrays and the inference of expression correlation. BMC Bioinformatics. 2007;8:461.

Cavalieri EL, Devanesan P, Bosland MC, Badawi AF, Rogan EG. Catechol estrogen metabolites and conjugates in different regions of the prostate of Noble rats treated with 4-hydroxyestradiol: implications for estrogen-induced initiation of prostate cancer. Carcinogenesis. 2002;23:329-333.

Chandran UR, Dhir R, Ma C, Michalopoulos G, Becich M, Gilbertson J. Differences in gene expression in prostate cancer, normal appearing prostate tissue adjacent to cancer and prostate tissue from cancer free organ donors. BMC Cancer. 2005;5:45.

Chang BL, Zheng SL, Isaacs SD *et al*. Polymorphisms in the CYP1B1 gene are associated with increased risk of prostate cancer. Br J Cancer. 2003;89:1524-1529.

Cheetham JE, Coleman PD, Chow N. Isolation of single immunohistochemically identified whole neuronal cell bodies from post-mortem human brain for simultaneous analysis of multiple gene expression. J Neurosci Methods. 1997;77:43-48.

Chen Z, Fan Z, McNeal JE *et al*. Hepsin and maspin are inversely expressed in laser capture microdissectioned prostate cancer. J Urol. 2003;169:1316-1319.

Chu TY, Shen CY, Lee HS, Liu HS. Monoclonality and surface lesion-specific microsatellite alterations in premalignant and malignant neoplasia of uterine cervix: a local field effect of genomic instability and clonal evolution. Genes Chromosomes Cancer. 1999;24:127-134.

Churchill GA. Fundamentals of experimental design for cDNA microarrays. Nat Genet. 2002;32 Suppl:490-495.

Chute CG, Panser LA, Girman CJ *et al*. The prevalence of prostatism: a population-based survey of urinary symptoms. J Urol. 1993;150:85-89.

Chuu CP, Hiipakka RA, Kokontis JM, Fukuchi J, Chen RY, Liao S. Inhibition of tumor growth and progression of LNCaP prostate cancer cells in athymic mice by androgen and liver X receptor agonist. Cancer Res. 2006;66:6482-6486.

Cimino D, Fuso L, Sfiligoi C *et al*. Identification of new genes associated with breast cancer progression by gene expression analysis of predefined sets of neoplastic tissues. Int J Cancer. 2008;123:1327-1338.

Colombel M, Vacherot F, Diez SG, Fontaine E, Buttyan R, Chopin D. Zonal variation of apoptosis and proliferation in the normal prostate and in benign prostatic hyperplasia. Br J Urol. 1998;82:380-385.

Colombo P, Patriarca C, Alfano RM *et al*. Molecular disorders in transitional vs. peripheral zone prostate adenocarcinoma. Int J Cancer. 2001;94:383-389.

Crick F. Central dogma of molecular biology. Nature. 1970;227:561-563.

Crnogorac-Jurcevic T, Efthimiou E, Nielsen T *et al*. Expression profiling of microdissected pancreatic adenocarcinomas. Oncogene. 2002;21:4587-4594.

Dakhova O, Ozen M, Creighton CJ *et al*. Global gene expression analysis of reactive stroma in prostate cancer. Clin Cancer Res. 2009;15:3979-3989.

Das D, Dasgupta P, Chandra A. alpha-acylmethyl co-enzyme A racemase: a tumour marker for the 21st century? BJU Int. 2005;96:3-4.

Dash A, Maine IP, Varambally S, Shen R, Chinnaiyan AM, Rubin MA. Changes in differential gene expression because of warm ischemia time of radical prostatectomy specimens. Am J Pathol. 2002;161:1743-1748.

Dennis GJ, Sherman BT, Hosack DA *et al*. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. 2003;4:P3.

Dhanasekaran SM, Barrette TR, Ghosh D *et al*. Delineation of prognostic biomarkers in prostate cancer. Nature. 2001;412:822-826.

Diboun I, Wernisch L, Orengo CA, Koltzenburg M. Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma. BMC Genomics. 2006;7:252.

Ding Y, Xu L, Chen S *et al*. Characterization of a method for profiling gene expression in cells recovered from intact human prostate tissue using RNA linear amplification. Prostate Cancer Prostatic Dis. 2006;9:379-391.

Dobson JF. Herophilus of Alexandria. Proc R Soc Med. 1925;18:19-32.

Dong JT. Chromosomal deletions and tumor suppressor genes in prostate cancer. Cancer Metastasis Rev. 2001;20:173-193.

Dubbink HJ, de Waal L, van Haperen R, Verkaik NS, Trapman J, Romijn JC. The human prostate-specific transglutaminase gene (TGM4): genomic organization, tissue-specific expression, and promoter characterization. Genomics. 1998;51:434-444.

Dyrskjot L, Thykjaer T, Kruhoffer M *et al*. Identifying distinct classes of bladder carcinoma using microarrays. Nat Genet. 2003;33:90-96.

Eberwine J, Yeh H, Miyashiro K *et al*. Analysis of gene expression in single live neurons. Proc Natl Acad Sci U S A. 1992;89:3010-3014.

Egevad L, Montorsi F, Montironi R. Donald F. Gleason, 1920-2008. Eur Urol. 2009

Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A. 1998;95:14863-14868.

El-Serag HB, Nurgalieva ZZ, Mistretta TA *et al*. Gene expression in Barrett's esophagus: laser capture versus whole tissue. Scand J Gastroenterol. 2009;44:787-795.

Emmert-Buck MR, Bonner RF, Smith PD *et al*. Laser capture microdissection. Science. 1996;274:998-1001.

Emmert-Buck MR, Roth MJ, Zhuang Z *et al*. Increased gelatinase A (MMP-2) and cathepsin B activity in invasive tumor regions of human colon cancer samples. Am J Pathol. 1994;145:1285-1290.

Erbersdobler A, Fritz H, Schnoger S *et al*. Tumour grade, proliferation, apoptosis, microvessel density, p53, and bcl-2 in prostate cancers: differences between tumours located in the transition zone and in the peripheral zone. Eur Urol. 2002a;41:40-46.

Erbersdobler A, Huhle S, Palisaar J *et al*. Pathological and clinical characteristics of large prostate cancers predominantly located in the transition zone. Prostate Cancer Prostatic Dis. 2002b;5:279-284.

Ernst T, Hergenhahn M, Kenzelmann M *et al*. Decrease and gain of gene expression are equally discriminatory markers for prostate carcinoma: a gene expression analysis on total and microdissected prostate tissue. Am J Pathol. 2002;160:2169-2180.

Febbo PG, Lowenberg M, Thorner AR, Brown M, Loda M, Golub TR. Androgen mediated regulation and functional implications of fkbp51 expression in prostate cancer. J Urol. 2005;173:1772-1777.

Febbo PG, Thorner A, Rubin MA *et al*. Application of oligonucleotide microarrays to assess the biological effects of neoadjuvant imatinib mesylate treatment for localized prostate cancer. Clin Cancer Res. 2006;12:152-158.

Feldman AL, Costouros NG, Wang E *et al*. Advantages of mRNA amplification for microarray analysis. Biotechniques. 2002;33:906-12, 914.

Fend F, Raffeld M. Laser capture microdissection in pathology. J Clin Pathol. 2000;53:666-672.

Florell SR, Coffin CM, Holden JA *et al*. Preservation of RNA for functional genomic studies: a multidisciplinary tumor bank protocol. Mod Pathol. 2001;14:116-128.

Florl AR, Steinhoff C, Muller M *et al*. Coordinate hypermethylation at specific genes in prostate carcinoma precedes LINE-1 hypomethylation. Br J Cancer. 2004;91:985-994.

Forrester JC, Ury HK. The Signed-Rank (Wilcoxon) test in the rapid analysis of biological data. Lancet. 1969;1:239-241.

Franklin WA, Gazdar AF, Haney J *et al*. Widely dispersed p53 mutation in respiratory epithelium. A novel mechanism for field carcinogenesis. J Clin Invest. 1997;100:2133-2137.

Franks LM. Benign nodular hyperplasia of the prostate; a review. Ann R Coll Surg Engl. 1953;14:92-106.

Friedman N. Inferring cellular networks using probabilistic graphical models. Science. 2004;303:799-805.

Fujita A, Gomes LR, Sato JR *et al*. Multivariate gene expression analysis reveals functional connectivity changes between normal/tumoral prostates. BMC Syst Biol. 2008;2:106.

Fujiuchi Y, Nagakawa O, Murakami K, Fuse H, Saiki I. Effect of hepatocyte growth factor on invasion of prostate cancer cell lines. Oncol Rep. 2003;10:1001-1006.

Fukuchi J, Hiipakka RA, Kokontis JM *et al*. Androgenic suppression of ATP-binding cassette transporter A1 expression in LNCaP human prostate cancer cells. Cancer Res. 2004;64:7682-7685.

Gerard MA, Krol A, Carbon P. Transcription factor hStaf/ZNF143 is required for expression of the human TFAM gene. Gene. 2007;401:145-153.

Giaginis C, Georgiadou M, Dimakopoulou K *et al*. Clinical significance of MCM-2 and MCM-5 expression in colon cancer: association with clinicopathological parameters and tumor proliferative capacity. Dig Dis Sci. 2009;54:282-291.

Giordano TJ, Shedden KA, Schwartz DR *et al*. Organ-specific molecular classification of primary lung, colon, and ovarian adenocarcinomas using gene expression profiles. Am J Pathol. 2001;159:1231-1238.

Gleason DF, Mellinger GT. the Veterans Administration Cooperative Urological Research Group (1974) Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. J Urol. 1974;111:58–64.

Glinsky GV, Glinskii AB, Stephenson AJ, Hoffman RM, Gerald WL. Gene expression profiling predicts clinical outcome of prostate cancer. J Clin Invest. 2004;113:913-923.

Golub TR, Slonim DK, Tamayo P *et al*. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science. 1999;286:531-537.

Grayhack JT, Lee C, Oliver L, Schaeffer AJ, Wendel EF. Biochemical profiles of prostatic fluid from normal and diseased prostate glands. Prostate. 1980;1:227-237.

Greco D, Somervuo P, Di Lieto A *et al*. Physiology, pathology and relatedness of human tissues from gene expression meta-analysis. PLoS One. 2008;3:e1880.

Gronberg H, Isaacs SD, Smith JR *et al*. Characteristics of prostate cancer in families potentially linked to the hereditary prostate cancer 1 (HPC1) locus. JAMA. 1997;278:1251-1255.

Grotzer MA, Patti R, Geoerger B, Eggert A, Chou TT, Phillips PC. Biological stability of RNA isolated from RNAlater-treated brain tumor and neuroblastoma xenografts. Med Pediatr Oncol. 2000;34:438-442.

Grubb RL, Calvert VS, Wulkuhle JD *et al*. Signal pathway profiling of prostate cancer using reverse phase protein arrays. Proteomics. 2003;3:2142-2146.

Guo D, Catchpoole DR. Isolation of intact RNA following cryosection of archived frozen tissue. Biotechniques. 2003;34:48-50.

Hanna-Morris A, Badvie S, Cohen P, McCullough T, Andreyev HJ, Allen-Mersh TG. Minichromosome maintenance protein 2 (MCM2) is a stronger discriminator of increased proliferation in mucosa adjacent to colorectal cancer than Ki-67. J Clin Pathol. 2009;62:325-330.

Harr B, Schlotterer C. Comparison of algorithms for the analysis of Affymetrix microarray data as evaluated by co-expression of genes in known operons. Nucleic Acids Res. 2006;34:e8.

Harrell JC, Dye WW, Harvell DM, Sartorius CA, Horwitz KB. Contaminating cells alter gene signatures in whole organ versus laser capture microdissected tumors: a comparison of experimental breast cancers and their lymph node metastases. Clin Exp Metastasis. 2008;25:81-88.

Hartigan JA, Wong MA. A k-means Clustering Algorithm. Applied Statistics. 1979;28:100-108.

Hayward SW, Baskin LS, Haughney PC *et al*. Epithelial development in the rat ventral prostate, anterior prostate and seminal vesicle. Acta Anat (Basel). 1996;155:81-93.

He WW, Sciavolino PJ, Wing J *et al*. A novel human prostate-specific, androgen-regulated homeobox gene (NKX3.1) that maps to 8p21, a region frequently deleted in prostate cancer. Genomics. 1997;43:69-77.

Hedenfalk IA. Gene expression profiling of hereditary and sporadic ovarian cancers reveals unique BRCA1 and BRCA2 signatures. J Natl Cancer Inst. 2002;94:960-961.

Herman JG, Meadows GG. Transferrin reverses the anti-invasive activity of human prostate cancer cells that overexpress sema3E. Int J Oncol. 2007;31:1267-1272.

Holter NS, Mitra M, Maritan A, Cieplak M, Banavar JR, Fedoroff NV. Fundamental patterns underlying gene expression profiles: simplicity from complexity. Proc Natl Acad Sci U S A. 2000;97:8409-8414.

Horoszewicz JS, Leong SS, Kawinski E *et al*. LNCaP model of human prostatic carcinoma. Cancer Res. 1983;43:1809-1818.

Hosack DA, Dennis GJ, Sherman BT, Lane HC, Lempicki RA. Identifying biological themes within lists of genes with EASE. Genome Biol. 2003;4:R70.

Howe HL, Wingo PA, Thun MJ *et al*. Annual report to the nation on the status of cancer (1973 through 1998), featuring cancers with recent increasing trends. J Natl Cancer Inst. 2001;93:824-842.

Hricak H, Dooms GC, McNeal JE *et al*. MR imaging of the prostate gland: normal anatomy. AJR Am J Roentgenol. 1987;148:51-58.

Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4:44-57.

Huang da W, Sherman BT, Tan Q *et al*. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol. 2007;8:R183.

Huang J, Qi R, Quackenbush J, Dauway E, Lazaridis E, Yeatman T. Effects of ischemia on gene expression. J Surg Res. 2001;99:222-227.

Hubank M. Gene expression profiling and its application in studies of haematological malignancy. Br J Haematol. 2004;124:577-594.

Huggins C. Endocrine-induced regression of cancers. Science. 1967;156:1050-1054.

Huggins C, Webster WO. Duality of human prostate in response to oestrogen. Journal of Urology. 1948;59:258.

Hughes C, Murphy A, Martin C, Sheils O, O'Leary J. Molecular pathology of prostate cancer. J Clin Pathol. 2005;58:673-684.

Hunter J. A treatise on the venereal disease. London: No. 13, Castle-Street, Leicester-Square; 1786:398 p.: plates.

Hutch JA. Anatomy and physiology of the Bladder, Trigone and Urethra. Butterworth; 1972

Imbeaud S, Graudens E, Boulanger V *et al*. Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. Nucleic Acids Res. 2005;33:e56.

Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res. 2003a;31:e15.

Irizarry RA, Hobbs B, Collin F *et al*. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003b;4:249-264.

Ishiguchi H, Izumi H, Torigoe T *et al*. ZNF143 activates gene expression in response to DNA damage and binds to cisplatin-modified DNA. Int J Cancer. 2004;111:900-909.

Jemal A, Siegel R, Xu J, Ward E. Cancer Statistics, 2010. CA Cancer J Clin. 2010

John K, Ragavan N, Pratt MM *et al*. Quantification of phase I/II metabolizing enzyme gene expression and polycyclic aromatic hydrocarbon-DNA adduct levels in human prostate. Prostate. 2009;69:505-519.

Jost-Albrecht K, Hofstetter W. Gene expression by human monocytes from peripheral blood in response to exposure to metals. J Biomed Mater Res B Appl Biomater. 2006;76:449-455.

Kaighn ME, Kirk D, Szalay M, Lechner JF. Growth control of prostatic carcinoma cells in serum-free media: interrelationship of hormone response, cell density, and nutrient media. Proc Natl Acad Sci U S A. 1981;78:5673-5676.

Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27-30.

Kanehisa M, Goto S, Hattori M *et al*. From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res. 2006;34:D354-7.

Khan J, Wei JS, Ringner M *et al*. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nat Med. 2001;7:673-679.

Khanna S, Cheng G, Gong B, Mustari MJ, Porter JD. Genome-wide transcriptional profiles are consistent with functional specialization of the extraocular muscle layers. Invest Ophthalmol Vis Sci. 2004;45:3055-3066.

Khatri P, Bhavsar P, Bawa G, Draghici S. Onto-Tools: an ensemble of web-accessible, ontology-based tools for the functional design and interpretation of high-throughput gene expression experiments. Nucleic Acids Res. 2004;32:W449-56.

Khatri P, Draghici S. Ontological analysis of gene expression data: current tools, limitations, and open problems. Bioinformatics. 2005;21:3587-3595.

Kim JW, Cheng Y, Liu W *et al*. Genetic and epigenetic inactivation of LPL gene in human prostate cancer. Int J Cancer. 2009;124:734-738.

King C, Guo N, Frampton GM, Gerry NP, Lenburg ME, Rosenberg CL. Reliability and reproducibility of gene expression measurements using amplified RNA from laser-microdissected primary breast tissue with oligonucleotide arrays. J Mol Diagn. 2005;7:57-64.

Kirchhofer D, Peek M, Lipari MT, Billeci K, Fan B, Moran P. Hepsin activates pro-hepatocyte growth factor and is inhibited by hepatocyte growth factor activator inhibitor-1B (HAI-1B) and HAI-2. FEBS Lett. 2005;579:1945-1950.

Kitahara O, Furukawa Y, Tanaka T *et al*. Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia. Cancer Res. 2001;61:3544-3549.

Kobayashi S, Demura T, Nonomura K, Koyanagi T. Autoradiographic localization of alpha 1-adrenoceptors in human prostate: special reference to zonal difference. J Urol. 1991;146:887-890.

Kohonen T. Analysis of a Simple Self-Organizing Process. Biological Cybernetics. 1982;43:59-69.

Koivisto P, Kononen J, Palmberg C *et al*. Androgen receptor gene amplification: a possible molecular mechanism for androgen deprivation therapy failure in prostate cancer. Cancer Res. 1997;57:314-319.

Koshida S, Kobayashi D, Moriai R, Tsuji N, Watanabe N. Specific overexpression of OLFM4(GW112/HGC-1) mRNA in colon, breast and lung cancer tissues detected using quantitative analysis. Cancer Sci. 2007;98:315-320.

Kourtidis A, Jain R, Carkner RD, Eifert C, Brosnan MJ, Conklin DS. An RNA interference screen identifies metabolic regulators NR1D1 and PBP as novel survival factors for breast cancer cells with the ERBB2 signature. Cancer Res. 2010;70:1783-1792.

Krill D, DeFlavia P, Dhir R *et al*. Expression patterns of vitamin D receptor in human prostate. J Cell Biochem. 2001;82:566-572.

Kube DM, Savci-Heijink CD, Lamblin AF *et al*. Optimization of laser capture microdissection and RNA amplification for gene expression profiling of prostate cancer. BMC Mol Biol. 2007;8:25.

Kuhn K, Baker SC, Chudin E *et al*. A novel, high-performance random array platform for quantitative gene expression profiling. Genome Res. 2004;14:2347-2356.

Kumar-Sinha C, Chinnaiyan AM. Molecular markers to identify patients at risk for recurrence after primary treatment for prostate cancer. Urology. 2003;62 Suppl 1:19-35.

Lapointe J, Li C, Higgins JP *et al*. Gene expression profiling identifies clinically relevant subtypes of prostate cancer. Proc Natl Acad Sci U S A. 2004;101:811-816.

Laukkanen JA, Laaksonen DE, Niskanen L, Pukkala E, Hakkarainen A, Salonen JT. Metabolic syndrome and the risk of prostate cancer in Finnish men: a population-based study. Cancer Epidemiol Biomarkers Prev. 2004;13:1646-1650.

Lazar MA, Hodin RA, Cardona G, Chin WW. Gene expression from the c-erbA alpha/Rev-ErbA alpha genomic locus. Potential regulation of alternative splicing by opposite strand transcription. J Biol Chem. 1990;265:12859-12863.

Le Duc IE. The anatomy of the prostate and pathology of early benign hypertrophy. J Urol. 1939;42:1217-1241.

Lee DK, Chang C. Endocrine mechanisms of disease: Expression and degradation of androgen receptor: mechanism and clinical implication. J Clin Endocrinol Metab. 2003;88:4043-4054.

Lee ML, Kuo FC, Whitmore GA, Sklar J. Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. Proc Natl Acad Sci U S A. 2000;97:9834-9839.

Leethanakul C, Patel V, Gillespie J *et al*. Gene expression profiles in squamous cell carcinomas of the oral cavity: use of laser capture microdissection for the construction and analysis of stage-specific cDNA libraries. Oral Oncol. 2000;36:474-483.

Leung CS, Srigley JR. Distribution of lipochrome pigment in the prostate gland: biological and diagnostic implications. Hum Pathol. 1995;26:1302-1307.

Leung YF, Cavalieri D. Fundamentals of cDNA microarray data analysis. Trends Genet. 2003;19:649-659.

Lexander H, Franzen B, Hirschberg D *et al*. Differential protein expression in anatomical zones of the prostate. Proteomics. 2005;5:2570-2576.

Li C, Wong WH. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc Natl Acad Sci U S A. 2001;98:31-36.

Li L, Roden J, Shapiro BE *et al*. Reproducibility, fidelity, and discriminant validity of mRNA amplification for microarray analysis from primary hematopoietic cells. J Mol Diagn. 2005;7:48-56.

Lim WK, Wang K, Lefebvre C, Califano A. Comparative analysis of microarray normalization procedures: effects on reverse engineering gene networks. Bioinformatics. 2007;23:i282-8.

Lin X, Tascilar M, Lee WH *et al*. GSTP1 CpG island hypermethylation is responsible for the absence of GSTP1 expression in human prostate cancer cells. Am J Pathol. 2001;159:1815-1826.

Liu JJ, Cutler G, Li W *et al*. Multiclass cancer classification and biomarker discovery using GA-based algorithms. Bioinformatics. 2005;21:2691-2697.

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25:402-408.

Lockhart DJ, Dong H, Byrne MC *et al*. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol. 1996;14:1675-1680.

Lowsley OS. The development of the human prostate gland with reference to the development of other structures at the neck of the urinary bladder. Journal of Anatomy. 1912;13:299.

Lu N, Hu N, Li WJ *et al*. Microsatellite alterations in esophageal dysplasia and squamous cell carcinoma from laser capture microdissected endoscopic biopsies. Cancer Lett. 2003;189:137-145.

Luciani MF, Denizot F, Savary S, Mattei MG, Chimini G. Cloning of two novel ABC transporters mapping on human chromosome 9. Genomics. 1994;21:150-159.

Luo J, Duggan DJ, Chen Y *et al*. Human prostate cancer and benign prostatic hyperplasia: molecular dissection by gene expression profiling. Cancer Res. 2001;61:4683-4688.

Luo J, Zha S, Gage WR *et al*. Alpha-methylacyl-CoA racemase: a new molecular marker for prostate cancer. Cancer Res. 2002;62:2220-2226.

Luo L, Salunga RC, Guo H *et al*. Gene expression profiles of laser-captured adjacent neuronal subtypes. Nat Med. 1999;5:117-122.

Luzzi V, Mahadevappa M, Raja R, Warrington JA, Watson MA. Accurate and reproducible gene expression profiles from laser capture microdissection, transcript amplification, and high density oligonucleotide microarray analysis. J Mol Diagn. 2003;5:9-14.

Ma XJ, Salunga R, Tuggle JT *et al*. Gene expression profiles of human breast cancer progression. Proc Natl Acad Sci U S A. 2003;100:5974-5979.

Mackey TJ, Borkowski A, Amin P, Jacobs SC, Kyprianou N. bcl-2/bax ratio as a predictive marker for therapeutic response to radiotherapy in patients with prostate cancer. Urology. 1998;52:1085-1090.

Magee JA, Araki T, Patil S *et al*. Expression profiling reveals hepsin overexpression in prostate cancer. Cancer Res. 2001;61:5692-5696.

Marionneau C, Couette B, Liu J *et al*. Specific pattern of ionic channel gene expression associated with pacemaker activity in the mouse heart. J Physiol. 2005;562:223-234.

Masuda N, Ohnishi T, Kawamoto S, Monden M, Okubo K. Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples. Nucleic Acids Res. 1999;27:4436-4443.

McFadyen MC, McLeod HL, Jackson FC, Melvin WT, Doehmer J, Murray GI. Cytochrome P450 CYP1B1 protein expression: a novel mechanism of anticancer drug resistance. Biochem Pharmacol. 2001;62:207-212.

McNeal JE. Regional morphology and pathology of the prostate. Am J Clin Pathol. 1968;49:347-357.

McNeal JE. The zonal anatomy of the prostate. Prostate. 1981;2:35-49.

McNeal JE. Normal histology of the prostate. Am J Surg Pathol. 1988;12:619-633.

McNeal JE, Leav I, Alroy J, Skutelsky E. Differential lectin staining of central and peripheral zones of the prostate and alterations in dysplasia. Am J Clin Pathol. 1988a;89:41-48.

McNeal JE, Redwine EA, Freiha FS, Stamey TA. Zonal distribution of prostatic adenocarcinoma. Correlation with histologic pattern and direction of spread. Am J Surg Pathol. 1988b;12:897-906.

Mead JR, Irvine SA, Ramji DP. Lipoprotein lipase: structure, function, regulation, and role in disease. J Mol Med. 2002;80:753-769.

Meehan KL, Holland JW, Dawkins HJ. Proteomic analysis of normal and malignant prostate tissue to identify novel proteins lost in cancer. Prostate. 2002;50:54-63.

Mehrotra J, Varde S, Wang H *et al*. Quantitative, spatial resolution of the epigenetic field effect in prostate cancer. Prostate. 2008;68:152-160.

Meng MV, Grossfeld GD, Williams GH *et al*. Minichromosome maintenance protein 2 expression in prostate: characterization and association with outcome after therapy for cancer. Clin Cancer Res. 2001;7:2712-2718.

Micke P, Ohshima M, Tahmasebpoor S *et al*. Biobanking of fresh frozen tissue: RNA is stable in nonfixed surgical specimens. Lab Invest. 2006;86:202-211.

Miller GJ, Miller HL, van Bokhoven A *et al*. Aberrant HOXC expression accompanies the malignant phenotype in human prostate. Cancer Res. 2003;63:5879-5888.

Mirowitz SA, Hammerman AM. CT depiction of prostatic zonal anatomy. J Comput Assist Tomogr. 1992;16:439-441.

Mobley JA, Leav I, Zielie P *et al*. Branched fatty acids in dairy and beef products markedly enhance alpha-methylacyl-CoA racemase expression in prostate cancer cells in vitro. Cancer Epidemiol Biomarkers Prev. 2003;12:775-783.

Mohr S, Bottin MC, Lannes B *et al*. Microdissection, mRNA amplification and microarray: a study of pleural mesothelial and malignant mesothelioma cells. Biochimie. 2004;86:13-19.

Murray GI, Melvin WT, Greenlee WF, Burke MD. Regulation, function, and tissue-specific expression of cytochrome P450 CYP1B1. Annu Rev Pharmacol Toxicol. 2001;41:297-316.

Mutter GL, Zahrieh D, Liu C *et al*. Comparison of frozen and RNALater solid tissue storage methods for use in RNA expression microarrays. BMC Genomics. 2004;5:88.

Myslinski E, Gerard MA, Krol A, Carbon P. Transcription of the human cell cycle regulated BUB1B gene requires hStaf/ZNF143. Nucleic Acids Res. 2007;35:3453-3464.

Myslinski E, Krol A, Carbon P. ZNF76 and ZNF143 are two human homologs of the transcriptional activator Staf. J Biol Chem. 1998;273:21998-22006.

Narita S, Tsuchiya N, Wang L *et al*. Association of lipoprotein lipase gene polymorphism with risk of prostate cancer in a Japanese population. Int J Cancer. 2004;112:872-876.

Noel EE, Ragavan N, Walsh MJ *et al*. Differential gene expression in the peripheral zone compared to the transition zone of the human prostate gland. Prostate Cancer Prostatic Dis. 2008;11:173-180.

Noguchi M, Stamey TA, Neal JE, Yemoto CE. An analysis of 148 consecutive transition zone cancers: clinical and histological characteristics. J Urol. 2000;163:1751-1755.

Nomura M, Shimizu S, Sugiyama T *et al*. 14-3-3 Interacts directly with and negatively regulates pro-apoptotic Bax. J Biol Chem. 2003;278:2058-2065.

Nupponen NN, Kakkola L, Koivisto P, Visakorpi T. Genetic alterations in hormone-refractory recurrent prostate carcinomas. Am J Pathol. 1998;153:141-148.

O'Malley BW, Conneely OM. Orphan receptors: in search of a unifying hypothesis for activation. Mol Endocrinol. 1992;6:1359-1361.

Oliphant A, Barker DL, Stuelpnagel JR, Chee MS. BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. Biotechniques. 2002;Suppl:56-8, 60-1.

Ono K, Tanaka T, Tsunoda T *et al*. Identification by cDNA microarray of genes involved in ovarian carcinogenesis. Cancer Res. 2000;60:5007-5011.

Pan W. Incorporating gene functions as priors in model-based clustering of microarray gene expression data. Bioinformatics. 2006;22:795-801.

Park WD, Stegall MD. A meta-analysis of kidney microarray datasets: investigation of cytokine gene detection and correlation with rt-PCR and detection thresholds. BMC Genomics. 2007;8:88.

Perou CM, Sorlie T, Eisen MB *et al*. Molecular portraits of human breast tumours. Nature. 2000;406:747-752.

Petrovics G, Liu A, Shaheduzzaman S *et al*. Frequent overexpression of ETS-related gene-1 (ERG1) in prostate cancer transcriptome. Oncogene. 2005;24:3847-3852.

Phillips J, Eberwine JH. Antisense RNA Amplification: A Linear Amplification Method for Analyzing the mRNA Population from Single Living Cells. Methods. 1996;10:283-288.

Phillips SM, Barton CM, Lee SJ *et al*. Loss of the retinoblastoma susceptibility gene (RB1) is a frequent and early event in prostatic tumorigenesis. Br J Cancer. 1994;70:1252-1257.

Polz MF, Cavanaugh CM. Bias in template-to-product ratios in multitemplate PCR. Appl Environ Microbiol. 1998;64:3724-3730.

Prakash K, Pirozzi G, Elashoff M *et al*. Symptomatic and asymptomatic benign prostatic hyperplasia: molecular differentiation by using microarrays. Proc Natl Acad Sci U S A. 2002;99:7598-7603.

Preitner N, Damiola F, Lopez-Molina L *et al*. The orphan nuclear receptor REV-ERBalpha controls circadian transcription within the positive limb of the mammalian circadian oscillator. Cell. 2002;110:251-260.

Prevo LJ, Sanchez CA, Galipeau PC, Reid BJ. p53-mutant clones and field effects in Barrett's esophagus. Cancer Res. 1999;59:4784-4787.

Price D. Comparative aspects of development and structure in the prostate. Natl Cancer Inst Monogr. 1963;12:1-27.

Quackenbush J. Computational analysis of microarray data. Nat Rev Genet. 2001;2:418-427.

Radford DM, Fair K, Thompson AM *et al*. Allelic loss on a chromosome 17 in ductal carcinoma in situ of the breast. Cancer Res. 1993;53:2947-2949.

Ragavan N, Hewitt R, Cooper LJ *et al*. CYP1B1 expression in prostate is higher in the peripheral than in the transition zone. Cancer Lett. 2004;215:69-78.

Ramaswamy S, Tamayo P, Rifkin R *et al*. Multiclass cancer diagnosis using tumor gene expression signatures. Proc Natl Acad Sci U S A. 2001;98:15149-15154.

Ramnath N, Hernandez FJ, Tan DF *et al*. MCM2 is an independent predictor of survival in patients with non-small-cell lung cancer. J Clin Oncol. 2001;19:4259-4266.

Rasband W. ImageJ Software. 1997

Reese JH, McNeal JE, Goldenberg SL, Redwine EA, Sellers RG. Distribution of lactoferrin in the normal and inflamed human prostate: an immunohistochemical study. Prostate. 1992;20:73-85.

Reese JH, McNeal JE, Redwine EA, Samloff IM, Stamey TA. Differential distribution of pepsinogen II between the zones of the human prostate and the seminal vesicle. J Urol. 1986;136:1148-1152.

Reese JH, McNeal JE, Redwine EA, Stamey TA, Freiha FS. Tissue type plasminogen activator as a marker for functional zones, within the human prostate gland. Prostate. 1988;12:47-53.

Rennie PS, Nelson CC. Epigenetic mechanisms for progression of prostate cancer. Cancer Metastasis Rev. 1998;17:401-409.

Rhodes DR, Barrette TR, Rubin MA, Ghosh D, Chinnaiyan AM. Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. Cancer Res. 2002;62:4427-4433.

Rhodes DR, Kalyana-Sundaram S, Mahavisno V *et al*. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. Neoplasia. 2007a;9:166-180.

Rhodes DR, Kalyana-Sundaram S, Tomlins SA *et al*. Molecular concepts analysis links tumors, pathways, mechanisms, and drugs. Neoplasia. 2007b;9:443-454.

Rhodes DR, Yu J, Shanker K *et al*. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia. 2004;6:1-6.

Richter E, Srivastava S, Dobi A. Androgen receptor and prostate cancer. Prostate Cancer Prostatic Dis. 2007;10:114-118.

Rohr HP, Bartsch G. Human benign prostatic hyperplasia: a stromal disease? New perspectives by quantitative morphology. Urology. 1980;16:625-633.

Rokman A, Ikonen T, Mononen N *et al*. ELAC2/HPC2 involvement in hereditary and sporadic prostate cancer. Cancer Res. 2001;61:6038-6041.

Romanuik TL, Ueda T, Le N *et al*. Novel biomarkers for prostate cancer including noncoding transcripts. Am J Pathol. 2009;175:2264-2276.

Rossi MC, Zetter BR. Selective stimulation of prostatic carcinoma cell proliferation by transferrin. Proc Natl Acad Sci U S A. 1992;89:6197-6201.

Roy S, McPherson RA, Apolloni A *et al*. 14-3-3 facilitates Ras-dependent Raf-1 activation in vitro and in vivo. Mol Cell Biol. 1998;18:3947-3955.

Rubin MA, Dunn R, Strawderman M, Pienta KJ. Tissue microarray sampling strategy for prostate cancer biomarker analysis. Am J Surg Pathol. 2002;26:312-319.

Santarius T, Bignell GR, Greenman CD *et al*. GLO1-A novel amplified gene in human cancer. Genes Chromosomes Cancer. 2010;49:711-725.

Santel T, Pflug G, Hemdan NY *et al*. Curcumin inhibits glyoxalase 1: a possible link to its anti-inflammatory and anti-tumor activity. PLoS One. 2008;3:e3508.

Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science. 1995;270:467-470.

Schuchhardt J, Beule D, Malik A *et al*. Normalization strategies for cDNA microarrays. Nucleic Acids Res. 2000;28:E47.

Segal E, Shapira M, Regev A *et al*. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nat Genet. 2003;34:166-176.

Seth RM, Burger AM, Kahn HJ. Analysis of the HER2/neu gene amplification in microdissected breast cancer tumour samples. Anticancer Res. 2006;26:927-931.

Sgroi DC, Teng S, Robinson G, LeVangie R, Hudson JRJ, Elkahloun AG. In vivo gene expression profile analysis of human breast cancer progression. Cancer Res. 1999;59:5656-5661.

Shand RL, Gelmann EP. Molecular biology of prostate-cancer pathogenesis. Curr Opin Urol. 2006;16:123-131.

Shen R, Ghosh D, Chinnaiyan A, Meng Z. Eigengene-based linear discriminant model for tumor classification using gene expression microarray data. Bioinformatics. 2006;22:2635-2642.

Shibata D, Hawes D, Li ZH, Hernandez AM, Spruck CH, Nichols PW. Specific genetic analysis of microscopic tissue after selective ultraviolet radiation fractionation and the polymerase chain reaction. Am J Pathol. 1992;141:539-543.

Shyamsundar R, Kim YH, Higgins JP *et al*. A DNA microarray survey of gene expression in normal human tissues. Genome Biol. 2005;6:R22.

Singh D, Febbo PG, Ross K *et al*. Gene expression correlates of clinical prostate cancer behavior. Cancer Cell. 2002;1:203-209.

Slaughter DP, Southwick HW, Smejkal W. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. Cancer. 1953;6:963-968.

Slonim DK. From patterns to pathways: gene expression data analysis comes of age. Nat Genet. 2002;32 Suppl:502-508.

Sobin LH. TNM classification of malignant tumours. Wiley-Blackwell; 2009

Sommer FG, McNeal JE, Carrol CL. MR depiction of zonal anatomy of the prostate at 1.5 T. J Comput Assist Tomogr. 1986;10:983-989.

Sorlie T, Perou CM, Tibshirani R *et al*. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A. 2001;98:10869-10874.

Srikantan V, Valladares M, Rhim JS, Moul JW, Srivastava S. HEPSIN inhibits cell growth/invasion in prostate cancer cells. Cancer Res. 2002;62:6812-6816.

Srinivasan M, Sedmak D, Jewell S. Effect of fixatives and tissue processing on the content and integrity of nucleic acids. Am J Pathol. 2002;161:1961-1971.

Stamey TA, Caldwell MC, Fan Z *et al*. Genetic profiling of Gleason grade 4/5 prostate cancer: which is the best prostatic control tissue? J Urol. 2003;170:2263-2268.

Stamey TA, Warrington JA, Caldwell MC *et al*. Molecular genetic profiling of Gleason grade 4/5 prostate cancers compared to benign prostatic hyperplasia. J Urol. 2001;166:2171-2177.

Stanbrough M, Bubley GJ, Ross K *et al*. Increased expression of genes converting adrenal androgens to testosterone in androgen-independent prostate cancer. Cancer Res. 2006;66:2815-2825.

Stoyanova R, Upson JJ, Patriotis C *et al*. Use of RNA amplification in the optimal characterization of global gene expression using cDNA microarrays. J Cell Physiol. 2004;201:359-365.

Stuart RO, Wachsman W, Berry CC *et al*. In silico dissection of cell-type-associated patterns of gene expression in prostate cancer. Proc Natl Acad Sci U S A. 2004;101:615-620.

Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP. GSEA-P: a desktop application for Gene Set Enrichment Analysis. Bioinformatics. 2007;23:3251-3253.

Subramanian A, Tamayo P, Mootha VK *et al*. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545-15550.

Takahashi T, Habuchi T, Kakehi Y *et al*. Clonal and chronological genetic analysis of multifocal cancers of the bladder and upper urinary tract. Cancer Res. 1998;58:5835-5841.

Takeshima Y, Amatya VJ, Daimaru Y, Nakayori F, Nakano T, Inai K. Heterogeneous genetic alterations in ovarian mucinous tumors: application and usefulness of laser capture microdissection. Hum Pathol. 2001;32:1203-1208.

Tanaka T, Knapp D, Nasmyth K. Loading of an Mcm protein onto DNA replication origins is regulated by Cdc6p and CDKs. Cell. 1997;90:649-660.

Tenedini E, Fagioli ME, Vianelli N *et al*. Gene expression profiling of normal and malignant CD34-derived megakaryocytic cells. Blood. 2004;104:3126-3135.

Group TTABPW. Expression profiling--best practices for data generation and interpretation in clinical trials. Nat Rev Genet. 2004;5:229-237.

Thielen JL, Volzing KG, Collier LS, Green LE, Largaespada DA, Marker PC. Markers of prostate region-specific epithelial identity define anatomical locations in the mouse prostate that are molecularly similar to human prostate cancers. Differentiation. 2007;75:49-61.

Thomas JG, Olson JM, Tapscott SJ, Zhao LP. An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles. Genome Res. 2001;11:1227-1236.

Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A. 2002;99:6567-6572.

Tisell LE, Salander H. The lobes of the human prostate. Scand J Urol Nephrol. 1975;9:185-191.

Tomlins SA, Mehra R, Rhodes DR *et al*. Integrative molecular concept modeling of prostate cancer progression. Nat Genet. 2007;39:41-51.

Tomlins SA, Rhodes DR, Perner S *et al*. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science. 2005;310:644-648.

Tsurusaki T, Aoki D, Kanetake H *et al*. Zone-dependent expression of estrogen receptors alpha and beta in human benign prostatic hyperplasia. J Clin Endocrinol Metab. 2003;88:1333-1340.

Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A. 2001;98:5116-5121.

van der Heul-Nieuwenhuijsen L, Hendriksen PJ, van der Kwast TH, Jenster G. Gene expression profiling of the human prostate zones. BJU Int. 2006;98:886-897.

Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD, Eberwine JH. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. Proc Natl Acad Sci U S A. 1990;87:1663-1667.

van Meer G, Halter D, Sprong H, Somerharju P, Egmond MR. ABC lipid transporters: extruders, flippases, or flopless activators? FEBS Lett. 2006;580:1171-1177.

Varambally S, Yu J, Laxman B *et al*. Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. Cancer Cell. 2005;8:393-406.

Vasioukhin V. Hepsin paradox reveals unexpected complexity of metastatic process. Cell Cycle. 2004;3:1394-1397.

Verhamme KM, Dieleman JP, Bleumink GS *et al*. Incidence and prevalence of lower urinary tract symptoms suggestive of benign prostatic hyperplasia in primary care--the Triumph project. Eur Urol. 2002;42:323-328.

Villers A, Terris MK, McNeal JE, Stamey TA. Ultrasound anatomy of the prostate: the normal gland and anatomical variations. J Urol. 1990;143:732-738.

Vincek V, Nassiri M, Knowles J, Nadji M, Morales AR. Preservation of tissue RNA in normal saline. Lab Invest. 2003;83:137-138.

Wakasugi T, Izumi H, Uchiumi T *et al*. ZNF143 interacts with p73 and is involved in cisplatin resistance through the transcriptional regulation of DNA repair genes. Oncogene. 2007;26:5194-5203.

Wallard MJ, Pennington CJ, Veerakumarasivam A *et al*. Comprehensive profiling and localisation of the matrix metalloproteinases in urothelial carcinoma. Br J Cancer. 2006;94:569-577.

Wang E, Miller LD, Ohnmacht GA, Liu ET, Marincola FM. High-fidelity mRNA amplification for gene profiling. Nat Biotechnol. 2000;18:457-459.

Wang H, Owens JD, Shih JH, Li MC, Bonner RF, Mushinski JF. Histological staining methods preparatory to laser capture microdissection significantly affect the integrity of the cellular RNA. BMC Genomics. 2006;7:97.

Wang Q, Symes AJ, Kane CA *et al*. A novel role for wnt/ca signaling in actin cytoskeleton remodeling and cell motility in prostate cancer. PLoS One. 2010;5:e10456.

Wein AJ, Kavoussi LR, Novick AC, Partin AW, Peters CA. Campbell-Walsh Urology e-dition: Text with Continually Updated Online Reference, 4-Volume Set. Saunders; 2006

Welsh JB, Sapinoso LM, Su AI *et al*. Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. Cancer Res. 2001a;61:5974-5978.

Welsh JB, Zarrinkar PP, Sapinoso LM *et al*. Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. Proc Natl Acad Sci U S A. 2001b;98:1176-1181.

Wen-Xin L, Xi-Shan H. Application of laser capture microdissection and differential display technique for screening of pathogenic genes involved in endometrial carcinoma. Int J Gynecol Cancer. 2007;17:1224-1230.

Williams JA, Martin FL, Muir GH, Hewer A, Grover PL, Phillips DH. Metabolic activation of carcinogens and expression of various cytochromes P450 in human prostate tissue. Carcinogenesis. 2000;21:1683-1689.

Wilson JD. The pathogenesis of benign prostatic hyperplasia. Am J Med. 1980;68:745-756.

Wolffe AP, Matzke MA. Epigenetics: regulation through repression. Science. 1999;286:481-486.

Won J, Kim DY, La M, Kim D, Meadows GG, Joe CO. Cleavage of 14-3-3 protein by caspase-3 facilitates bad interaction with Bcl-x(L) during apoptosis. J Biol Chem. 2003;278:19347-19351.

Wu MS, Lin YS, Chang YT, Shun CT, Lin MT, Lin JT. Gene expression profiling of gastric cancer by microarray combined with laser capture microdissection. World J Gastroenterol. 2005;11:7405-7412.

Wu Z, Irizarry R, R G, Murillo F, F S. A Model Based Background Adjustment for Oligonucleotide Expression Arrays. Technical Report [dissertation]. Baltimore: John Hopkins University; 2003.

Xu L, Tan AC, Naiman DQ, Geman D, Winslow RL. Robust prostate cancer marker genes emerge from direct integration of inter-study microarray data. Bioinformatics. 2005;21:3905-3911.

Yeung KY, Ruzzo WL. Principal component analysis for clustering gene expression data. Bioinformatics. 2001;17:763-774.

Yin L, Wu N, Curtin JC *et al*. Rev-erbalpha, a heme sensor that coordinates metabolic and circadian pathways. Science. 2007;318:1786-1789.

Yu JS, Koujak S, Nagase S *et al*. PCDH8, the human homolog of PAPC, is a candidate tumor suppressor of breast cancer. Oncogene. 2008;27:4657-4665.

Yu YP, Landsittel D, Jing L *et al*. Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy. J Clin Oncol. 2004;22:2790-2799.

Zeeberg BR, Qin H, Narasimhan S *et al*. High-Throughput GoMiner, an 'industrial-strength' integrative gene ontology tool for interpretation of multiple-microarray experiments, with application to studies of Common Variable Immune Deficiency (CVID). BMC Bioinformatics. 2005;6:168.

Zhang H, Yu CY, Singer B, Xiong M. Recursive partitioning for tumor classification with gene expression microarray data. Proc Natl Acad Sci U S A. 2001;98:6730-6735.

Zhang J, Liu WL, Tang DC *et al*. Identification and characterization of a novel member of olfactomedin-related protein family, hGC-1, expressed during myeloid lineage development. Gene. 2002;283:83-93.

Zhang J, Yang J, Roy SK *et al*. The cell death regulator GRIM-19 is an inhibitor of signal transducer and activator of transcription 3. Proc Natl Acad Sci U S A. 2003;100:9342-9347.

Zhang X, Huang Q, Yang Z, Li Y, Li CY. GW112, a novel antiapoptotic protein that promotes tumor growth. Cancer Res. 2004;64:2474-2481.

Zheng SL, Chang BL, Faith DA *et al*. Sequence variants of alpha-methylacyl-CoA racemase are associated with prostate cancer risk. Cancer Res. 2002;62:6485-6488.