Analysis of Messenger RNAs Detectable in Medium Conditioned by Tumour Cells and in Serum from Breast Cancer Patients

A thesis submitted for the degree of Ph.D.

Dublin City University

By

Elaine Kenny B.Sc. (Biotechnology)

The research work described in thesis was performed under the supervision of Prof. Martin Clynes and Dr Lorraine O'Driscoll

National Institute for Cellular Biotechnology

Dublin City University

June 2006

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Ph.D. is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: Elaine Tannyn (Candidate) ID No.: 97470015

Date: 24th August 2006

Acknowledgements

I would like to sincerely thank Professor Martin Clynes and Dr. Lorraine O'Driscoll for taking a chance on me and letting me do some research work before I finished my degree. It helped me to realise that research work is were I want to be (and encouraged me to study harder for my exams so that I could do a PhD here!). Throughout my PhD the constant encouragement (and crossed fingers!) got me through many dark days and confusing results.

Thanks Paddy ("the human encyclopaedia"), for all the help and training when I first started. Joe, I'm convinced the place would come to a standstill without you, thanks for all the "special runs" I know we're a demanding lot! To all who have advised me on techniques/ software and listened patiently to my (sometimes insane) ideas; Jai, Niall, Paudie, Ann Marie and Norma; thanks a mill, I couldn't have done it without you.

A special thanks also to Irene, you're the best coffee/ McDonalds mate a girl could ever ask for. You and Bella made it so easy for me at the end. To my lunch buddies Olga (and Yegor for 8 ½ months!), Jason, Ann Marie and Helena, thanks. To Yvonne, Carol, Mairead, Eadaoin, Sweta, Mick, Finbar, Paula, Will, Laura and everyone else who's friendship made working in the NCTCC/ NICB such a happy experience for me thank-you.

Thanks especially to my family, Mam, Dad, Ed, Paul, Trev, my twinny Darren, sisters(-ish) Hilary, Lorraine and Meabh for all your encouragement and support and yes, I'm finally going to get a real job! And above all, to Paul, without your encouragement I may never have started (or finished!) this work. This achievement is as much yours as it is mine, thanks for putting up with all the crankiness and tears and for always being there for me. Bring on Hawaii!!

ABBREVIATIONS

A	-	Absent
AFP	-	Alpha feto protein
ATCC	-	American Tissue Culture Collection
cDNA	-	Complementary DNA
CDS	-	Coding Sequence
CEA	-	Carcinoembryonic antigen
CK	-	Cytokeratin
CM	-	Conditioned Media
COV	-	Coefficient of Variation
cRNA	-	Complementary RNA
DEPC	-	Diethyl Pyrocarbonate
DMEM	-	Dublecco's Minimum Essential Medium
DMSO	-	Dimethyl Sulfoxide
EDTA	-	Ethylenediaminetetraacetic Acid
EST	-	Expressed sequence tag
FCS	-	Foetal Calf Serum
H	-	NCI H1299
HCC	-	Hepatocellular carcinoma
hCG	-	Human chorionic gonadotropin-β
HNSCC	-	Head and neck squamous cell carcinoma
HT	-	NCI H1299 TAX
IHC	-	Immunohistochemistry
LOH	-	Loss of Heterozygosity
M	-	MDA-F
MA	-	MDA-F-ADR
MMLV-RT	-	Moloney Murine Leukaemia Virus- Reverse Transcriptase
mRNA	-	Messenger RNA
MT	-	MDA-F-TAX
NPC	-	Nasopharyngeal carcinoma
NSCLC	-	Non-small cell lung cancer

P or M	-	Present or Marginal
PBS	-	Phosphate Buffered Saline
Post	-	Post-operation
Pre	-	Pre-operation
R	-	RPMI 2650
rcf	-	relative centrifugal field (also= xg)
RM	-	RPMI 2650 ML
Rpm	-	Revolutions per minute
RT	-	RPMI 2650 TAX
RT-PCR	-	Reverse Transcription Polymerase Chain Reaction
SD	-	Standard Deviation
Tris	-	Tris (hydroxymethyl) aminomethane
TV	-	0.25% Trypsin/ 0.01% EDTA solution in PBS
UTR	-	Untranslated Region

ABSTRACT

Biomarkers have great potential for cancer detection and monitoring, with much of this type of analysis being carried out using tumour biopsies. This approach requires invasive procedures to obtain suitable specimens and only allows analysis of gene expression at one particular time point in the history of a cancer and from one location in the body.

Biomarkers detectable in readily accessible body fluids, such as serum, saliva or urine would allow on-going/ sequential monitoring of the course of disease (progression, response to therapy, *etc.*). A small number of studies have indicated the possibility of amplifying extracellular mRNA from the serum and/or plasma of cancer patients supporting the potential of this route of analysis. Limitations of these studies include the small numbers of serum/plasma specimens analysed; limited numbers of gene transcripts analysed; and discrepancies in protocols used, leading to a possibility that cells circulating in the bloodstream may be included in the RNA isolations; and the fact that a standard technique in general, was not employed.

This thesis aims to address these issues by firstly establishing the possibility of routinely extracting and amplifying extracellular mRNA from the conditioned media (CM) of cultured cells. Having established, in principle, that amplifiable RNA could be isolated, comparison and optimisation of methods of extracting RNA was carried out to identify a reliable and reproducible method. Once this was established, CM samples from different cancer cell lines were investigated for expression of known tumour- related mRNAs by RT-PCR and microarray technology was employed to investigate the global expression of mRNAs with both known and unknown functions.

A slightly modified version of this method was also applied to mRNA extracted from serum of breast cancer patients and normal volunteers, enabling approx. 55,000 transcripts/ variants represented on the whole human genome chips to be analysed to identify mRNAs suitable for further analysis as potential future biomarkers.

1 INTRODUCTION

1.1 Tumour Markers and Biomarkers in Cancer

Biomarkers, defined as biochemical features that may be used to measure the progression of disease or the effects of treatment, have great potential for cancer detection and monitoring. The term tumour marker is applied to all substances produced and released either by tumour cells or by host cells and whose presence may be detected in the serum or other biological fluids, behaving as an indicator of the presence of the tumour, but not all tumour markers are suitable biomarkers.

There are three types of biomarkers, diagnostic, prognostic and predictive. A diagnostic marker is able to accurately diagnose disease i.e. presence of a diagnostic biomarker should identify with a disease-state. A prognostic marker should be able to distinguish (once a disease state has been diagnosed), the likely outcome associated with the presence/ quantity of the marker irrespective of therapy. Candidates for prognostic markers include EGFR and node status. The concentration of or presence/ absence of a biomarker in response to therapy is a predictive biomarker. It should accurately predict a patients response to therapy e.g. HER2/neu.

The ideal tumour biomarker marker should have both a high sensitivity and a high specificity so that cancer patients can be effectively distinguished from healthy subjects or patients with non- cancerous concentrations. Effective biomarkers can serve as molecular signposts of the physiological state of a cell, which is essential for treatment decisions. Examples of prognostic markers include hormone receptors, proliferation markers, proteases, markers of angiogenesis etc.

Predictive markers can help make decisions between different alternative treatments, e.g. breast cancer patients that have estrogen receptor positive tumours are usually treated with anti-estrogen compounds whereas estrogen receptor negative tumour patients are treated with chemotherapy.

Traditionally cancer diagnosis is based on microscopical assessments of morphologic alterations of cell and tissues. In many cases, there is a clear distinction between benign and malignant cells/ tissues.

The TNM staging system circulated by the American Joint Committee on Cancer (AJCC) focuses on the anatomical extent of the disease, which is assessed by combining tumour size/ depth (T), lymph node spread (N), and presence or absence of metastasis (M). Its function is to provide clinicians with the ability for prediction of survival, choice of initial treatment, stratification of patients in clinical trials, accurate communication among healthcare providers resulting in uniform reporting of outcomes (Ludwig and Weinstein 2005).

Far more common though is the situation whereby the diagnosis may be ambiguous and determination of prognosis is difficult (Alaiya *et al* 2005). There is a need to find more ways to classify cancer using tumour markers/ biomarkers because individual molecular markers and patterns of markers are successfully subdividing traditional tumour classes into subsets that behave differently from each other. Also chemotherapeutic and biological agents are more widely used now and more effective than when TNM staging was introduced. And lastly, emerging now is the fact that many target agents are effective only if their respective molecular markers are mutated or expressed at sufficient levels. A good example of this in breast cancer is discovering the oestrogen receptor (ER) and Her2/*neu* status. Independently of TNM staging ER positive status improves prognosis and directs clinicians to use hormone targeting agents like tamoxifen. Originally Her2/*neu* positivity was considered as a negative prognostic indicator independent of TNM stage until the arrival of Herceptin (Ludwig and Weinstei, 2005).

Many advances have been made in the understanding of malignant disease but reliable prognostic and treatment predictive biomarkers have yet to be found for many cancers.

1.2 Serum/ Tissue Tumour Markers

Serum tumour markers can be detected based on the levels of tumour specific proteins/ nucleic acids in serum by identification techniques such as ELISA (protein) and PCR (nucleic acids). Tissue tumour markers are identified by techniques such as immunohistochemistry (IHC) for detection of protein over expression or fluorescent-in-situ-hybridisation (FISH) for detection of gene amplification. Some of the more common serum protein markers and tissue markers used for the identification/ classification of cancer are detailed.

1.2.1 α-Fetoprotein

α-Fetoprotein (AFP), as a serum protein marker for Hepatocellular carcinoma (HCC), was found to have a sensitivity of 39-65%, a specificity of 76-94%, and a positive predictive value of 9-50% when results from several studies from 1995 to 2002 were considered (Daniele *et al* 2004). The higher the cut-off level for AFP detection selected, the higher the specificity and the lower the sensitivity.

A serum concentration of >500μg/L, in an area with high incidence of HCC and in the appropriate clinical setting, is usually considered diagnostic of HCC. HCC patients with high AFP levels at diagnosis tend to have greater tumour size, bi-lobar involvement, massive or diffuse types, and portal vein thrombosis (Tangkijvanich *et al* 2000).

Even so, an increase from 10- 500μg/L is also common in non-malignant chronic liver disease (Poon *et al* 2002) and patients with chronic viral hepatitis can also have elevated levels of AFP with no HCC (Daniele *et al* 2004) so that the specificity of the AFP test for HCC patients is low.

A variant of AFP (AFP-L3), the fucosylated variant, has a high affinity to the sugar chain of *Lens culinaris*, and has been shown to be a more useful indicator of distant metastasis for

HCC than total AFP (Johnson *et al* 2000). An AFP-L3 positive tumour is an advanced tumour regardless of small tumour size and lower serum AFP levels (Qin and Tang 2004). It has a 9-12 month lead-time in early recognition when compared to imaging techniques, showing a sensitivity of 56% and a specificity of over 95% (Li, D., *et al* 2001).

Work by Johnson *et al* (2000) identified a monosialyated form of AFP (msAFP) which had earlier been described as allowing early, even preclinical diagnosis of HCC in high –risk patients (Johnson *et al* 1997). They went on to develop an assay to detect this msAFP and found that it had more diagnostic potential than non-diagnostic AFP i.e. serum AFP levels between 50- 500μg/L (Poon *et al* 2002).

1.2.2 Human chorionic gonadotropin- β (hCG)

There are four members of the glycoprotein hormone family, the placental product hCG, thyroid-stimulating hormone, follicle stimulating hormone and LH, and all excluding hCG, are pituitary products. The hormones share a common glycoprotein hormone α-subunit non-covalently linked to a functionally distinct β-subunit, consequently forming four different intact heterodimeric hormones (Butler and Iles 2003). The β-subunits are homologous and allow each hormone, despite performing discrete functions, to act on the same receptor.

It is the free β-subunit (hCGβ) that is produced by epithelial tumours and is independent of glycoprotein hormone α-subunit gene expression (Iles *et al* 1990). Ectopic expression of hCGβ by bladder carcinoma occurs in ~35% of cases and serves as a model for such ectopic expression by common epithelial cancers. There is also a correlation of hCGβ expression by these tumours with grade, stage and prognosis (Butler and Iles 2003). hCGβ expression has also been significantly associated with failure to respond to therapy (Moutzouris *et al* 1993).

1.2.3 CA 125

CA 125 is produced by a variety of cells, but particularly by ovarian cancer cells. It is an antigenic epitope on a high-molecular weight glycoprotein and was discovered in 1981 by immunising mice with ovarian cancer cells. It is expressed in coelomic epithelium during embryonic development and recognised by a monoclonal antibody developed against a human ovarian carcinoma cell line (Bast *et al* 1981).

In patients with endometrial carcinoma, preoperative serum CA 125 levels correlate with stage, depth of myometrial invasion, histologic grade, cervical invasion, peritoneal cytology, and lymph node status (Gadducci *et al* 2004).

1.2.4 CEA

Carcinoembryonic antigen (CEA) was one of the first tumour markers to be identified and characterised. It is a single- chain glycoprotein of 641 amino acids with a molecular mass of 180- 200kDa containing 45-55% carbohydrate. It consists of a large family of at least 30 closely related cell surface glycoproteins encoded by about ten genes/ pseudogenes located on the long arm of chromosome 19q13.2. The domain structure of CEA proteins are similar to the γ heavy chain of the immunoglobulin IgG indicating that it may be a part of the immunoglobulin super family (Seregni *et al* 2004).

CEA molecules can be divided into three subgroups, the first subgroup containing seven expressed genes, another containing 11 expressed pregnancy –specific glycoproteins (PSG) and a third containing six pseudogenes (Aquino *et al* 2004).

CEA, as a tumour-associated antigen, is expressed in over 95% of colorectal, gastric and pancreatic carcinomas. In addition to this it is also expressed in approximately 70% of non-small cell lung cancer, 50% of breast cancer, in mucinous ovarian carcinoma, adenocarcinoma of the penis and endometrial adenocarcinoma (Aquino *et al* 2004).

Serum CEA is elevated in 30-50% of patients with symptomatic metastatic breast cancer, and a positive correlation has been shown between changes in serum CEA and therapeutic response in patients with metastatic breast cancer (Cheung *et al* 2000).

An early study by Thompson *et al* (1969) described the elevation of CEA in serum of almost all patients with colorectal cancer and rarely in healthy subjects was promising but subsequent studies conflicted with these findings. For example Fletcher (1986) concluded that using a cut-off point of 2.5μg/l, the sensitivity of CEA for early colorectal cancer was only 30-40% resulting in a 250 false positive tests for every patient with cancer. Also possibly more important, 60% of the cancers would remain undetected.

CEA can also be elevated in many benign disorders. As CEA is metabolised by the liver, any benign liver disease that would impair liver function preventing clearance of CEA, would consequently result in an increased serum levels (Duffy 2001). Studies of serum CEA levels in male and female smokers and non- smoker groups has also shown that smoking can almost double the serum levels of CEA (Wilson *et al* 1999).

Cell associated CEA has been used as a tumour marker to trace circulating cancer cells for prediction of recurrence (Lopez-Guerrero *et al* 1999) or detection of micrometastasis in lymph nodes and bone marrow (Liefers *et al* 1998, Weitz *et al* 1999). CEA has been used to target anticancer agents by coupling to tumour-selective anti-CEA monoclonal antibodies (mAbs) (Greiner *et al* 1987) and in immunotherapy, by production of recombinant vaccines able to elicit MHC- restricted immune responses against immunogenic CEA-derived peptides (Foon *et al* 1997).

1.2.5 Cytokeratins

Cytokeratins are one of seven classes of intermediate filaments that, together with microtubules and actin microfilaments, make up the cytoskeleton of most eukaryotic cells.

Different epithelial cells express characteristic, differentiation- dependent combinations of two or more cytokeratins indicating tissue- specific expression (Sergni *et al* 2004). Epithelial cells express between two and ten cytokeratin isotypes. Studies have shown that cytokeratin 19 (CK19) plays an important role in the regulation of cell migration and invasion (Uenishi *et al* 2003a), and can predict early postoperative recurrence of HCC (Uenishi *et al* 2003b).

Cytokeratins also exist in soluble form and can be detected serum using monoclonal antibodies. Serum CK19 fragments can be detected using an enzyme- immunoassay called CYFRA 21-1. Pre-operative serum CYFRA21-1 levels in patients with ovarian carcinoma correlated with carcinoma status and tumour stage. Also pathologic complete response rate was significantly greater in patients with low pre operative CYFRA21-1 (Gadducci *et al* 2004).

The expression of cytokeratins 8, 18, 19 and 20, at the protein or mRNA level, have been evaluated for use as bladder cancer markers. Cytokeratin 20 (CK20) had a sensitivity of 78-87% but a low specificity of 55-63% and cytokeratin 19 showed a sensitivity of 75-97% with a specificity of 67-71%. Cytokeratin markers have high false positive rates among individuals with a wide range of clinical disorders which questions their clinical utility (Lokeshwar *et al* 2005).

Attallah et al (2006), using an office-based dot-enzyme immunobinding assay (dot-EIA), evaluated Cytokeratin 10 (CK-10). CK-10 was shown to have a high sensitivity (97%) and specificity (94%) in detecting CK-10 from the urine of 192 bladder cancer patients and 72 normal controls. Attallah *et al* (2006) found that CK-10 could not be detected in patients showing response to radiotherapy.

Gacci *et al* (2006) used a Urinary Bladder Cancer (UBC) test on the urine of 36 bladder cancer patients. The urine samples were taken pre and at four time-points post transurethral

resection. The UBC test tests for urinary fragments of cytokeratin 8 and 18. They found a correlation between the postoperative modifications of UBC levels and the risk of tumour recurrence during the first year of follow-up.

The importance of carrying out immunohistochemical analysis of CK1, CK7 and CK14 was found by Coletta *et al* (2006) to be very useful in distinguishing basaloid squamous carcinoma (BSC). Basaloid squamous carcinoma is a relatively rare and aggressive variant of squamous cell carcinoma. Correct diagnosis of this tumour as the treatment and prognosis of BSC are different to other neoplasms e.g. adenosquamous carcinoma or small cell undifferentiated carcinoma. Analysis of combinations of cytokeratin expression has also been found useful in diagnosing specific cancer types from each other. Combined immunohistochemistry analysis of beta-catenin, CK7 and CK20 was found by Ikeda *et al* (2006) to aid in the discrimination of primary lung adenocarcinoma from metastatic colorectal.

1.2.6 EGF receptor family

The EGF receptor family has four members, EGFR1 (HER1), ErbB2 (HER2), HER3 and HER4. The members of the family can form four homodimers and six heterodimers to give a total of 10 distinct states. Receptor dimerisation, protein kinase activation, *trans*-autophosphorylation and initiation of signalling cascades results from the binding of ligands to the ectodomain (Schelessinger 2000). The four members have the potential to stimulate the Raf-MEK-ERK protein kinase cascade (Serrano- Oliver *et al* 2006).

The EGF family of ligands are expressed in the extracellular domain of transmembrane proteins and are generated by regulated proteolysis to yield growth factors that contain 49-85 amino acids. These growth factors have different specificities for the receptor family members and also the transforming potential and signalling pathways activated by different

dimmers are distinctive. Examples include the high affinity for ErbB1 that both epidermal growth factor (EGF) and transforming growth factor α (TGFα) have, and the fact that ErbB1-ErbB2 heterodimers are associated with a more robust signal than ErbB1-ErbB1 homodimers (Serrano-Olivera *et al* 2006).

1.2.6.1 Epidermal growth factor receptor

Epidermal growth factor receptor (EGFR) is a cell membrane tyrosine kinase receptor that belongs to the ErbB family. When it is activated by its natural ligands it initiates a complex intracellular signal transduction cascade that promotes cancer cell division and migration, angiogenesis, and apoptosis inhibition (Galizia *et al* 2006).

Galizia *et al* (2006) studied the expression of EGFR using immunohistochemistry (IHC) on tumour tissues for 149 colon cancer patients in relation to other molecular markers, p27, p53, and vascular endothelial growth factor (VEGF). The prognostic role of EGFR expression in 126 colon cancer patients that were undergoing potentially curative surgery was also examined. They found EGFR expression in 36% of tumour samples and it did not correlate with tumour grade, growth pattern and microsatellite status but did correlate with Dukes' stages. In patients with early-stage colon cancers (classified as Dukes' A and B), that were not treated with adjuvant chemotherapy, and were positive for EGFR, showed a reduction of nearly 50% in 5 year survival probability compared to EGFR negative patients. In Dukes' C and D stages, the five year survival rate of EGFR positive tumours was cut to 20.3%. The study suggests a potential for treating colorectal cancer with anti- EGFR targeted therapies although the authors noted that previous studies looking at the expression of EGFR have shown great variability.

As the human epidermal growth factor receptor 2 (HER2) is known to heterodimerise preferentially with the EGFR, Emlet *et al* (2006) sought to investigate whether HER2 had a

role as a predictive marker for responsiveness to EGFR targeted therapies as tumour EGFR expression was previously found not to be useful as a predictive marker of clinical response to EGFR- targeted therapies. They tested EGFR, HER2 and HER3 markers. HER2 overexpression was found to be the best single predictive marker, but combinations of two markers provided additional predictive information.

1.2.6.2 HER2/NEU

Human epidermal growth factor receptor 2 (HER2) is an 185kDa glycoprotein normally expressed in the epithelia of numerous organs for example lungs, bladder pancreas, breast and prostate. HER2 is a member of the epidermal growth factor receptor (EGFR) family, which is comprised of 4 family members. The HER2/*neu* receptor has three domains: an intracellular tyrosine kinase portion, a hydrophobic transmembrane domain, and an extracellular domain (ECD), which is the ligand-binding portion of the receptor. The ECD undergoes proteolytic cleavage from the full-length protein by metalloproteases, and is shed into the blood as a circulating antigen (Codony-Servat *et al* 1999). This serum HER2/*neu* is a glycoprotein of 97 to 115kDa and is also known as p105 (Lüftner *et al* 2003). Proteolytic cleavage of the ECD represents a ligand-independent activation mechanism, which leads to HER2/*neu* phosphorylation and active HER2/*neu* signalling (Köstler *et al* 2004).

Immunohistochemistry can be used to detect the degree of HER2/*neu* protein over expression in paraffin-embedded tissue samples using mono- or polyclonal antibodies. These antibodies bind to the HER-2*neu* expressed on the cell membranes. Loss of antigenic determinants during the tissue fixation process can compromise the results.

As mentioned earlier, the ECD of HER2/*neu* can be proteolytically cleaved and this 105kDa molecule shed into serum. Circulating HER2 ectodomain can be detected in about 80% of patients with tumours over expressing HER2/*neu* compared to 3% of tumours not

over expressing the oncoprotein (Seregni *et* al 2004). A negative prognostic effect of high serum levels of HER2/*neu* seems to be related to resistance to chemotherapy (Pegram *et al* 1998, Mehta *et al* 1998).

Elevated serum HER2/*neu* levels in breast cancer patients following primary surgery or immediately after the completion of adjuvant chemotherapy provide prognostic information. Retrospective studies have shown that patients with elevated levels detectable in serum may benefit from specific anti-HER2/*neu*- target treatment options early in the course of their disease (Lüftner *et al* 2003).

Köstler *et al* (2004) found that patients with elevated serum ECD levels were more likely to respond to Herceptin- based treatment. Herceptin is a monoclonal antibody that targets the HER2/*neu* oncoprotein. The authors also found that changing ECD levels throughout the treatment could not only parallel but could precede the clinical course of disease. This allowed for a significant prediction of response, clinical benefit, and progression-free survival as early as after 1 and 2 weeks of treatment, respectively.

HER2/*neu* gene amplification and/ or protein over expression are detected in approximately 25 to 30% of primary breast cancers (Slamon *et al* 1987) and are associated with poor prognosis, short survival and recurrences (Seregni *et al* 2004). Fluorescent-in situ-hybridisation (FISH) methods can be used to measure the HER-2/*neu* gene copy number per nucleus to identify if there is DNA amplification.

1.2.7 Cancer cell- secreted proteins

Wu *et al* (2005) examined the conditioned media of nasopharyngeal cancer cell lines for proteins secreted by the cancer cells in an attempt to identify biomarkers for early nasopharyngeal carcinoma (NPC) diagnosis. MALDI-TOF mass spectrometry was used to identify secreted proteins. The secreted proteins detected were then assessed in NPC

biopsies using IHC. Three proteins, fibronectin, Mac-2 binding protein (MAC-2 BP) and plasminogen activator inhibitor 1 (PAI-1) were found to be highly expressed in NPC biopsies but weakly or not expressed in normal nasopharyngeal tissue. Analysis of serum levels of the proteins in samples from 46 NPC patients and 47 healthy controls using ELISA identified that the three proteins were significantly higher in the NPC patients than the normal controls.

1.2.8 Exploitation of the immune system for cancer biomarkers

The release/ secretion of proteins from tumours triggers a response in cancer patients. These tumour antigens arise from several mechanisms, including tumour-specific alterations in protein expression, mutation, folding, degradation, or intracellular localisation. Antibody immune responses show promise as clinical biomarkers because antibodies have long half-lives in serum, are easy to measure, and are stable in blood samples.

One of the most well studied auto-antigen in cancer is the p53 protein. The half-life of mutated p53 is much longer than wild type p53 (several hours compared to several minutes). The p53 auto antibodies are dependent on the type of p53 mutation but the accumulation of protein, rather than the mutations results in autoantibody generation (Anderson and LaBaer 2005). p53 auto antibodies have also been detected in heavy smokers prior to the diagnosis of lung cancer (Trivers *et al* 1996).

1.3 Detection of circulating tumour cells in cancer patients.

Disseminated tumour cells can be detected in blood by using RT-PCR methods to detect certain tumour–associated antigen gene transcripts or cell lineage-specific markers. Such markers however (for example cytokeratins which identify cells of an epithelial origin), are not tumour specific and may be expressed by non-neoplastic cells (Vogel and Kaltoff

2001). Another concern is that the increasingly more sensitive RT-PCR techniques are capable of detecting illegitimate transcription i.e. expression of transcripts at such low levels that the antigen is not expressed and so cannot be relevant for cellular function (Zieglschmid *et al* 2005). Although if detection of the transcript alone is useful as a biomarker, relevance to cellular function may not be important.

There are many studies assessing tumour markers by RT-PCR methods on whole blood. Many of these studies make the assumption that presence or absence of a tumour-related transcript in whole blood from cancer patients is indicative of circulating cancer cells. Studies by Vasioukhin *et al* (1994) and Kopreski *et al* (1997), however, have shown that presence of extracellular tumour DNA can be detected in the absence of circulating cancer cells and in patients with early disease and low tumour burden. The following is a summary of some recent studies employing RT-PCR methods to detect mRNA transcripts in whole blood samples.

Ijichi *et al* (2002) looked at the clinical significance of using AFP mRNA as a predictive marker of HCC cells disseminated into the circulation and for metastatic recurrence. They found that patients with consistent positive AFP mRNA showed the highest recurrence rate (85%) and a trend to distant or multiple recurrences. If the patients AFP mRNA status went from positive pre treatment to negative post-adjuvant treatment, their overall, and disease-free, survival rates were much better than those with permanent positive AFP mRNA.

Minata *et al* (2001) found that the presence of AFP mRNA pre-operative was significantly correlated with tumour size and presence of AFP mRNA post-operatively indicated metastatic recurrence. Meta-analysis carried out by Ding *et al* (2005) on several publications regarding the expression of AFP mRNA one week after surgery (including the study mentioned above by Ijichi *et al*), and they concluded that expression of AFP mRNA in peripheral blood 1 week after surgery correlated with the recurrence of HCC and was a

useful predictor for tumour recurrence.

Although other reports have noted that circulating AFP mRNA was transiently detectable in patients with cirrhosis no predictive value was found for HCC development and recurrence (Gross-Goupil *et* al 2003).

In 2003 Schamhart *et al* performed a multi-centre study to investigate methods for improving prostate-specific antigen (PSA) mRNA detection, by defining criteria required to reduce false positive results. These criteria included positive results in two or more PCR tests depending on the number of PCR cycles used. The nested PCR assay was carried out using either 2 x 25 PCR cycles or 2 x 35 cycles. Analysis of the results showed that for any specimen that was positive after 2 x 35 cycles of PCR, a diagnostic specificity could only be assigned if the specimen was positive in three or more additional tests. The pre-defined diagnostic specificity was set at ≥ 98% and was weighted by the best estimates of the mean, λ (SD), of the expected Poisson frequency distributions of the number of positive tests. The authors were able to show that compared to the high sensitivity test (35 x 2 PCR cycles, λ=1.0), the less sensitive but more relevant (25 x 2 PCR cycles λ= 0.2) test, was affected to a lesser extent by illegitimate PSA transcripts without a reduction in detection of true positives.

Human mammaglobin mRNA (hMAM) was studied by Roncella *et al* (2005) to evaluate its use as a molecular marker for detection of occult breast cancer cells in peripheral blood. The nested RT-PCR assay detected hMAM in 156/165 (95%) of fresh breast cancer tissues analysed, but in only 12% (16/ 137) of peripheral blood specimens. They found that their assay was sensitive down to one tumour cell in 10^6normal cells. No hMAM transcripts were detected in either 66 healthy blood donors or 151 patients with benign breast disease. Although the assay proved to be very sensitive, the authors noted that the 5% of stage I-II

patients was lower than the expected percentage of these patients to have circulating cancer cells.

Another study performed in 2006 by Mercatali *et al* evaluated the presence of mammaglobin B mRNA (homolog of hMAM mRNA) in conjunction with maspin mRNA and CK-19 mRNA in peripheral blood of 140 breast cancer patients and 27 healthy volunteers. Evaluation of CK-19 mRNA was abandoned at an early stage, due to a high number of false positive results. Maspin mRNA was detected in 24% of breast cancer patients but also in 11% of healthy volunteers. Mammaglobin mRNA was detected in 7% of breast cancer patients but in none of the blood samples from healthy volunteers.

Higher levels of CK-19 in both bone marrow and peripheral blood has also been shown to be a significant predictive marker for breast cancer recurrence (Nakata *et* al 2000, Nogi *et al* 2003), and for nasopharyngeal carcinoma micrometastasis (Lin *et al* 2002).

In a large study performed by Soeth *et al* 2005, a correlation was found between cytokeratin 20 (CK 20) expression in pancreatic ductal carcinoma patients and poor survival. They analysed bone marrow and venous blood in 172 patients with pancreatic ductal carcinoma. Of 135 bone marrow samples, 45 (33.3%) were positive for CK 20 mRNA expression and 52 of the 154 (33.8%) blood samples were positive for CK 20 expression. The authors also found that detection rates increased with tumour stage. The authors measured detection of CK 20 mRNA and expression was determine to be indicative of disseminated tumour cells but they failed to address the false positive results. Whilst the expression of CK 20 was found to be associated with stage of tumour progression, 11.8% of bone marrow samples (8/68) and 16.7% of blood samples (9/54) from patients with non malignant diseases were positive for CK 20 expression. Also 5% (1/20) blood samples from healthy volunteers were positive for CK 20 expression.

Koyanagi *et al* (2005) developed a multi-marker quantitative real-time PCR method to

detect circulating melanoma cells. The multi-marker panel included MART-1, GalNAc-T, PAX-3 and MAGE-A3 mRNA transcripts. The study was carried out by firstly assessing the detection limits and imprecision of the assays by diluting melanoma cells originating from melanoma cell lines expressing all four markers into peripheral blood samples from volunteers. This was repeated with 3 different cell lines. 1 melanoma cell could be detected in 10^7 peripheral blood leukocytes. Detection rates of the multi-marker quantitative real-time PCR assay were higher than any of the individual markers alone. Detection rates were also significantly different depending on stage of disease, 15% detection of at least one marker in stage I patients compared to 86% detection of at least one marker in stage IV patients. The markers were not detected in healthy donor blood using the optimised assay.

Microphtalmia transcription factor (Mitf) mRNA which is important in melanocyte development and melanoma growth, was assessed using real-time quantitative RT-PCR for expression in peripheral blood as a marker for detection of circulating tumour cells in melanoma patients. Employment of cut-off values aided in discrimination of healthy samples from cancer samples. The authors found that in 90 melanoma patients tested, the rate of Mitf mRNA detection was higher with increasing cancer stage. Comparisons were made with samples taken post bio chemotherapy (consisting of cisplatin, dacarbazine, vinblastine, interleukin-2, α-IFN and granulocyte colony-stimulating factor at different concentrations on different days) and post surgery and Mitf mRNA detection decreased with treatment. Detection of the mRNA post treatment was associated with a significantly lower relapse-free and overall survival.

CEA mRNA was found to correlate with depth of tumour invasion, vessel invasion, TNM stage and postoperative metastasis in a study by Wang *et al* 2006, in which four mRNA transcripts were evaluated for significance in predicting colorectal cancer. The other three markers evaluated were human telomerase reverse transcriptase (hTERT), CK-19, and CK-

20. Detection of hTERT, CK-19, CK-20 and CEA mRNA transcripts were found in 69.4%, 66.7%, 52.8% and 72.2% of 72 patients respectively. No hTERT or CEA mRNA transcripts were detected in samples from 30 healthy volunteers however; CK-19 or CK-20 mRNA transcripts were detected in 2 of the 30 volunteers. The authors found that only detection of positive CEA mRNA correlated directly with postoperative metastases however only 58% of the CEA positive patients metastasised in the course of the study.

Uen *et al* (2006) evaluated MUC1 and c-Met mRNA for use as markers for detection of circulating tumour cells in patients with gastric cancer by. A multiplex PCR studied was designed for analysis of RNA extracted from whole blood. Detection of c-Met or MUC1 mRNA was significantly correlated with the depth of tumour invasion, lymph node metastasis, TNM stage, vessel invasion, peri-neural involvement and post operative metastasis. 5.6% and 8.3% of blood samples from healthy donors were positive for c-Met and MUC1 mRNA respectively.

Stathopoulou *et al* (2006) investigated whether the detection of CK-19 in healthy individuals in their prior studies was due to poor primer design. New primers were designed for CK-19 which avoided amplification of contaminating genomic DNA and CK-19 pseudogenes. Real-time RT-PCR was carried out using RNA extracted from mononuclear cells isolated by gradient centrifugation with Ficoll Hypaque-1077 from 222 peripheral blood samples. Of 62 samples from healthy volunteers, none were positive for CK-19 in comparison to 2 positive using previous poorly designed primers with a further 47 showing late amplification of CK-19 negating the need to employ a cut-off level to distinguish cancer patients from normal patients. 20% of the samples from breast cancer patients were positive for CK-19 mRNA.

Detection of CK-19 mRNA in the peripheral blood of breast cancer patients was used as a marker to assess the effective targeting of circulating tumour cells by the anti-HER-2

antibody, Herceptin. The authors noted that chemotherapy resistant tumour cells found in circulation expression of HER-2 on the disseminated tumour cells had previously been found to be an independent prognostic factor associated with increased risk of relapse and survival. The authors found that before Herceptin infusions, CK-19 positive cells were found in either the peripheral blood or bone marrow or in both samples corresponding to 25/30 breast cancer patients. HER2 mRNA expression was found in the CK-19 positive samples. After infusion treatment with Herceptin, 93% of patients became negative by nested RT-PCR and 67% by real-time PCR. After a median follow-up of 6 months the median duration of CK-19 mRNA negativity was 9, 12 and 6 months for stage I/II, III and IV disease respectively.

It is evident from the studies mentioned that the incidence of mRNA transcripts being detected in the whole blood of healthy volunteers is to a certain extent accepted. The re-design of CK-19 primers by Stathopoulou *et al* (2006) indicates the possibility that other studies carried out showing detection of mRNA transcripts in healthy donor blood may also be a reflection of poor primer design and/ or presence of splice variants/pseudogenes not yet identified.

A lot of the studies, as mentioned earlier, state that detection of tumour-related transcripts was indicative of circulating tumour cells but few studies have actually confirmed that the mRNA detected originates from circulating tumour cells and not from extracellular mRNA transcripts. Most of the studies mentioned above make the assumption, rightly or wrongly, that the presence of tumour-related mRNAs in the serum of cancer patients indicates the presence of cancer cells. Whilst this assumption may be okay to make if the results obtained indicated that the mRNAs were cancer-specific i.e. no detection of the transcript in healthy controls/ detection under a predefined threshold, many of the studies mentioned above suffered from positive results in the healthy controls. Whilst this is fine in a

retrospective study, a false discovery rate of 5% is too high for prospective studies meaning that for every 1000 people screened for the presence of the tumour-related biomarker, 50 healthy people will be falsely identified.

To confirm that transcripts found in circulating blood represent circulating tumour cells, the ideal method is to separate the cells found in the blood from serum and analyse and carry out RT-PCR for detection of tumour-related mRNAs on these cells and not blindly on whole blood samples.

Detection of erbB2 mRNA transcripts in the plasma of breast cancer patients was found to be associated with circulating tumour cells and negative estrogen and progesterone receptor status by Xu *et al* (2006). In this study the authors separated the plasma of the breast cancer patients from the mononuclear cells by Ficoll. RT-PCR was carried out for erbB2 mRNA on the plasma sample and detection of hMAM transcripts in RNA extracted from the mononuclear cells was taken to be indicative of circulating tumour cells mixed in with mononuclear cells. ErbB2 mRNA transcripts were detected in 46% of patients with cancer and in 10% of plasma taken from healthy volunteers.

Expression of hMAM mRNA in the mononuclear cell fraction was coupled in 7/7 cases by erbB2 expression in the corresponding plasma sample. This study however, despite separation of the mononuclear blood cells from the plasma fraction, was slightly flawed in that firstly the plasma samples were not re-centrifuged/ filtered to prove there were no tumour/ mononuclear cells still present in the plasma sample. The second flaw is that they did not assess both samples for the detection of both mRNAs. This would have highlighted whether the positive detection of erbB2 in the plasma of 43% of breast cancer patients corresponded to positive detection of erbB2 transcripts from the mononuclear cell fraction of the blood samples and vice versa for the detection of hMAM.

A study by Stathopoulos *et al* (2005) found that detection of CK-19 mRNA transcripts in

cells that were separated from whole blood using Ficoll-Hypaque did not indicate relapse of breast cancer patients. They found that adjuvant chemotherapy and hormone treatment resulted in the disappearance of the CK-19 positive cells in all node negative patients and in 15/31 node positive patients. However after a 40 month follow-up, all node negative patients that were positive for CK-19 mRNA transcripts were relapse-free.

Traditional enrichment techniques using conventional density gradients like Ficoll result in a large number of cells in the enriched fraction however consisting mainly of blood mononuclear cells. In 2005, Müller *et al* tested the efficiency of an enhanced density gradient system using a device called the OncoQuick density gradient system (Greiner Bio-One GmbH). The OncoQuick system comprises a liquid density separation medium and a porous barrier membrane optimised for the enrichment of circulating tumour cells. The enriched cells were then cytocentrifuged at a concentration of $5x10^5$ cells per $240mm^2$ areas. The authors studied the slides for recovery of MCF 7 breast cancers cells that were spiked into 18mls of blood from healthy donor. Serum taken from breast cancer patients were also examined for the presence of the tumour marker CA15.3 and bone marrow aspirations from the same patients were analysed using the OncoQuick enrichment gradient system. The authors found that 39.7% of patients with metastatic disease showed circulating tumour cells and a positive correlation was found with the serum levels of the tumour marker CA15.3. Circulating tumour cells were found more frequently in patients with progressive disease than in patients with stable disease or remission (87.5% compared to 43.8%) as assessed by frequent sampling of the patients.

Other methods have also been developed to enrich blood samples for tumour cells. These methods include antibody based methods which are dependent on the ability of antibodies to distinguish between cells of different tissue origin. The most widely used antibodies are those that target cytokeratins as they are markers of epithelial cell origin and would not

normally be expected to be found in the blood of healthy people. Analysis of cytokeratin positive staining cells in blood samples has traditionally been carried out using immunocytochemistry, where the cells are fixed to glass slides and analysis is carried out. However the low reported sensitivities of immunocytochemistry detection of disseminated tumour cells at approximately one cell in every 10^5 or 10^6 mononuclear cells may ultimately mean that circulating tumour cells are missed in these analyses because so many slides have to be screened in order to detect circulating tumour cells (Müller *et al* 2005).

In a study by Pachmann *et al* (2005), non-specific staining of fixed cells using anti-epithelial cell specific adhesion molecule (anti-EpCAM) was overcome by using a live cell analysis system called MAINTRAC®. This method enables the relocation of cell for visual examination and quantification, and for taking fluoromicrographs. They found that they were able to monitor the response of the whole tumour in breast cancer patients to adjuvant therapy by using the MAINTRAC® system to analyse the response of circulating epithelial cells. A reduction in the number of epithelial cells in circulation accurately predicted final tumour reduction at surgery.

In 2004, Demel *et al* reported a new sensitive and specific immuno-molecular assay for the detection of malignant cells in the peripheral blood of patients with solid tumours. With their combined antibody-RT-PCR technique they were able to detect 2 tumour cells in 5mls of blood. This was carried out by firstly enriching 5 mls of blood, from healthy donors that had been spiked with 2 breast cancer cells, by using 100µl of an antibody cocktail fixed on immunomagnetic beads. This antibody cocktail included 3 different monoclonal antibodies, ber-ep 4 (an epithelial membrane marker), Muc1 (a mucking marker) and GP1.4 (a glycoprotein marker). After incubation with the antibodies (positive immunomagnetic separation), and washing of cells, the cells were lysed and mRNA recovered by a magnetic separation. Subsequent multi-plex PCR was carried out on the reverse transcribed cDNA

for 4 tumour markers, Ga 733.2, Ca 15.3, Her 2 and Claudin 7. All four markers were amplifiable by PCR in the "spiked" blood samples and negative in blood samples where no breast cancer cells were added. This test is commercially available as AdnaTest BreastCancerDetect (AdnaGen) and all reagents e.g. magnetic beads and pre-mixed primer/ PCR constituents are provided.

Zieglschmid *et al* (2005) reported a similar success with the use of multiple monoclonal antibody enrichment for tumour cells. The cells were enriched by using combinations of monoclonal antibodies coupled to Dynabeads® to target the tumour cells; MOC31, 8B6 and Ber-EP4 (for the enrichment of testicular cancer cells) and MOC31 and Ber-EP4 (for the enrichment of colorectal cancer). Detection of testicular and colorectal cancer cells spiked into normal healthy blood was possible by using multiplex PCR. The PCR primers contained different amounts of primer pairs depending on the cancer cells under investigation. The authors reported a synergistic effect with enrichment using a combination of antibodies over enrichment using single antibodies.

Guo *et al* (2005) used an immunomagnetic assay to enrich peripheral blood specimens from colorectal cancer patients for tumour cells. The expression of CK-20 was evaluated in the enriched cell samples. CK-20 mRNA expression was found to correlate with tumour diameter, lymphatic metastasis and hepatic metastasis.

Zigeuner *et al* (2003) tested methods to positively select epithelial cells with the anti epithelial antibody BER-EP4 and deplete mononuclear cells with an anti-leukocyte antibody CD45 in order to detect tumour cells that had been added to whole blood specimens. The authors found that enriching for tumour cells using either positive or negative selection detected tumour cells at a rate significantly higher than compared to unenriched cell populations by immunocytochemistry. The tumour cell recovery was 43.5% for positive selection and 32.6% for negative selection.

Ulmer *et al* (2004) immunomagnetically isolated circulating melanoma cells using a murine monoclonal antibody targeted to the melanoma-associated chondroitin sulphate proteoglycan in blood specimens taken from 164 melanoma patients and 50 volunteers without malignant disease. In addition to this single-cell comparative genomic hybridisation (SCOMP) was carried out on 15 individual cells isolated from 7 patients to define their chromosomal aberrations, and so prove their malignant origin. The authors found that detection of two or more cells correlated significantly with a reduced survival of patients with metastatic melanoma. All cells analysed by SCOMP displayed multiple chromosomal changes and carried aberrations typical for melanoma.

A negative immunomagnetic assay was utilised by Wiedswang *et al* (2006) to enrich peripheral blood samples for the presence of circulating tumour cells resulting in increased sensitivity. The negative immunomagnetic assay is carried out by incubating peripheral blood mononuclear cells (isolated by Ficoll density centrifugation) with anti-CD45 conjugated M450 Dynabeads and placing into a neodymium magnetic particle concentrator. Cells were then collected and cytospins were prepared. Of 34 patients that were found to be positive for circulating cancer cells, 9 relapsed (7 of whom systematically relapsed) with death resulting from breast cancer occurring in 6 of those patients. However the authors found that detection of patients at risk for relapse was superior using immunocytochemistry for detection of tumour cells in the bone marrow.

1.4 Cell-free Nucleic Acids in Serum/ Plasma

Mandel and Métais reported the finding of extracellular nucleic acids in plasma in 1948. DNA and RNA were observed in the plasma of healthy and sick individuals although the significance was not recognised at the time.

1.4.1 Serum / Plasma DNA

1.4.1.1 Cancer Testing

Studies in more recent years showed that the concentration of circulating DNA in patients with metastatic disease was higher, a level at which for some patients decreased with successful anticancer therapy (Leon *et al* 1977). It has since been proven that tumour cells can release DNA into circulation. Two reports in 1994 demonstrated the presence of tumour-derived oncogene mutations in the plasma of patients with pancreatic cancer, myelodysplastic syndrome or acute myeologenous leukaemia (Sorenson *et al* 1994, Vasioukhin *et al* 1994). In 1996, two groups reported the presence of tumour- associated microsatellite alterations in the plasma and serum of cancer patients, suggesting that, in these cases most of the plasma/ serum DNA is tumour-derived (Chen *et al* 1996, Nawroz *et al* 1996).

Studies carried out using serum DNA from cancer patients can be grouped as follows (1) Quantification of total DNA, (2) Quantification of gene expression (qPCR), (3) Detection of microsatellite alterations including loss of heterozygosity (LOH), (4) Detection of mutations in specific genes and (5) Detection of aberrantly methylated gene promoter regions.

The case number studied in the literature varies greatly (6- 100) along with the control group number (0- 100). At the moment there is only one study in which the control group have been age, sex and smoker matched (lung cancer study- Sozzi *et al* 2003).

1.4.1.1.1 Quantification of total DNA

Sozzi *et al* carried out two studies based on quantifying the total amount of DNA in the plasma of lung cancer patients. In the first paper published in 2001, 81 patients with NSCLC were studied during follow-up to see if their mean plasma DNA concentrations

were higher than in the 38 controls. They found that in 35 clinically relapse-free individuals their mean plasma DNA levels were significantly lower than those recorded at time of surgery (~345ng/ ml before surgery to ~34ng/ml). Four patients that showed a 2-20 fold increase in the amount of plasma DNA during follow-up went on to present with liver metastases (2), recurrence of carcinoma (1) or new primary tumour of the liver (1).

A study carried out by Gal *et al* (2004), showed that the level of serum DNA in different groups of breast cancer patients was significantly higher than healthy controls. 96 primary breast cancer patients showed a mean serum DNA concentration of 221ng/ml compared to 63ng/ml from the 24 healthy controls. The serum DNA levels were elevated independently of the size of primary tumour or lymph node metastases.

1.4.1.1.2 Quantification of gene expression

Quantification of MYCN DNA in serum/ plasma was found by Combaret *et al* (2002) to help identify disease progression in neuroblastoma patients during follow- up after chemotherapy. They found that the levels of circulating MYCN DNA were comparable at diagnosis and relapse.

A second study by Sozzi *et al* in 2003 relied on the quantification of the human telomerase reverse transcriptase gene (hTERT). 100 NSCLC were studied and included 100 controls. The amount of plasma DNA corresponding to the hTERT gene was quantified using real-time quantitative polymerase chain reaction (QPCR). They found that the amount of hTERT plasma DNA in the cancer patients was almost eight times the value detected in controls (24.3ng/ml compared to 3.1ng/ml).

1.4.1.1.3 Detection of microsatellite alterations including LOH

Narow *et al* and Chen *et al* first reported the presence of microsatellite DNA in serum in 1996. Narow *et al* studied the DNA from serum and corresponding lymphocytes and tumours of patients with head and neck cancers. They were then scored for alterations, either presence of new alleles or LOH at each of 12 markers. Six out of 21 patients (29%) had one or more microsatellite alteration, which matched the primary tumours. Chen *et al* looked at the plasma of SCLC patients. They found that microsatellite alterations were present in 16/21 (76%) tumours and in 15/21 (71%) corresponding plasma samples. Thirty of 49 (61%) of NSCLC patient plasma studied by Sozzi *et al* (1999) showed a change in plasma DNA that correlated to the corresponding tumour. The plasma DNA abnormalities were detectable in approximately 44% of either stage I cases or those with tumours up to 2cm in diameter showing that this type of study is applicable to early and late cancers.

Microsatellite alterations and LOH studies have been carried out on tumours of various

cancers. Panels of markers have been studied depending on the cancer type. One of the most common microsatellite alterations studied in the serum/ plasma of lung cancer patients is the fragile histidine triad (FHIT). FHIT is a tumour suppressor gene that is often has a high rate of LOH in lung cancer patients especially in smokers (Sozzi *et* al 1999). A lot of the analyses published since follow a similar trend (Bruhn *et al* 2000, Sozzi *et al* 2001, Chang *et al* 2002, Beau-Faller *et al* 2003, and Adriani *et al* 2004). The repeated studies suggest that the more microsatellite alterations and LOH that are found in the plasma/ serum DNA of cancer patients after surgery/ chemotherapy the poorer the prognosis.

1.4.1.1.4 Detection of mutations in specific genes

Two of the most widely studied mutated genes detectable in the serum of cancer patients are the TP53 gene and the K-ras oncogene. Most of the studies involving detection of TP53 gene mutations in the plasma/ serum of a number of cancers look at its detection in combination to microsatellite alterations and LOH. Gonzalez *et al* (2000) found that 71% of lung cancer patients studied had either TP53 mutations or microsatellite modifications. Gocke *et al* (2000) found it was possible to detect p53 mutations in the plasma of colorectal cancers. Silva *et al* (2002) looked at the microsatellite alterations and TP53 mutations in the plasma of breast cancer patients to see if presence of the tumour-related DNA was a predictive factor relating to recurrence and disease-free survival. They found that 61/142 (42.9%) patients showed molecular alterations in their plasma DNA. During follow-up they found that 17/23 (74%) of patients that recurred showed presence of tumour-related DNA persistent in their plasma. Adriani *et al* (2004) used a panel of genetic markers to study to plasma DNA of lung cancer patients. They found that at least 1 of the three genetic markers studied (p53 mutation, FHIT and 3p LOH) was identified in the plasma of 51.6% of all patients and 60.7% of stage I patients.

Kopreski *et al* reported in 1997 that they could detect mutant K-ras in the plasma/ serum of 12/31 (39%) of colorectal patients and in none of 28 control samples. A later study again by Kopreski *et al* (2000) was carried out on the plasma of 232 colorectal cancer patients and that K-ras mutations were detectable in 28% of these patients. Out of the patients that had tissue available (135) K-ras mutations were found in 35 (26%) of these patients with 29 (83%) of the correlating plasma samples positive also for the mutation. Detection of K-ras mutation in serum post curative resection surgery has also been shown to correlate with disease recurrence (Ryan *et al* 2003).

1.4.1.1.5 Detection of aberrantly methylated DNA

Promoter methylation studies of plasma/ serum DNA have been used to quantify the methylation status of a number of oncogenes. Esteller *et al* (1999) probed for detection of aberrant methylation status of tumour suppressor genes in serum DNA of NSCLC patients. The authors looked at the methylation status of genes that were found to be commonly hypermethylated in tumours of NSCLC patients. The methylation status of p16, DAP kinase, GSTP1 and MGMT promoters were studied in 22 NSCLC patients using primary tumours and paired normal tissue and serum specimens. Of the 15 primary tumours that showed aberrant methylation of at least one of the promoters studied, 11 also had abnormal methylated DNA in the matched serum specimens with no false positives. The detection rate in plasma/ serum of ~50% of methylated promoter DNA (corresponding to the tumour) was also noted by Usadel *et al* 2002. They found that out of 95 lung cancer patients that had detectable methylation of the Adenomatous Polyposis Coli (APC) promoter, 42 also had methylated APC DNA in their serum/ plasma.

Aberrant methylation of the p16 gene has been detected in the plasma/ serum of a wide variety of cancers: Liver (Wong *et al* 1999), NSCLC (Esteller *et al* 1999), Head and Neck

(Sanchez- Cespedes *et al* 2000, Wong *et al* 2003), Esophageal SCC (Hibi *et al* 2001), Glioblastoma (Ramirez *et al* 2003), Colorectal (Nakayama *et al* 2003) and Nasopharyngeal (Wong *et al* 2004). The occurrence of promoter methylation of this gene in plasma/ serum specimens varied from 23- 66% in the different cancer groups. Where no aberrant methylation of the gene was found in the tumour, the plasma/ serum specimens also showed no methylated p16 DNA. None of the healthy controls in any of these studies showed aberrant methylation of the p16 gene. Wong *et al* (2004) also found a correlation in the detection of aberrant methylation post surgery and recurrence of disease. None of the patients they studied that were in remission showed evidence of serum p16 methylation, while 38% of those that recurred had detectable levels of p16 methylation in their plasma.

Most of the publications use methylation- specific PCR (MSP) to determine the methylation status of gene promoters. Either the promoter sequence is methylated, unmethylated or both. Some of the more recent publications report a quantitative method using real-time PCR (Usadel *et al* 2002, Wong *et al* 2003a, Wong *et al* 2003b, Widschwendter *et al* 2004, and Wong *et al* 2004). This new method has been shown to have a 10 fold greater sensitivity compared to traditional MSP methods (Usadel *et al* 2002).

1.4.1.1.6 Detection of viral sequences

Quantitative analysis of cell-free Epstein-Barr virus DNA in the plasma of patients with nasopharyngeal carcinoma (NPC) has been found to be a useful way of monitoring NPC patients. Advanced-stage patients were found to have higher plasma EBV DNA levels than those with early-stage disease. A month after radiotherapy was completed, plasma EBV DNA could not be detected in 7 of 15 subjects but levels were still high in the remaining 8 patients. After further clinical examination all of the seven with no detectable levels of the virus had complete tumour regression while six of the eight patients with high levels of

plasma EBV DNA showed tumour persistence or distant metastasis (Lo *et al* 1999).

Epstein-Barr virus (EBV) DNA has been detected in the serum of patients with nasopharyngeal carcinoma (NPC). A more recent study has shown that these EBV DNA molecules are "naked" and not contained in virions. Also the DNA fragments were found to be mostly shorter than 181bp (Chan *et al* 2003). Other cancers associated with EBV have shown a correlation between EBV levels and disease progression (Ngan *et al* 2002). Quantitative detection of human papillomavirus (HPV) DNA sequences in patients' sera was found to be a potential marker of early metastatic disease in HPV-associated head and neck squamous cell carcinoma (HNSCC) (Capone *et al* 2000). Detection of HPV DNA in plasma of cervical cancer patients was associated with disease metastasis (Pornthanakasem *et al* 2001). HPV DNA detection in serum of early-stage cervical cancer patients was found to be associated with poor prognosis factors such as lymphovascular invasion and deep stromal invasion. The serum from 112 patients with cervical cancer was probed for presence of the virus, 27 of the samples were positive (24.1%) while none of the patients with carcinoma in situ or benign disease were positive (Hsu *et al* 2003). Ho *et al* (2005) found a link between the detection of viral DNA in the blood of cervical cancer patients within three months of treatment and recurrence of the disease, 87.5% of positive samples involved distant metastasis a summary of some of the publications relating to serum DNA are listed in Table 1.4.1.1.

Mutation/ virus	Reference	Tumour type	positive cases
Microsatellie alterations	Nawroz *et al* (1996)	HNSCC	68/152 (45%)
	Gonzalez *et al* (1999)	SCLC	35/87 (40%)
	Taback *et al* (2001)	Breast	12/56 (21%)
	Sozzi *et al* (2001)	Lung	9/33 (27%)
	Silva *et al* (2002)	Breast	561/142 (42.9%)
	Andriani *et al*(2004)	Lung	19/40 (47.5%)
Hypermethylation	Esteller *et al* (1999)	NSCLC	11/22 (50%)
	Wong *et al* (1999)	Liver	13/22 (59%)
	Sanchez-Cespedes *et al* (2000)	HNSCC	21/50 (42%)
	Lee *et al* (2002)	Gastric	45/54 (83%)
	Usadel *et al*(2002)	Lung	42/89 (47%)
	Müller *et al* (2003)	Breast	11/62 (17.7%)
	Yamaguchi *et al* (2003)	Colorectal	3/14 (21%)
	Wong *et al* (2004)	Nasopharyngeal	29/41 (71%)
	Fiegl *et al* (2005)	Breast	
K-Ras mutations	Kopreski *et al* (2000)	Colorectal	27/70 (38%)
	Ryan *et al* (2003)	Colorectal	32/78 (41%)
	Camps *et al* (2005)	NSCLC	20/60 (30%)
APC mutations	Diehl *et al* (2005)	Colorectal	>60%
EBV DNA	Mutirangura *et al* (1998)	NPC	13/42 (31%)
	Lo *et al* (1999)	NPC	55/57 (96%)
	Gallagher *et al* (1999)	Hodgkins	30/33 (91%)
	Chan *et al* (2002)	NPC	155/170 (91%)

Table 1.4.1.1 Summary Table of serum/ plasma DNA cancer studies

1.4.1.2 Prenatal Diagnosis

For almost 40 years it has been known that foetal nucleated haematopoietic cells circulate within the maternal blood (Walknowska *et al* 1969). Conventional prenatal diagnostic procedures like chorionic villus sampling (CVS) or amniocentesis carry a risk to the unborn child so the discovery of foetal DNA amplifiable from maternal plasma and serum by Lo *et* al (1997) has given rise to the opportunity for non-invasive prenatal diagnosis and monitoring. Honda *et al* (2002) accurately detected with 100% sensitivity, Y-chromosomal signals in maternal plasma using real-time PCR in samples taken from women in their fifth week of pregnancy.

Prenatal diagnosis of certain neurological disorders, foetal chromosomal aneuploidies, sex-linked disorders and foetal rhesus D (RhD) status using foetal DNA from maternal plasma

has been shown to be possible. Non-invasive diagnosis of RHD genotype is now part of routine prenatal care in the United Kingdom, France, and the Netherlands (Bianchi *et al* 2005).

1.4.2 Serum/ Plasma RNA

1.4.2.1 Cancer testing

In 1999, Kopreski *et al* extracted and amplified mRNA from the serum of patients with malignant melanoma. They successfully detected tyrosinase mRNA in 4/6 patients samples with all 20 control samples remaining negative. The study also included some serum RNA stability tests whereby the effects of successive rounds of freeze-thawing aliquots of serum, the effect of leaving thawed serum at 4°C and also the usefulness of adding RNase inhibitors prior to freezing were assessed. Tyrosinase and c-abl mRNA were found to be detectable after the first round of freeze thawing, but a notable reduction was seen by the second round of freeze-thaw in at least one sample. By the third round, it was no longer possible to detect tyrosinase mRNA gene transcripts. The addition of RNasin prevented the degradation of RNA for one additional freeze-thaw only. With the serum specimens that were stored at 4°C once thawed, aliquots taken at 15 minute intervals showed it was possible to amplify RNA at only the 0 and 15min time samples.

Since the study reported by Kopreski *et al;* there have been a few studies relating to the extraction and amplification of mRNA from the plasma/ serum of cancer patients have been reported. Chen *et al* (2000) were able to amplify telomerase RNA from the serum of breast cancer patients that had small, well-differentiated tumours without either nodal or distant spread. Normal somatic cells have little or no detectable telomerase levels, while cancers have detectable telomerase activity in 85-100% of cases. They detected mRNA corresponding to the two subunits of human telomerase: telomerase RNA template (hTR)

and its catalytic component, telomerase reverse transcriptase protein (hTERT). hTR was found in 94% of the tumours and 28% of the corresponding serum. hTERT was also found in 94% of the tumours and in 25% of the corresponding tumours. All controls including two patients with benign disease had undetectable levels of both mRNAs.

Silva *et al* (2001) performed a study using the plasma of breast cancer patients. They chose mammaglobin and CK-19 mRNA as their target transcripts. Of the 45 patients, 27 were positive for mammaglobin and 22 were positive for CK-19. In the control samples 3/25 were positive for mammaglobin and 5/25 were positive for CK-19. The differences were statistically significant and presence of both transcripts in the plasma was associated with tumour size and proliferative index.

5T4 mRNA, which is the transcript that relates to a trophoblast glycoprotein frequently overexpressed in epithelial malignancies, was found to be expressed in 2/5 breast cancer sera and 6/14 lung cancer sera and 3/25 control sera (Kopreski *et al* 2001).

In 2001 Fleischhacker *et al* looked at the serum of lung cancer patients and screened them for detection of a number of mRNAs including CK- 19, MAGE- 2, TTF-1, PGP9.5, Her2/neu and HnRNPB1. CK-19 was found to be expressed in control samples (in this case all controls were positive). The combination testing of Her2/neu and HnRNPB1 were found to be the most useful in that one or both of these markers detected all lung tumours. TTF-1 and MAGE-2 were negative in all serum specimens while PGP9.5 was positive in only one serum sample.

Plasmsa β-catenin mRNA levels were measured in colorectal carcinoma and adenoma patients by Wong *et al* (2004) and compared to levels in the plasma of healthy controls. The gene transcript was found to have a median concentration (as determined by qPCR) to be 8737, 1218 and 291 in the colorectal carcinoma, colorectal adenoma and healthy subjects respectively. The plasma concentration of β-catenin mRNA was found to be

correlated with tumour stage. In addition to this, the concentration of β-catenin decreased significantly after tumour removal in 16/19 patients.

In 2005, Miura *et al* developed a quantitative detection method for detection of hTERT and AFP mRNA in the serum of hepatocellular carcinoma (HCC) patients. A total of 104 patients, were tested (64 with HCC, 20 with liver cirrhosis, and 20 with chronic hepatitis) were tested against 50 healthy individuals. Serum hTERT mRNA was detected at higher levels in patients with HCC over those with chronic liver diseases. Expression of hTERT mRNA was found to be correlated with tumour size, number, and degree of differentiation.

It has also been found possible to amplify virus-related RNA from the plasma of patients with virus- related cancer (Lo *et al* 1999). In this study the small EBV-encoded RNAs that are expressed in all NPC cases, and that are widely used for the detection of EBV-associated human tumours, were expressed in 23/26 (88.5%) plasma samples from NPC patients. All control samples were negative.

Wong *et al* (2006) carried out the first reported microarray study on RNA extracted from the serum of 10 oral cancer patients and 10 normal controls. They found that a higher number of probe sets were called Present in the cancer patient serum in comparison to the normal control serum specimens (2,623± 868 probe sets compared to 1,792± 165 probe sets). Five of ten cancer-related gene transcripts were validated by qPCR on serum specimens to be significantly differentially expressed between the two sample groups. They achieved a sensitivity of 91% and specificity of 71% in distinguishing oral squamous cell carcinoma.

A lot of the work to date regarding the testing of blood samples of cancer patients has related to detection of circulating malignant melanoma cells, the amount of circulating cancer cells has been shown to correlate with tumour burden, with lower rates seen in those with early disease. Extracellular DNA, however, has been shown to be detectable even in

the absence of circulating cancer cells (Kopreski *et al* 1997). The finding that extracellular mRNA is detectable in cancer patients (see Table 1.4.2.1. for summary) indicates the potential for a more sensitive, non-invasive means of diagnosing cancer patients at an early stage of disease. However, techniques to detect both cellular and extracellular mRNA are quite complementary and thought should be given to the analysis of both. Whilst analysis of whole blood specimens may yield conflicting results, separate analysis of tumour-related mRNA expression in circulating tumour cells and extracellular mRNA may aid in better definition of detection thresholds, sensitivities and specificities *etc.*

Marker (mRNA)	Reference	Cancer Type	No. of patients	positive cases	No. of controls	No. of controls +ve
Tyrosinase	Kopreski *et al* (1999)	Malignant melanoma	6	4 (67%)	20	0
EBV assoc.	Lo *et al* (1999)	Nasopharyngeal	26	23(88.5%)	29	6(20.7%)
hTR	Chen *et al* (2000)	Breast	18	5 (28%)	23	0
hTERT			16	4 (25%)	23	0
Mammaglobin	Silva *et al* (2000)	Breast	45	27 (60%)	25	3 (12%)
CK-19			45	22 (49%)	25	5 (20%)
CK- 19	Fleischhacker *et al* (2001)	Lung	18	15 (83%)	12	12 (100%)
HnRNP B1			18	14 (78%)	12	0
HER2/ neu			18	8 (44%)	12	0
TTF -1			18	0	12	0
PGP 9.5			18	1 (5%)	12	0
MAGE -2			18	0	12	0
5T4	Kopreski *et al* (2001)	Lung + Breast	19- (5 breast, 14 Lung)	2 (40%) 6 (43%)	25	3 (12%)
β-catenin	Miura *et al* (2005)	Colorectal	58 CRC 49 CRA	58 (100%) 49 (100%)	43	36 (84%)

Table 1.4.2.1 Summary Table of serum/ plasma RNA studies, CRC= colorectal carcinoma, CRA= colorectal adenoma, +ve= positive

1.5 AIMS OF THESIS

The aims of this thesis were to

i. Establish if it is possible to routinely extract and amplify extracellular mRNA from the conditioned media of cultured human cancer cells,

ii. Compare methods of extracting the mRNA to find a reliable and reproducible method.

iii. Investigate if any known tumour-related mRNAs are expressed and, if so, at what levels.

iv. Establish if it is possible to reliably and reproducibly amplify, label and hybridise CM cRNA to Affymetrix U133A Plus 2.0 array chips.

v. To compare CM obtained from various cell lines and their drug selected variants to find out if there are any changes in CM mRNA expression between the groups

vi. To study the mRNA expression profile of breast cancer patients using microarray technology by extracting mRNA from serum pre and post lumpectomy surgery and comparing it to mRNA extracted from the matched tumour and normal tissue.

vii. To study the nucleotide sequence of mRNAs detectable in serum/ CM to find motifs/ domains that may have an effect on the stability of the mRNA.

2 Materials & Methods

2.1 WATER

Ultrapure (UHP) water was used in the preparation of all media and solutions. This water was purified by a reverse osmosis system (Millipore Milli-RO 10 Plus, Elgastat UHP) to a standard of 12 - 18 MΩ/cm resistance.

2.1.1 DEPC treated water

All water used in RNA extractions/ RT-PCR reactions was either commercially available RNase- free water or, UHP water treated with Diethyl Pyrocarbonate (DEPC -Sigma D5758). DEPC inactivates RNases in solution at about 0.1%v/v ensuring that there were no endogenous RNases present in the water which would affect the experiments. 100µl of DEPC solution was added to 100ml of UHP. The solution was then stirred, left overnight at 37°C to allow the DEPC inactivate RNases. The next morning the solution was autoclaved for 20 minutes at 121°C to inactivate the chemical.

2.2 GLASSWARE

Solutions required for cell culture and maintenance were prepared and stored in sterile glass bottles. Bottles (and lids) and all other glassware used for any cell-related work was prepared as follows: - all glassware and lids were soaked in a 2% (v/v) solution of RBS-25 (AGB Scientific) for at least 1 hour. Following scrubbing and several rinses in tap water, the bottles were then washed by machine using Neodisher detergent, an organic, phosphate-based acid detergent. The bottles were then rinsed twice with distilled water, once with ultrapure water and were sterilised by autoclaving.

2.3 STERILISATION

Water, glassware and all thermostable solutions were sterilised by autoclaving at 121^{0}C for 20 minutes (min) under pressure of 1bar. Thermolabile solutions were filtered through a 0.22μm sterile filter (Millipore, millex-gv, SLGV-025BS). Low protein-binding filters were used for all protein-containing solutions.

2.4 MEDIA PREPARATION

Medium was routinely prepared and sterility checked by Joe Carey. The basal media used during routine cell culture were prepared according to the formulations shown in Table 2.4.1. 10x media were added to sterile ultrapure water, buffered with HEPES and $NaHCO_3$ and adjusted to a pH of 7.45 - 7.55 using sterile 1.5M NaOH and 1.5M HCl. The media were then filtered through sterile 0.22μm bell filters (Gelman, 121-58) and stored in 500ml sterile bottles at 4^{0}C. Sterility checks were carried out on each 500ml bottle of medium as described in Section 2.5.6.

The basal media were stored at 4^{0}C up to their expiry dates as specified on each individual 10x medium container. Prior to use, 100ml aliquots of basal media were supplemented according to requirements (see Table 2.4.1) and this was used as routine culture medium. This was stored for up to 2 weeks at 4^{0}C, after which time, fresh culture medium was prepared.

COMPONENTS	DMEM (Gibco, 12501-029)	Hams F12 (Gibco, 21700-109)	RPMI 1640 (Gibco)
10X Medium	500ml	Powder	Powder
Ultrapure H_20	4300ml	4700ml	4700ml
1M HEPES* Sigma , H-9136	100ml	100ml	100ml
7.5% $NaHCO_3$ BDH, 30151	45ml	45ml	45ml

Table 2.4.1 Preparation of basal media

* HEPES = N-(2-Hydroxyethyl) piperazine-N'-(2-ethanesulfonic acid)

2.5 CELL LINES

All cell culture work was carried out in a class II down-flow re-circulating laminar flow cabinet (Nuaire Biological Cabinet). Strict aseptic techniques were adhered to at all times. The laminar flow cabinet was swabbed with 70% industrial methylated spirits (IMS) before and after use, as were all items used in the cabinet. Each cell line was assigned specific media and waste bottles. Only one cell line was worked with at a time in the cabinet, which was allowed to clear for 15min between different cell lines. The cabinet itself was cleaned each week with industrial detergents (alternating between Virkon, Antec. International and TEGO, TH. Goldschmidt Ltd.), as were the incubators. The cell lines used during the course of this study, their sources and their basal media requirements are listed in Table 2.5.1. Cell Lines were maintained in 25cm^2 flasks (Costar; 3050), 75cm^2 (Costar; 3075) or 175cm^2 flasks (Corning, 431079) at 37^{0}C (with 5% CO_2 if required) and fed every two to three days.

Cell line	Basal medium (Plus additives)	Cell type	Source
DLKP	ATCC (HAMS F12 + DMEM) (5%FCS)	Human poorly-differentiated lung carcinoma	Dr Geraldine Grant, NCTCC
RPMI 2650 MI (RPMI TAX)	MEM (5%FCS, 2mM L-glut, 1% Sodium Pyruvate, 1% NEAA, 5% CO_2)	Nasal septum carcinoma, melphalan selected variant	Dr Yizheng Liang, NICB
RPMI 2650 TAX (RPMI TAX)	MEM (5%FCS, 2mM L-glut, 1% Sodium Pyruvate, 1% NEAA, 5% CO_2)	Nasal septum carcinoma, taxol selected variant	Dr Yizheng Liang, NICB
MCF-7	MEM (5%FCS, 2mM L-glut, 1% Sodium Pyruvate, 5% CO_2)	Breast carcinoma	ATCC
MDA-F (MDA-MB-435S-F)	RPMI 1640 (10%FCS, 2mM L-glut, 5% CO_2)	Melanoma (formerly described as breast ductal carcinoma)	Dr Sharon Glynn, NICB
MDA-F-ADR-SI (MDA-MB-435-F/ADR-10p10p SI)	RPMI 1640 (10%FCS, 2mM L-glut,5% CO_2)	Melanoma (formerly described as breast ductal carcinoma) adriamycin selected variant	Dr Sharon Glynn, NICB
MDA-F-TAX-SI (MDA-MB-435-F/Taxol-10p4p SI)	RPMI 1640 (10%FCS, 2mM L-glut,5% CO_2)	Melanoma (formerly described as breast ductal carcinoma) taxol selected variant	Dr Sharon Glynn, NICB
BT474	DMEM (10%FCS, 1% Sodium Pyruvate)	Breast ductal carcinoma	ATCC
NCI H1299 (H1299)	RPMI 1640 (5% FCS 1% Sodium Pyruvate)	Large cell lung carcinoma cell line	ATCC
NCI H1299 TAX (H1299 TAX)	RPMI 1640 (5% FCS 1% Sodium Pyruvate)	Large cell lung carcinoma cell line, taxol selected variant.	Dr Laura Breen, NICB
HMEC-1001-10 (HMEC)	MEGM Bullet kit	Breast mammary cells	Biowhittaker

Table 2.5.1 Cell lines used during the course of this study

2.5.1 Subculture of Adherent Lines

During routine sub-culturing or harvesting of adherent lines, cells were removed from their flasks by enzymatic detachment.

To achieve this, waste medium was removed from the flasks and flasks were rinsed with a pre-warmed (37^0C) trypsin/EDTA (TV) solution (0.25% trypsin (Gibco, 25090-028), 0.01% EDTA (Sigma, EDS) solution in PBS (Oxoid, BR14a)). The purpose of this was to remove any naturally occurring trypsin inhibitor which may be present in residual serum. Fresh TV was then placed on the cells (2.5ml/75cm^2 flask) and the flasks incubated at 37^0C until the cells were seen to have detached. The trypsin was deactivated by addition of an equal volume of growth medium (*i.e.* containing serum). The entire solution was transferred to a 30ml sterile universal tube (Sterilin; 128a) and centrifuged at 900 rpm for 5 min. The resulting cell pellet was resuspended in pre-warmed (37^0C) fresh growth medium; cells were counted (Section 2.5.2) and used to re-seed a flask at the required cell density or to set up an assay.

2.5.2 Cell Counting

Cell counting and viability determinations were carried out using a trypan blue (Gibco, 15250-012) dye exclusion technique.

An aliquot of trypan blue was added to a sample from a single cell suspension in a ratio of 1:5. After 3 min incubation at room temperature, a sample of this mixture was applied to the chamber of a haemocytometer over which a glass coverslip had been placed. Cells in the 16 squares of the four outer corner grids of the chamber were counted microscopically, an average count per corner grid was calculated with the dilution factor being taken into account, and final cell numbers were multiplied by 10^4 to determine the number of cells per ml. The volume occupied by the chamber is 0.1cm x 0.1cm x 0.01cm *i.e.* 0.0001cm^3.

Therefore cell number x 10^4 is equivalent to cells per ml. Non-viable cells were those that stained blue while viable cells excluded the trypan blue dye and appeared unstained.

2.5.3 Cell Freezing

To allow long-term storage of cell stocks, cells were frozen and cryo-preserved in liquid nitrogen at temperatures below -180^{0}C. Once frozen properly, such stocks should last indefinitely.

Cells to be frozen were harvested in the log phase of growth (*i.e.* actively growing and approximately 50 - 70% confluent) and counted as described in Sections 2.5.2. Pelleted cells were re-suspended in serum. An equal volume of a DMSO/serum (1:9, v/v) was slowly added drop wise to the cell suspension to give a final concentration of at least $5x10^6$ cells/ml. The careful performance of this step was very important, as DMSO is toxic to cells. When added slowly, the cells had a period of time to adapt to the presence of the DMSO, otherwise cells may have lysed. The suspension was then aliquoted into labelled cryovials (Greiner, 122 278) which were then quickly placed in the vapour phase of a liquid nitrogen container (approximately -80^{0}C). After 2.5 to 3.5 hours, the cryovials were lowered down into the liquid nitrogen, where they were stored until required.

2.5.4 Cell Thawing

Prior to the removal of a cryovial from the liquid nitrogen stores for thawing, a sterile universal tube containing growth medium (5ml) was prepared and placed at 37°C for approximately 30mins prior to use. The rapid thawing and subsequent dilution of thawed cells is required to reduce their exposure time to the DMSO freezing solution, which is toxic at room temperature. The cell suspension was centrifuged at 900rpm for 5 min, the DMSO-containing supernatant removed and the cell pellet re-suspended in fresh growth medium. Thawed cells were placed into tissue culture flasks with the appropriate volume of

medium (5ml/25cm^2 flask, 8-10ml/75cm^2 flask and 15-18ml/175cm^2 flask) and allowed to attach overnight. After 24 hours, the cells were re-fed with fresh medium to remove any residual traces of DMSO and non-viable cells.

2.5.5 Sterility Checks

Sterility checks were routinely carried out on all media, supplements and trypsin used for cell culture. Samples of basal media were inoculated into Columbia (Oxoid, CM331) blood agar plates, Sabauraud (Oxoid, CM217) dextrose and Thioglycollate (Oxoid, CM173) broths which detect most contaminants including bacteria, fungus and yeast. Growth media (*i.e.* supplementedas shown in Table 2.5.1) were sterility checked at least 2 days prior to use by incubating samples at 37^{0}C. These were subsequently examined for turbidity and other indications of contamination.

2.6 MYCOPLASMA ANALYSIS

Mycoplasma examinations were carried out routinely (at least every 3 months) on all cell lines used in this study. These analyses were performed by Aine Adams and Michael Henry.

2.6.1 Indirect Staining Procedure

In this procedure, *Mycoplasma*-negative NRK cells (a normal rat kidney fibroblast cell line) were used as indicator cells. These cells were incubated with supernatant from test cell lines and then examined for *Mycoplasma* contamination. NRK cells were used for this procedure because cell integrity is well maintained during fixation. A fluorescent Hoechst stain was utilised which binds specifically to DNA and so will stain the nucleus of the cell in addition to any *Mycoplasma* DNA present. A *Mycoplasma* infection would thus be seen as small fluorescent bodies in the cytoplasm and sometimes the extracellular region of the NRK cells.

NRK cells were seeded onto sterile coverslips in sterile Petri dishes at a cell density of $2x10^3$ cells per ml and allowed to attach over night at 37^0C in a 5% CO_2 humidified incubator. 1ml of cell-free (cleared by centrifugation at 1,000 rpm for 5 min) supernatant from each test cell line was then inoculated onto a NRK Petri dish and incubated as before until the cells reached 20 - 50% confluency (4 - 5 days). After this time, the waste medium was removed from the Petri dishes, the coverslips washed twice with sterile PBS, once with a cold PBS/Carnoys (50/50) solution and fixed with 2ml of Carnoys solution (acetic acid:methanol-1:3) for 10 min. The fixative was then removed and after air-drying, the coverslips were washed twice in deionised water and stained with 2ml of Hoechst 33258 stain (BDH) (50ng/ml) for 10 min.

From this point on, work was carried out in the dark to limit quenching of the fluorescent stain. The coverslips were rinsed three times in PBS. They were then mounted (using 50% (v/v) glycerol in 0.05M citric acid and 0.1M disodium phosphate) and examined by fluorescence microscope under a UV filter.

2.6.2 Direct Staining

The direct stain for *Mycoplasma* involved a culture method where test samples were inoculated onto an enriched *Mycoplasma* culture broth (Oxoid, CM403) - supplemented with 16% serum, 0.002% DNA (BDH; 42026), 2mg/ml fungizone (Gibco,15290-026), $2x10^3$ units penicillin (Sigma, Pen-3) and 10ml of a 25% (w/v) yeast extract solution - to optimise growth of any contaminants. These were incubated at 37^0C for 48 hours. Samples of this broth were then streaked onto plates of *Mycoplasma* agar base (Oxoid, CM401) which had also been supplemented as above and the plates were incubated for 3 weeks at 37^0C in a CO_2 environment. The plates were viewed microscopically at least every 7 days; the appearance of small, "fried egg" -shaped colonies is indicates *Mycoplasma* infection.

2.7 Cell growth and Conditioned Media (CM)/ serum collection.

2.7.1 Identifying a suitable time point for CM collection

The experimental design to identify a suitable time point for CM collection and analysis is shown in Fig. 2.7.1, RPMI ML cells were set up in T75cm^2 flasks at a density of 2 x 10^{6}cells and 8mls of media was added. After each 24 hours, 1ml of conditioned media (CM) was taken off the cells, filtered with a 0.22µm filter and immediately added in 250µl aliquots to 750µl aliquots of Tri Reagent and stored at -80°C until RNA extraction (see Section 2.8.2.2) was carried out.

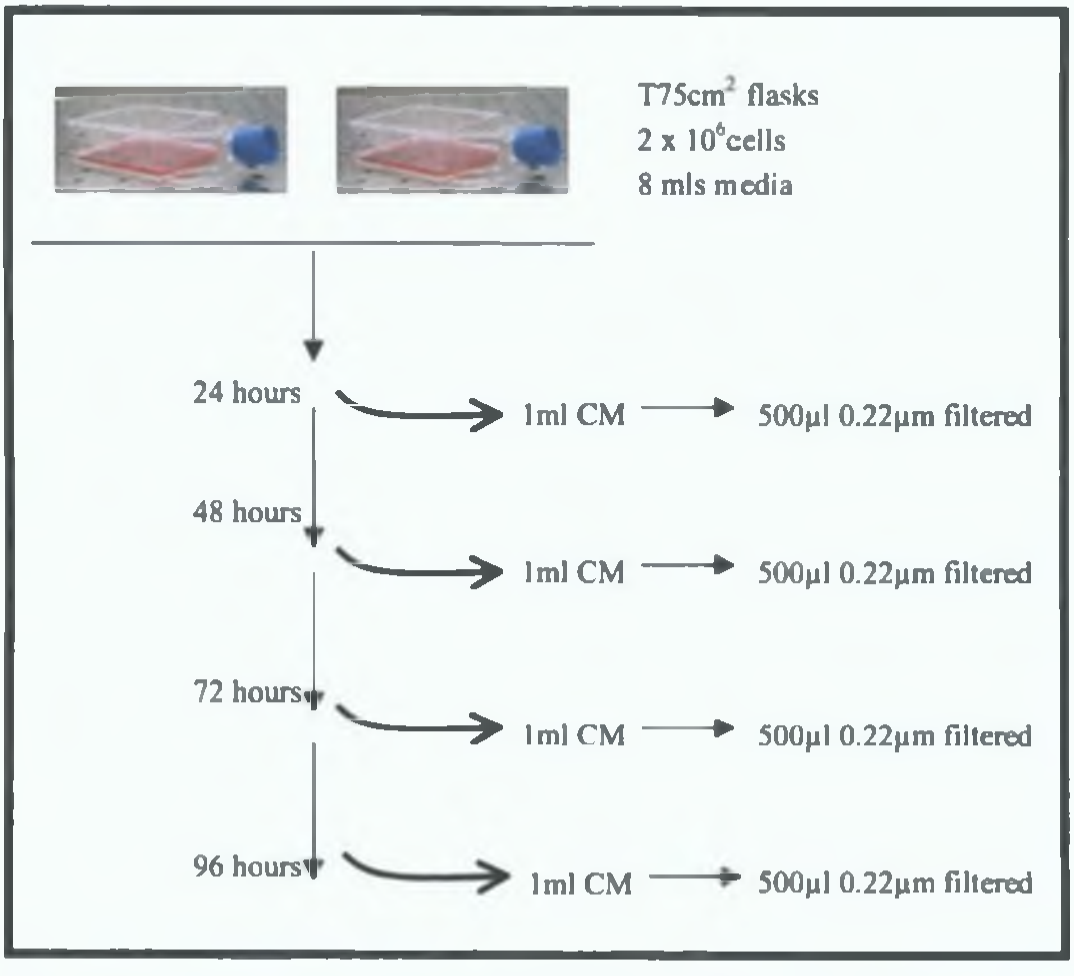


Figure 2.7.1 Experimental design of study to identify a suitable time point post cell seeding for CM collection and analysis

2.7.2 Study to find a reproducible minimum detection level

In this study, cells were set up as in the previous study, i.e. 2 x 10^6 cells/ flask. The cells were cultured for 48 hours, after which CM was taken and filtered through a 0.22µm filter. For RNA extraction using the Promega SV Total Isolation kit, 100µl of the CM was immediately added to the lysis buffer and processed according to the manufacturer's protocol (see Section 2.8.2.1). For samples whereby RNA was to be extracted using Tri Reagent, either 250µl of the filtered CM was immediately added to 750µl of Tri Reagent, or 100µl of filtered CM was added to 150µl of water and then added to 750µl of Tri Reagent, and either processed immediately according to Section 2.8.2.2 or stored at -80°c until required (see Fig 2.7.2 for overview).

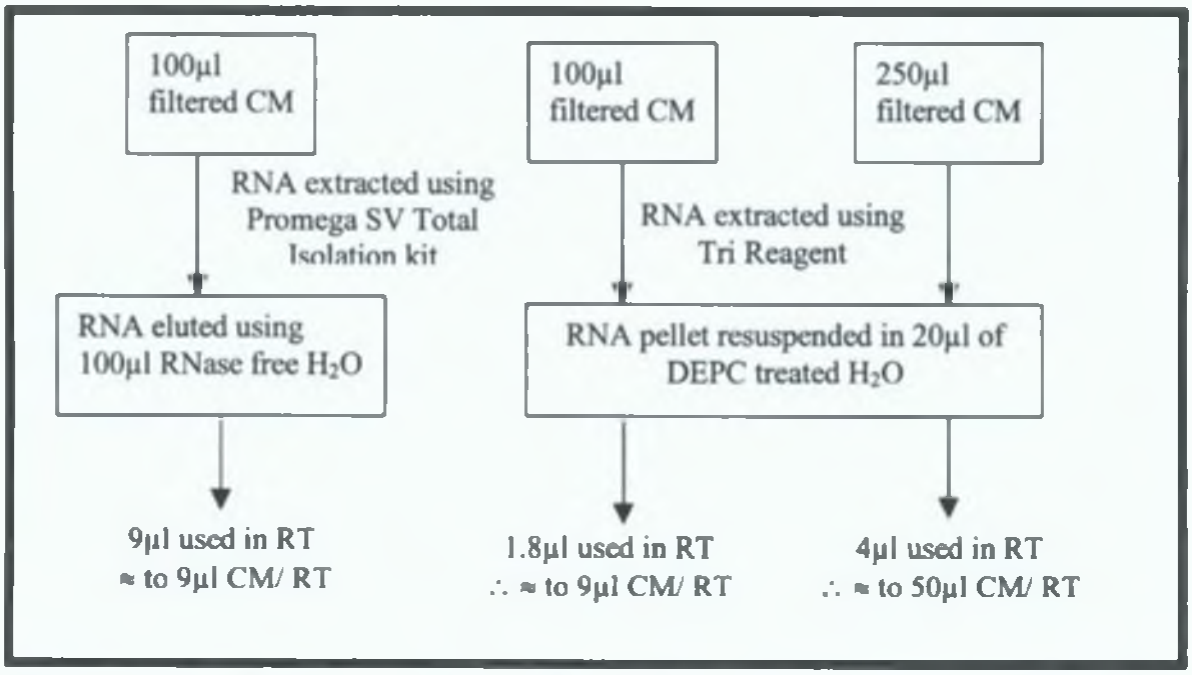


Figure 2.7.2.1 Experimental design of study to finddetermine conditions fro reproducible detection of mRNA in CM, RT= reverse transcription.

2.7.3 Analysis of transcripts in RNA isolated from cell lines and CM

Flasks of cells for this study ($75cm^2$) were set up at a cell density of 2×10^6 cells/ flask. 8mls of media (see Table 2.5.1 for details of cell lines and media) was added to each flask. After 48 hours CM was taken off the cells and split into two aliquots. One aliquot was filtered through a 0.22µm filter and the other aliquot was filtered through a 0.45µm filter. 250µl aliquots of filtered CM were immediately added to 750µl of Tri Reagent and stored at -80°C until required for RNA exraction.

2.7.4 Concentration protocol 1

DLKP cells were seeded at a density of 4×10^6 cells per $175cm^2$flask with two flasks set up, and 15 mls of media was added to each flask. After 48 hours, the conditioned media (CM) from the each pair of flasks was taken off and pooled. Half of the CM was filtered through a 0.22µm filter and the other half through a 0.45µm filter. A 250µl aliquot from each filtered sample was immediately mixed with 750µl aliquots of Tri Reagent (0.22B/ 0.45B) and placed in the -80°C freezer for RNA extraction at a later time using the protocol described in Section 2.7.2.2.

The remainder of the 15mls of the 0.45µm filtered or 0.22µm filtered CM was placed in Ultrafree-15 millipore concentrators. The samples were then centrifuged for 15 mins at 500rcf, 15 mins at 2000rcf and 15 mins at 3220rcf at which time the samples had been concentrated down to approximately 300µl (0.22µm filtered sample) and 500µl (0.24µm filtered sample). The centrifugal speed was increased to help aid with concentration of the CM sample. 250µl aliquots of each concentrated sample were immediately mixed with 750µl aliquots of Tri Reagent (0.22C/ 0.45C) and stored at -80°C for RNA extraction at a later time using the protocol described in Section 2.7.2.2.

Finally 250µl aliquots from each of the "flow- through" CM were also mixed with 750µl aliquots of Tri Reagent (0.22FT/ 0.45FT) for RNA extraction at a later time using the protocol described in Section 2.7.2.2 (see Fig 2.7.4.1 for experimental design of protocol 1)

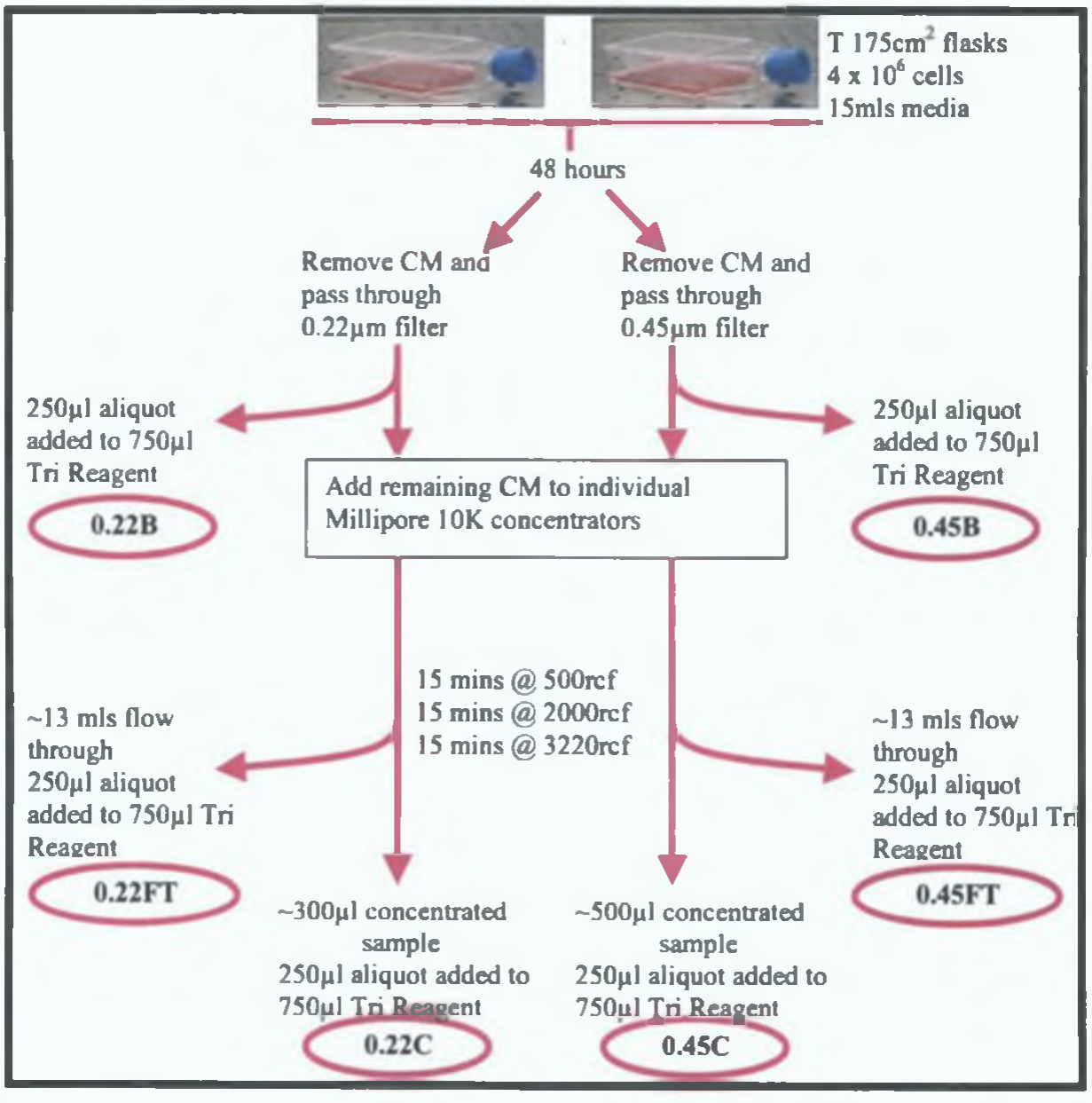


Figure 2.7.4.1 Experimental design of concentration protocol 1. 0.22/ 0.45 refer to the filter size the CM samples were filtered through. The suffixes B, C and FT refer to samples taken before concentration (B), taken after concentration (C), or taken from the CM that flowed- through the concentrator column (FT).

2.7.5 Concentration protocol 2

To establish if it was possible to further concentrate the samples obtained in the first concentration protocol (see Section 2.7.4), the procedure was repeated using 0.45µm filtered samples only, and at every sampling point, i.e. before concentration, concentrated and flow through. Four 250µl aliquots were taken at each of these sampling points (see Fig. 2.7.5.1 for experimental design). The centrifuge speed was also lowered to 500rcf until the sample was concentrated to approximately 1ml. The four aliquots of 250µl for each sample underwent the RNA extraction procedure as before (see Section 2.8.2.2) with the exception that the four RNA pellets for each sample were resuspended into one 15µl volume of DEPC treated water.

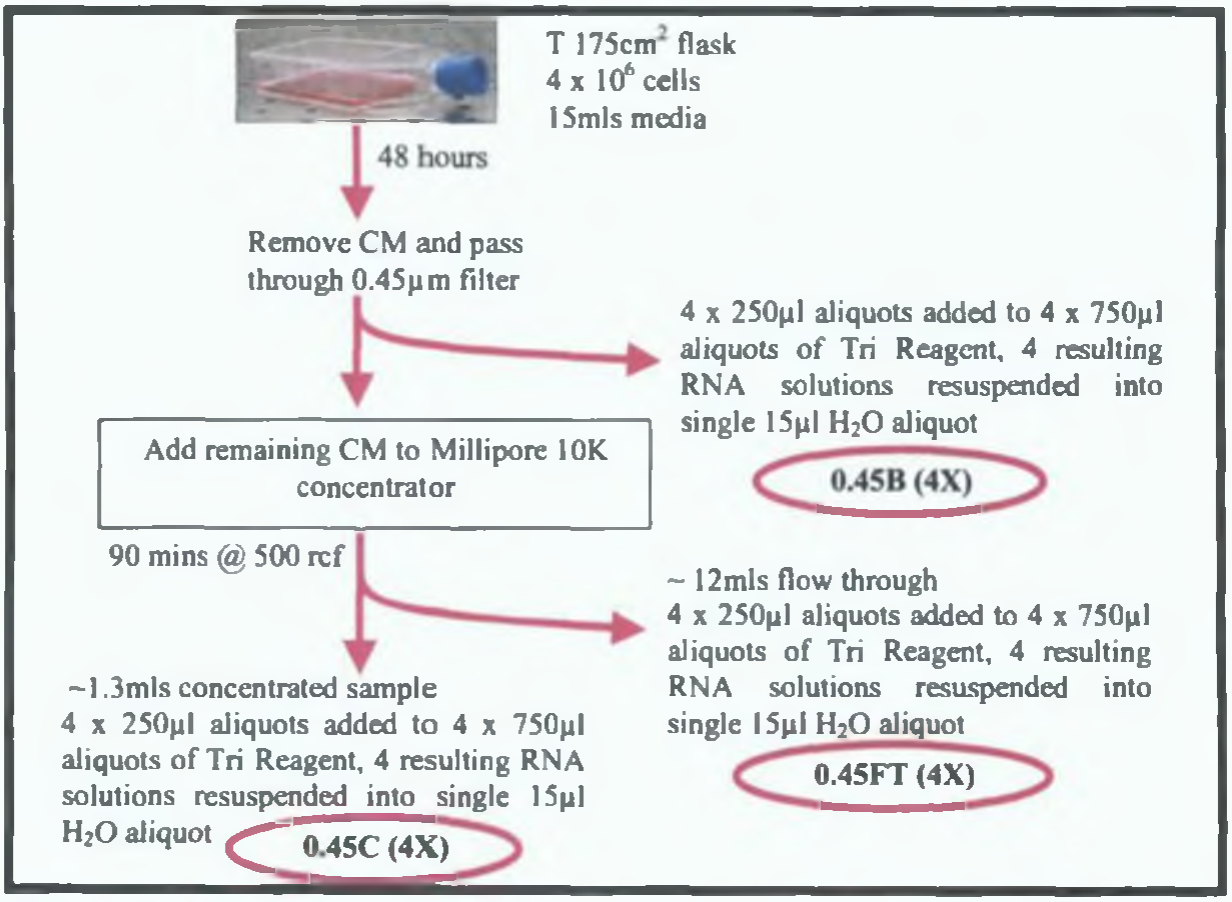


Figure 2.7.5.1 Experimental design for concentration protocol 2 0.45 refers to the filter size the CM samples were passed through. The suffixes B, C and FT refer to samples taken before concentration (B), taken after concentration (C), or taken from

the CM that flowed- through the concentrator column (FT). (4X) refers to the four individual RNA pellets resuspended into one RNA suspension of 15µl.

2.7.6 Concentration protocol 3

To investigate whether the addition of RNasin and/ or a decrease in centrifuge time would increase the integrity of the RNA yield, the procedure outlined in Section 2.7.5 was repeated again with two CM samples. Into one 3750 Units (~100µl) of RNase inhibitor (Sigma) was added and nothing was added into the second. The centrifuge speed was also increased to 3220rcf in order to shorten the time required for concentration (see Fig 2.7.6.1 for experimental design)

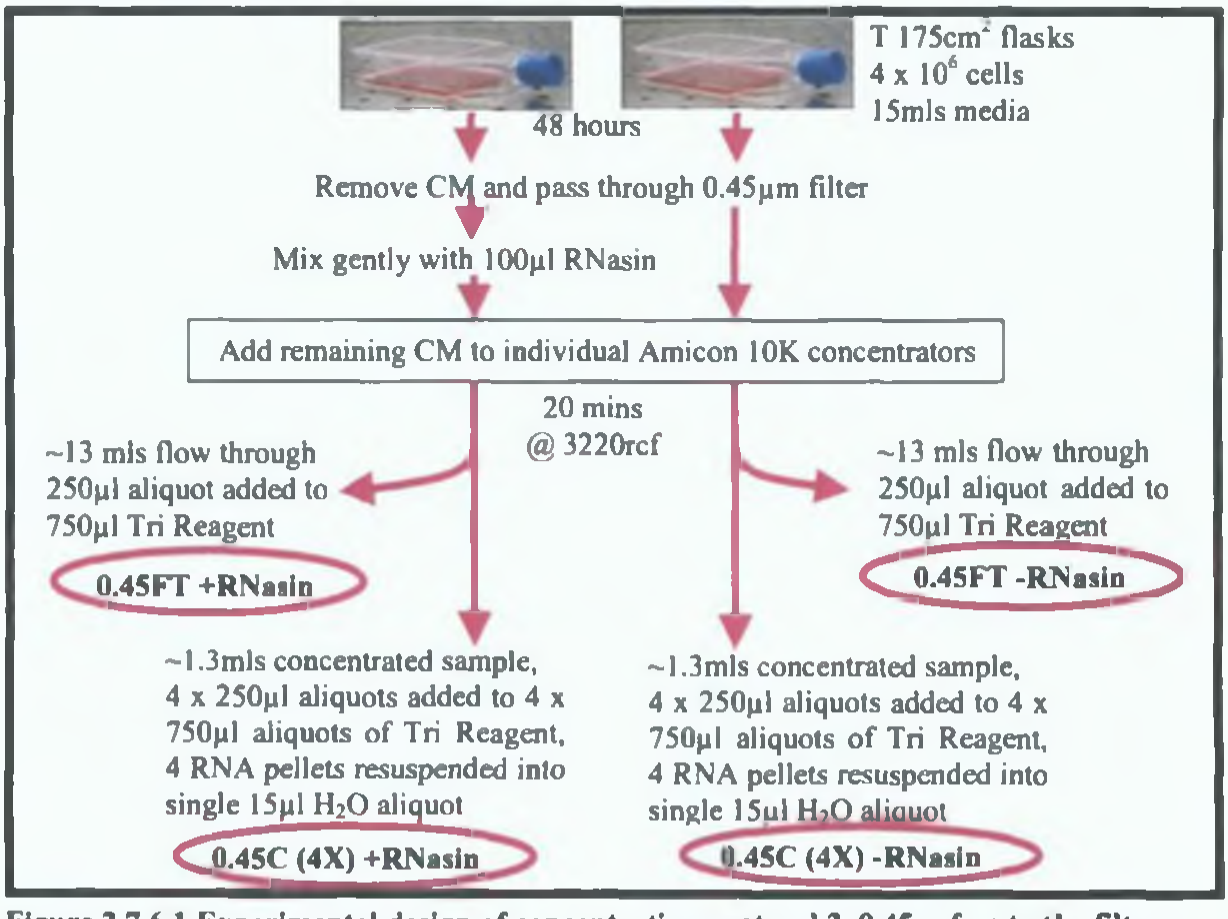


Figure 2.7.6.1 Experimental design of concentration protocol 3. 0.45 refers to the filter size that the CM samples were passed through. The suffixes B, C and FT refer to samples taken before concentration (B), taken after concentration (C), or taken from the CM that flowed- through the concentrator column (FT). (4X) refers to the fact

that four individual RNA pellets resulting from RNA extraction of 4x 250µl CM samples were resuspended into one RNA suspension of 15µl. +RNasin refers to the CM sample where RNasin was added to the CM after filtration and before concentration, compared to the CM sample were no RNasin was added (-RNasin).

2.7.7 Pilot DLKP microarray study

To investigate if extracellular RNA was suitable for global analysis of gene transcripts, a pilot study using RNA isolated from the CM of one cell line, DLKP was performed.

DLKP cells at passage 54 were seeded in 75cm^2 flasks at a density of 2 x 10^6 cells/ flask and 8mls of media was added. After 48 hours, CM was removed from the flask, filtered through a 0.45µm filter and six 250µl aliquots of CM were added to six 750µl aliquots of Tri Reagent respectively. RNA was extracted from all six aliquots individually (see Section 2.8.2.2) and the RNA from one of these samples (DLKP 1) was selected to undergo the Affymetrix two- cycle amplification and biotinylation protocol (see Section 2.12). The same procedure was carried out for DLKP cells set up in duplicate at passage 58. RNA was extracted from both of these samples (see Section 2.8.2.2) and one of the six aliquots of extracted RNA for both samples (assigned DLKP 2 and DLKP 3) were selected to undergo the Affymetrix two-cycle amplification and biotinylation protocol (see Section 2.12).

2.7.8 Analysis Of mRNA Expressed By Various CM Using Microarray

All cells for this study were set up in 175cm^2 flasks so that they were no more than 80% confluent after 48 hours in culture and fed with 18ml of media (for this study, the cells were seeded and maintained by other researchers, Dr Laura Breen, Isabella Bray and Helena Joyce). The flasks were set up in triplicate. After 48 hours in culture the conditioned media was removed and half was passed through a 0.22µm filter and the other half through a 0.45µm filter (only the CM passed through the 0.45µm filter was used in this study).

Within 10- 20 minutes, the filtered CM was added in 250µl aliquots to 750µl aliquots of Tri Reagent respectively and stored at -80°C until required.

2.7.9 Culture of normal breast cells

The normal breast cells, HMEC 1001-10 (HMEC) were cultured in a 25cm^2 flask until they were approximately 60% confluent. The media (see Table 2.5.1 for details) was then replaced with fresh media and the cells left in culture for a further 48 hours to reach an approximate confluency of 80%. After 48 hours, the CM was removed, filtered through a 0.45µm filter and 250µl aliquots were added to aliquots of 750µl Tri Reagent, respectively and the samples were stored at -80°C until required for RNA extraction, which was carried out using the protocol described in Section 2.8.2.2. The cells were lysed in 1ml of Tri Reagent and processed according to the manufacturer's protocol.

2.7.10 Microarray analysis of Breast cancer serum/ tissue samples

From each patient recruited with consent for the Breast cancer microarray study, 2 serum specimens and 2 tissue specimens were taken. On the morning of surgery to remove the breast tumour, approximately 10ml of blood was collected. These specimens (termed Pre-serum) were processed within four hours of blood draw. Tumour and matched normal tissue specimens were also collected at time of surgery and stored at -80°C until required. Approximately 2- 4 months post surgery, a second blood specimen was collected (termed Post- serum), and was processed within 4 hours of collection. Serum specimens from six normal volunteers were also collected and processed within four hours of collection.

All serum specimens were processed by removing the serum from the clotted blood and placing into a 10ml centrifuge tube. A balance tube was made by adding an equal amount

of water to another 10ml centrifuge tube. The tubes were then centrifuged at room temperature for 15 minutes at 500rcf. After centrifugation the cleared serum was carefully removed and passed through a 0.45µm filter to further ensure no particles/ platelets were retained. The serum was then aliquoted into 500µl/ 1ml aliquots and stored at -80°C until required. RNA extraction of the serum was carried out using a slightly modified version of the optimised Tri Reagent protocol developed here (see Section 2.8.2.2). In order to minimise the amount of time the serum was defrosting without protection from RNases, when required, 500µl of Tri Reagent was added to the frozen serum specimens and the samples were gently pipetted to ensure thorough mixing. This serum-Tri Reagent mix was then added in 500µl aliquots to clean eppendorf tubes containing 250µl of Tri Reagent, so that the final ratio of Tri Reagent to serum was 3:1 (i.e.750µl : 250µl).

Tumour and matched normal tissue specimens were cut into small sections and manually cut up in a petri dish containing 1ml of Tri Reagent. RNA was then extracted from these tissue specimens according to the manufacturer's protocol.

2.8 RNA EXTRACTION

2.8.1 Cells

Approximately 1 x 10^7 cells were trypsinised, washed once with PBS, pelleted and lysed using 1ml of Tri Reagent™ (Sigma, T-9424). The samples were allowed to stand for 5 min at room temperature to allow complete dissociation of nucleoprotein complexes and then stored at –80^{0}C. When required, RNA was extracted according to the manufacturer's protocol.

2.8.2 Conditioned Media

2.8.2.1 Promega SV Total RNA Isolation System

Chen *et al* (2000) used the Promega SV Total isolation system to extract RNA from 100µl of serum specimens and perform RT-PCR on the extracted RNA, so this kit was tested using CM samples.

CM was taken from flasks as required and filtered through a 0.22µm filter. 100µl of CM was added to 175µl of SV RNA lysis buffer that had β- mercaptoethanol added, as recommended by the manufacturer. The solutions were then mixed by inversion. To this, 350µl of SV RNA dilution buffer was added and again mixed by inversion and incubated at 70°C for 3 minutes. After the incubation, the samples were centrifuged for 10 minutes and to the cleared lysate (in this case there was little/no sediment) 200µl of 95% ethanol was added and the solution was mixed by pipetting. The mixture was then transferred to a spin basket supplied in the kit and centrifuged for 1 minute. After discarding the eluant, 600µl of SV RNA wash solution was added. Again the tubes were centrifuged for 1 minute and the eluant discarded. To destroy any possible DNA contamination, 50µl of DNase mix was applied to the membrane and incubated for 5 minutes at RT, after which 200µl of SV DNase stop solution was added and the tubes centrifuged for 1 minute. SV RNA wash solution (600µl) was added and the tubes were centrifuged for 1 minute. The flow through was discarded. 250µl of SV RNA wash solution was added to the spin baskets and the samples were centrifuged for 2 minutes. The spin baskets were then transferred to elution tubes and 100µl of nuclease-free water was added to the membrane. The RNA was eluted by centrifuging the tubes for 1 minute (all centrifugation steps carried out >13,000rpm).

2.8.2.2 Tri Reagent

CM was taken from flasks as required and filtered through either a 0.22/ 0.45µm filter. CM,

in volumes of 250µl (or 100µl of CM and 150µl H_2O) were added to 750µl of Tri Reagent™. As recommended when using Tri Reagent, the samples were allowed to stand for 5 min at room temperature to allow complete dissociation of nucleoprotein complexes and then stored at –80°C. When required, the samples were thawed and allowed to stand for 5 min before 0.2ml of chloroform was added per 750µl of Tri Reagent™ used. Samples were then shaken vigorously for 15 sec and allowed to stand for 15 min at room temperature, after which they were centrifuged at 13000rpm in a microfuge for 15 min at 4°C. This step separated the mixture into 3 phases; the RNA was contained in the colourless upper aqueous layer. This layer was then transferred to a fresh Eppendorf tube, to which glycogen (final concentration 120 µg/ml) and 0.5ml of ice-cold isopropanol were added. The samples were mixed and incubated at room temperature for 5-10 minutes and stored at –20°C over-night, to ensure maximum RNA precipitation. Eppendorf tubes were then allowed to stand at room temperature for 10 mins before centrifugation at 12,000rpm for 30 minutes and 4°C to pellet the precipitated RNA. Taking care not to disturb the RNA pellet, supernatant was removed and the RNA pellet was subsequently washed by the addition of 750 µl of 75% ethanol and vortexing. Following centrifugation at 7- 10,000 rpm for 5 minutes at 4°C, the supernatant was removed (this wash step was repeated). The RNA pellet was then allowed to air-dry for 5-10 minutes and subsequently was re-suspended in the required amount of RNase- free or DEPC-treated H_20. To facilitate dissolution, samples were mixed by repeated pipetting and incubated at 55-60°C for 10-15 minutes.

Volumes of water used to resuspend the RNA pellets varied depending on the study involved and expected yield of RNA.

- RNA used for RT-PCR study in Section 2.7.3 was resuspended in 20µl of DEPC treated water.
- RNA used in concentration of biological samples in Section 2.7.4- 2.7.6 was

resuspended in 15µl DEPC treated water.

- RNA used for CM microarray studies in Sections 2.7.7- 2.7.9 was resuspended into 6µl of DEPC treated water.

2.9 Quantification/ visualisation of RNA

2.9.1 The NanoDrop® ND-1000 Specrophotometer (Labtech International).

The NanoDrop® ND-1000 is a full-spectrum (220-750nm) spectrophotometer that measures 1 µl samples with high accuracy and reproducibility. It uses a sample retention technology that relies on surface tension alone to hold the sample in place eliminating the need for cuvettes and other sample containment devices. In addition, the NanoDrop has the capability to measure highly concentrated samples without dilution (50X higher concentration than the samples measured by a standard cuvette spectrophotometer).

To quantify an RNA sample, 1 µl of the sample is pipetted onto the end of a fiber optic cable (the receiving fiber, see Fig. 2.9.1.1 (a)). A second fiber optic cable (the source fiber, see Fig. 2.9.1.1(b)) is then brought into contact with the liquid sample causing the liquid to bridge the gap between the fiber optic ends. The gap is controlled to a 1mm path (see Fig.2.9.1.1 (c)). A pulsed xenon flash lamp provides the light source and a spectrometer utilising a linear CCD array is used to analyse the light after passing through the sample. The instrument is controlled by special software run from a computer, and the data is logged in an archive file on the computer (see Fig. 2.9.1.2 for screenshot).

When measurement of the sample is complete, the sample can be simply wiped away using a soft laboratory wipe. This is sufficient to prevent sample carryover because each measurement pedestal is a highly polished end of a fiber optic cable, with no cracks or crevices for residual sample to reside in.

Figure 2.9.1.1 Samples are quantified by loading 1µl onto the receiving fibre (A), the source fiber, connected to the sampling arm (B) is brought down into contact with the sample allowing a 1mm gap between the upper and lower pedestal (C), through which the light is passed. (Pictures adapted from ND-1000 Spectrophotometer users manual V 3.1.0).

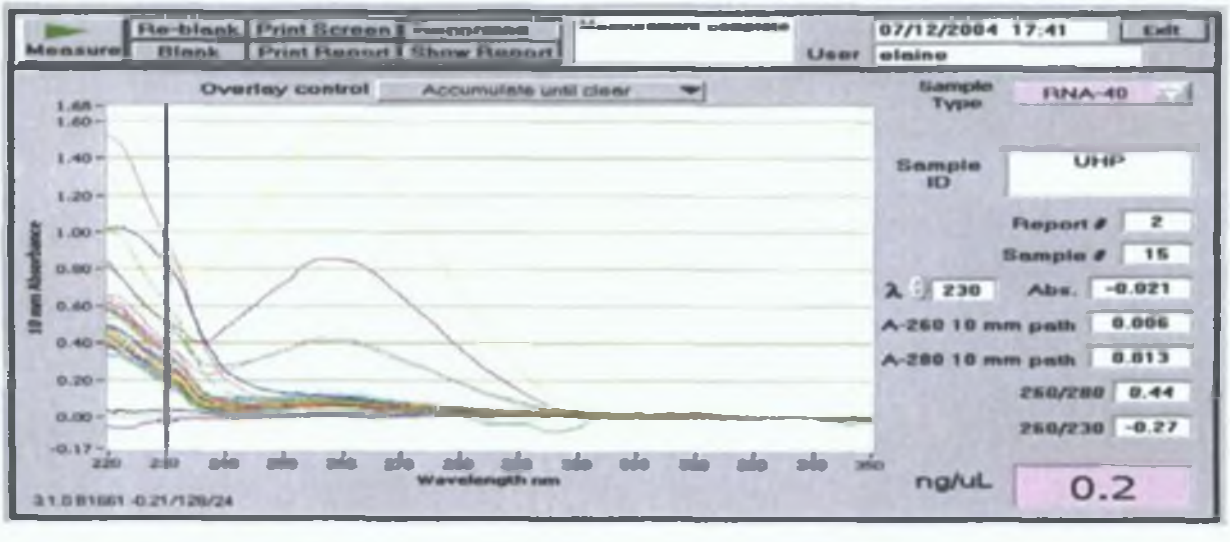


Figure 2.9.1.2 Example screen shot of Nanodrop readouts, wavelength is plotted on the X-axis and absorbance is plotted on the Y-axis, RNA absorbs at 260nm. Coloured lines correspond to measurements obtained for different RNA samples

2.9.2 Agilent 2100 bioanalyser

The Agilent 2100 Bioanalyser is a microfluidics-based platform for the analysis of proteins, DNA and RNA. The miniature chips are made from glass and contain a network of interconnected channels and reservoirs. The RNA 6000 Nano LabChip kit enables analysis of samples containing as little as 5ng of total RNA. The channels are firstly filled with a gel

matrix and the sample wells with buffer or sample, there are 12 sample wells per chip. 1µl of each sample is loaded into a sample well along with a fluorescent dye (marker). An RNA ladder is loaded into another sample well for size comparison. When all the samples are loaded, the chip is briefly vortexed and loaded onto the bioanalyser machine (see Fig. 2.9.2.1(A) for picture of chip). The machine is fully automated and electrophoretically separates the samples by injecting the individual samples contained in the sample wells into a separation chamber (see Fig. 2.9.1 (B)).

Figure 2.9.2.1 The RNA 6000 Nano chip, a picture of the front of the RNA Nano chip (A) and a diagram of the microchannels in the chip (B), the sample moves through microchannels (1) and is injected into separation chamber (2) were components are electrophoretically separated (3) and detected by their fluorescence and translated into gel-like images and electropherograms. Both pictures available from http://www.agilent.com/chem/labonachip.

The resulting data is presented as an electropherogram (See Fig 2.9.2 (a) and (b)). The fluorescensce is measured on the Y-axis and the time in seconds is measured on the X-axis. The smaller fragments are detected first and are shown on the left-hand side of the electrophoreogram. Fig. 2.9.2.2 (a) is an example of good quality cell RNA, the 18S and 28S ribosomal RNA peaks are quite sharp and the 28S is higher than the 18S peak. As RNA degrades, the 28S RNA peak decreases and smaller fragments are visible. An

example of slightly degraded cell RNA is shown in Fig. 2.9.2.2 (b). The 28S peak is smaller than the 18S peak and an extra peak is visible at approximately 25 seconds. This peak represents the fragmented RNA.

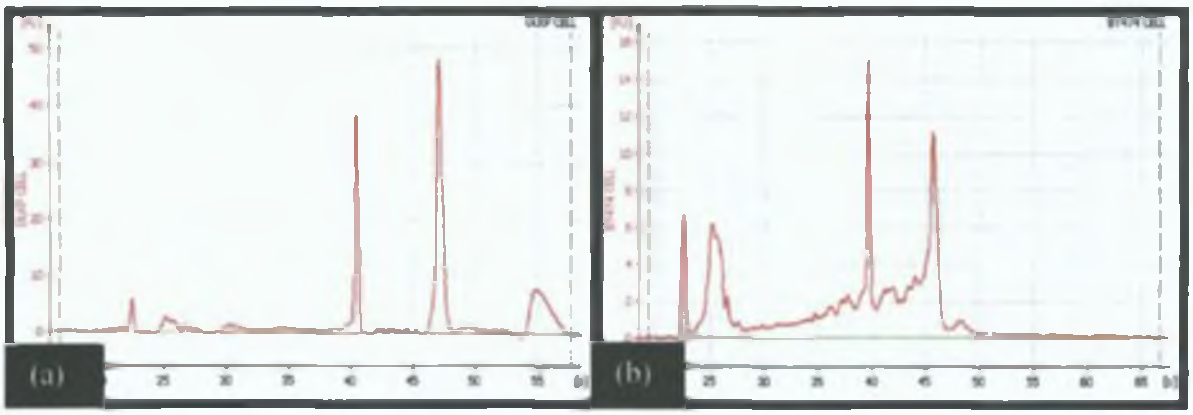


Fig. 2.9.2.2 Examples of electrophoreograms generated using the Agilent Bioanalyser. (a) intact DLKP cell RNA as evident by the sharp 18S and 28S ribosomal bands, (b) slightly degraded BT474 cell RNA as evident by the extra peak at approximately 20secs indicating the presence of smaller fragments of RNA.

2.10 Reverse transcriptase reaction

Reverse transcriptase (RT) reactions were carried out using micropipettes which were specifically allocated to this work. The micropipettes were cleaned down using RnaseZap (Ambion 9870) according to the manufacturer's instructions.

First-strand cDNA was synthesised using the following procedure:-

2µl oligo (dT) 18 primers (500ng/µl) (Sigma)

+ X µl RNA* (from CM)/ 1µg RNA from cultured cells

+ Xµl water*

*the volumes of CM RNA were different depending on the protocol employed (see note below), so water was added to bring the total volume to 6µl.

The contents were mixed in a 0.5ml Eppendorf tube (Anachem 96.9216.9.01), heated to 72^0C for 5 min and then chilled on ice. To this, the following were added:-

- 2µl of a 10x buffer (500 mM Tris-HCl, pH 8.3, with 500 mM KCl, 30 mM MgCl2, and 50 mM DTT, B 8559)
- 1µl RNasin (40U/µl) (Sigma R2520)
- 1µl dNTPs (10mM of each dNTP) (Sigma DNTP100)
- 9µl DEPC -treated water
- 1µl Moloney murine leukaemia virus-reverse transcriptase (MMLV-RT) (200U/µl) (M1427 Sigma).

The solutions were mixed and the RT reaction was carried out by incubating the Eppendorf tubes at 37^0C for 1 hour. The MMLV-RT enzyme was then inactivated by heating to 95^0C for 2 min. The cDNA was stored at -20^0C until required for use in PCR reactions as outlined in Section 2.11

*Note: With the exceptions of the study to find a reproducible minimum detection level (amounts shown in Fig. 2.7.2.1, Section 2.7.2) and the validation of microarray results (see below), the amount of RNA solution used for reverse transcription was 4µl.

Validation of microarray results

In all reverse transcription reactions carried out in the validation RT-PCRs on samples used for microarray analysis (see validation of microarray data Sections 3.8.7 and 3.8.11), the amount of RNA used was double the original volume used in the microarray amplification and labelling procedure (see Table 2.10.1 for summary of volumes used for microarray amplification protocol). The reason for this is that the minimum equivalent volume of CM required for reproducible RT-PCR was determined to be 50µl (see Section 3.2). The

minimum volume of 50µl would equate to 1.2µl of RNA solution for the samples used in the microarray study (because of the smaller volumes of DEPC treated water that the RNA samples were resuspended in, see Section 2.8.2.2). As the volume of RNA for these samples used in the microarray amplification procedure were below the 1.2µl volume (see Table 2.10.1), the volume of each RNA solution used for the validation of microarray results by RT-PCR was doubled (see Section 3.8.7 and 3.8.11 for these RT-PCR results). Ultimately this meant that the volume for each sample used in the microarray validation RT-PCRs was an exact reflection of the volumes used for the amplification and labelling procedures.

Sample		Amount used for amplification and labelling (Section 2.12) Volume(µl) = 100ng	Amount used for validation of microarray results using RT-PCR (µl)
H1299	1	0.9	1.8
	2	1	2
	3	1.2	2.4
H1299 TAX	1	1.3	2.6
	2	1	2
	3	1.1	2.2
RPMI	1	1.4	2.8
	2	1	2
	3	1.5	3
RPMI TAX	1	0.6	1.2
	2	0.7	1.4
	3	0.7	1.4
RPMI ML	1	1	2
	2	0.8	1.6
	3	0.8	1.6
MDA-F	1	0.9	1.8
	2	1.1	2.2
	3	0.8	1.6
MDA-F–TAX-SI	1	0.9	1.8
	2	0.9	1.8
	3	0.8	1.6
MDA-F-ADR-SI	1	0.9	1.8
	2	0.8	1.6
	3	0.9	1.8

Table 2.10.1 Volume of RNA extracted from CM and used for microarray amplification protocol and in subsequent RT-PCR validation studies.

2.11 Polymerase chain reaction

A modified polymerase chain reaction (PCR) procedure was followed in this study.

5 µl cDNA from CM or 2.5 µl cDNA from cells was amplified in a 25 µl PCR reaction

Each PCR tube contained the following:-

- 14.25µl$^{\Psi}$ UHP
- 2.5µl 10x buffer (100 mM Tris-HCl, pH 8.3, 500 mM KCl, Sigma D4545)
- 1.5µl 25mM-$MgCl_2$ (Sigma 4545)
- 0.5µl dNTPs (25mM each of dATP, dCTP, dGTP and dTTP) (Sigma DNTP100)
- 1µl each of first and second strand target primers (250ng/µl)
- 0.5µl of 5U/µl *Taq* DNA polymerase enzyme* (Sigma D4545)
- 5µl cDNA from CM/ 2.5µl cDNA from cells + 2.5µl UHP

$^{\Psi}$ If an endogenous control was required to be amplified along with gene of interest, this amount was reduced to 12.25µl and 0.5µl of each primer also added to the mix.

cDNA was amplified by PCR as follows:

95^{0}C for 5 min - to denature double-stranded cDNA

30 cycles*: 95^{0}C for 30 sec - denature

54^{0}C-58°C for 30 sec - anneal

72^{0}C for 45 sec - extend

72^{0}C for 10 min. – extend

The reaction tubes were then stored at 4^{0}C until analysed by gel electrophoresis as described in Section 2.11.2

* 30 cycles were sufficient when amplifying gene transcripts from cell RNA however, it was sometimes necessary to increase this to 45 cycles when using RNA extracted from

CM.

In the validation of microarray results using RT-PCR, it was found necessary to reduce the cycle number to 35 for the CM samples to prevent saturation of the PCR bands and allow semi-quantitative differences between samples to be ascertained. All primer sequences used in RT-PCR studies are listed in Tables 2.11.1.1 and 2.12.1.2.

2.11.1 Primer sequences used for RT-PCR studies

All primers listed in Table 2.11.1.1 and 2.11.1.2 (not published elsewhere), were designed to be RNA specific. Human Blat searches (http://genome.ucsc.edu/) were carried out on all transcripts of interest to identify exon- intron boundary regions (Kent, 2002). At least one primer of the primer pair set was designed to cross an exon- intron boundary junction, so that only amplification of an mRNA transcript was possible. Fig. 2.11.1.1 shows the regions of the mRNA transcripts amplified using the primers listed in Table 2.11.1.1

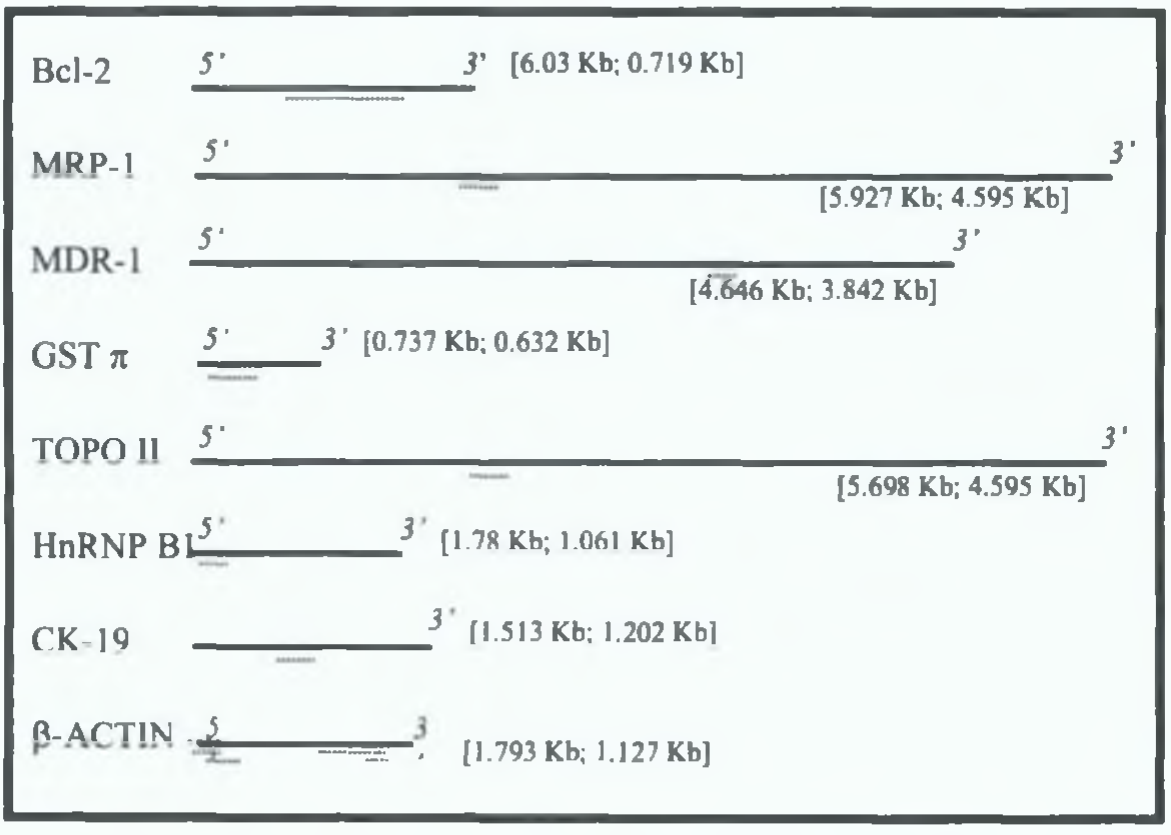


Figure 2.11.1.1 Location of regions selected for amplification. 5', central or 3' location of the regions selected for amplification (——) along the coding regions (————) of transcripts analysed, using primers described in Table 2.11.1.1. Information on size of full gene transcripts, as well as size of coding regions only – as illustrated here – are included for all gene products [ ] analysed.

cDNA	Primer Sequences	Amplified Product (bp)
bcl-2[x]	(F): 5' TCATGTGTGTGGAGAGCGTCAA 3' (R): 5' CTACTGCTTTAGTGAACCTTTTGC 3'	306
mrp-1[x]	(F): 5' GTACATTAACATGATCTGGTC 3' (R): 5' CGTTCATCAGCTTGATCCGAT 3'	203
mdr-1[x]	(F): 5' GTTCAAACTTCTGCTCCTGA 3' (R): 5' CCCATCATTGCAATAGCAGG 3'	157
GST-π[x]	(F): 5' ATGCTGCTGGCAGATCAG 3' (R): 5' GTAGATGAGGGAGATGTATTTGCA 3'	270
Topo II[x]	(F): 5' AACTTTGGCTGTTTCAGG 3' (R): 5' ATCATTATCTTCCCATAACGAAGCGT 3'	216
HnRNP B1[$]	(F): 5' GGAGAGGAAAAAGAGAGAAAAG 3' (R): 5' TTGATCTTTTGCTTGCAGGA 3'	155
CK-19[$]	(F): 5' GCGGGACAAGATTCTTGGTG 3' (R): 5' CTTCAGGCCTTCGATCTGCAT 3'	214
β-actin[x]	(F): 5' TGGACATCCGCAAAGACCTGTAC 3' (R): 5' TCAGGAGGAGCAATGATCTTGA 3'	142
β-actin	(F): 5' CTTTGCCGATCCGCCGCCCGTC 3' (R): 5' CATCACGCCCTGGTGCCTGGGG 3'	171
β-actin[$]	(F): 5' ACGGCTCCGGCATGTGCAAG 3' (R): 5' TGACGATGCCGTGCTCGATG 3'	196
β-actin[x]	(F): 5' GAAATCGTGCGTGACATTAAGGAGAAGCT 3' (R): 5' TCAGGAGGAGCAATGATCTTGA 3'	383

Table 2.11.1.1 Primer sequences used for CM RT-PCR study (see Section 3.1-3.4). *Note*: β-actin primers were designed to amplify a range of products (142, 171, 196 and 383) at a range of locations (5', central, or 3' along the length of the transcript). All PCRs carried out at 54°C annealing temperature and 45 PCR cycles. x/ $ = primers published elsewhere, x= O'Driscoll *et al* (1996), $= (with the exception of HnRNPB1 forward primer)primers published by Fleischhaker *et al* (2001).

For the primers designed to validate the microarray study, they were also designed to be as close to, or within the region of the mRNA transcript also targeted by the Affymetrix probes (Fig 2.11.1.3 for example and Table 2.11.1.3 for details).

cDNA	Primer Sequences (5'- 3')	Amplified Product (bp)	Annealing Temp (cycle no)
Hyp Protein FLJ32784	(F): ACTACATACCTCTGGTGTCTCTGCAG (R): CTTGGACTCAGACGGGCG	158	56°C (45 cycles)
SMAF 1	(F): ATTAGAGTTTGGTTTTCCTGGAGTCA (R): GGCTTTCACATGTCGGTCTGA	152	54°C (45 cycles)
Clone IMAGE: 5528716	(F): CATATCTCAGGTGACAAGAAGGTTACC (R): TTAGAGCTCCTGAAAGTGACACCA	152	54°C (35 cycles)
ABCB4	(F): GACTAGATATGCATATTACTACTCAGGATTG (R): TAATTTCTATATGAAAGTGTGACATTAACAATG	239	54°C (35 cycles)
Clone HLUNG 2003061	(F): CCCAGAGAGGAAGAGTAAAGAGGAC (R): GCAATACCAACGCCTTCATTTC	154	54°C (35 cycles)
NBR2	(F): GTTCATTTAGCAGCTCCGAGG (R): TTCGCGAAGCTGGGCT	179	56°C (35 cycles)
COMT	(F): TCTCTTGGAGGAATGTGGCC (R): CTCCCTGTATTCCAGGAACGAT	151	54°C (35 cycles)
HLXB9	(F): TTCAACTCCCAGGCGCAGT (R): CAAATCTTCACCTGGGTCTCG	180	56°C (45 cycles)
DDI2	(F): GGTCAAAGGGATCAGCAGAAAAC (R): TTGTTCTTTAGATGAACAGGTCTTTACC	163	54°C (35 cycles)
APIP	(F): CAAAACCATGTGTGAGTGTTATGACT (R): TTCTTTTGGCTTAGACAATTCCATT	119	54°C (35/ 45 cycles)*
SCD5	(F): GAAATACACAGCACATCCAGAAAGAA (R): GAGTGGTTACAATGAGTATGTAACAAAGCT	151	54°C (45 cycles)

Table 2.11.1.2 Primers used to validate microarray study data. *45 cycles used for validation of Present/ Absent call study (see Section 3.8.7.3), 35 cycles used for validation of trend change in expression (see Section 3.8.11).

Worked Example of How PCR Primers for Validation of Microarray Results Were Chosen

The Netaffx database (see section 2.13.9.3) was accessed to find the transcript assignment for the probe set 1570505_at. The sequence of the transcript assignment (BC020618) was then retrieved from the NCBI database. This full sequence was imported into BLAT to aid in selection of exon- intron boundary sites.

Fig. 2.11.1.2 shows the results of the BLAT search. The letters highlighted in light blue often indicates exon- intron boundary junctions. Any text in black indicates a region that does not align to the genome, as per example the poly A tail in Fig. 2.11.1.2. The exon-intron boundary circled in red in Fig 2.11.1.2 was chosen for incorporation into primer design, as it was the most 3' exon- intron boundary point identified.

Figure 2.11.1.2 BLAT results for transcript BC020618 (ABCB4). The letters in dark blue are the nucleotides that align exactly with the genome, the letters in light blue indicate points where the sequence aligns with the genome if a piece of genomic data is taken out (Example of one circled in red).

Next, the transcript sequence was aligned with both the Affymetrix target region and the proposed selected primers to establish if there was overlap using multalin software (http://prodes.toulouse.inra.fr/multalin/multalin.html). Affymetrix recommends picking primers in the same region that they target so that the resultsfor subsequent RT-PCR analysis have a higher probability of validating the microarray data.

In Fig. 2.11.1.3 the first (and longest) sequence shown is the transcript assignment BC020618. The second sequence if the Affymetrix target region (called HG-U133_PLUS; see left-hand side). The concensus/ overlap sequence between these is on the bottom. The part of the transcript highlighted in red is the part that overlaps with the Affymetrix target and it is clear that parts of this region is within the first 50 bases of the 3' end of the sequence. The dark blue undashed lines indicate the primer annealing sites and the dashed lines between then, the amplified region.

The region circled in red was the most 3' exon-intron boundary site identified. The closest region that could be targeted using mRNA specific primers (blue line) overlaps the start of the Affymetrix targeted region. Details of the regions targeted by Affymetrix and by our RT-PCR primers for the validation of microarray data study are listed in Table 2.11.1.3.

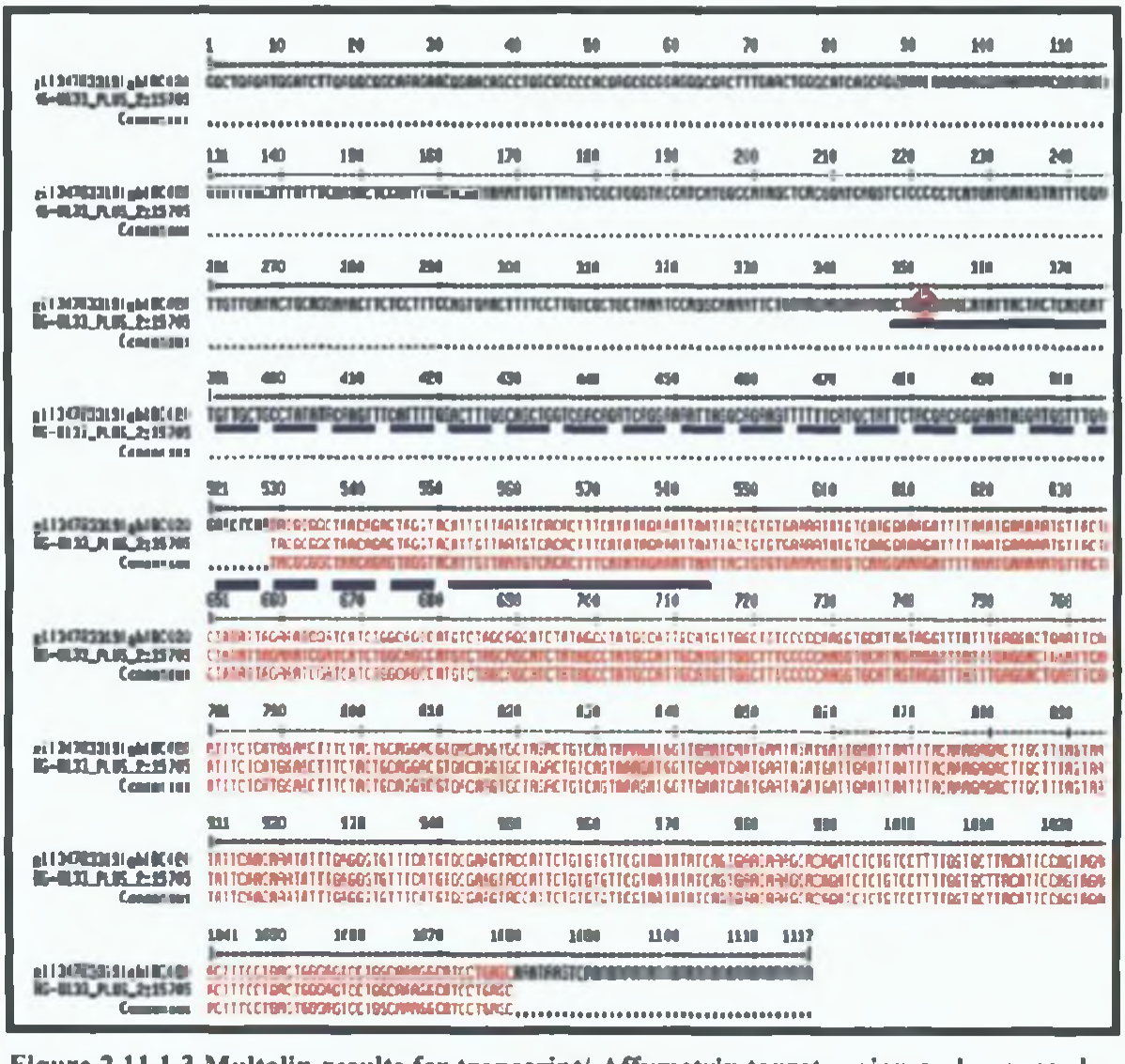


Figure 2.11.1.3 Multalin results for transcript/ Affymetrix target region and proposed primer positions. The first (and longest) sequence is the transcript assignment BC020618. The second sequence is the Affymetrix target region (called HG-U133_PLUS on the left hand side). The consensus/ overlap sequence is on the bottom. The part of the transcript highlighted in red is the part that overlaps with the Affymetrix target. The blue undashed lines indicate the primer annealing sites and the dashed lines between them, the amplified region. The region circled in red was the most 3' exon- intron boundary site that could be found (sequence GA, see also Fig. 2.11.1.2, this sequence is shaded light blue indicating that it is most likely a splice site).

Probe set ID HG U133A Plus 2.0	Accession No. (NCBI)	Gene Name	mRNA length (CDS)	PCR Amplicon Region	Affymetrix Target Region
1552853_at	NM_144623[1]	Hyp Protein FLJ32784	4181bp (1194-3398bp)	3673-3830	3619-3678
1554128_at	NM_001018082	small adipocyte factor 1 (SMAF1)	770bp (57-229bp)	403-554	191-285
1561573_at	BC039476	Clone IMAGE:5528716	1023bp (?)	287-438	439- 957
1570505_at	BC020618	ATP-binding cassette, subfamily (MDR/TAP), member 4 (ABCB4)	117bp (8- 547bp)	346-584	529-1079
1557996_at	AK091784	cDNA FLJ34465 fis, clone HLUNG2003061	1866bp (?)	564-717	1281-1829
207631_at	NM_005821	Neighbour of BRCA1 gene (NBR2)	1392bp (332-670bp)	975-1153	803-1293
208817_at	NM_000754[2]/ NM_007310	Catechol-O-methyltransferase (COMT)	1289bp (203-1018bp) 1067bp (131-796bp)	808- 958 586- 736	653-1000 431- 778
214614_at	NM_005515	Homeo box HB9 (HLXB9)	2176bp (304-1509bp)	987-1166	1643-2115
224449_at	BC006011	DNA- damage inducible protein 2 (DDI2)	1524bp (129-764bp)	625- 787	995- 1464
218698_at	NM_015957	APAF1 interacting protein (APIP)	1226bp (78- 806bp)	698- 816	627-1190

Table 2.11.1.3 Details of ten probe sets chosen for microarray confirmation RT-PCR analysis. CDS= coding sequence, [1]Sequence removed by NCBI Refseq staff since the study was carried out. [2]NM_000754 is 220 bases longer than NM_007310 at the 5' end, but the rest of the sequence is identical. As Affymetrix probe sets targets the identical part, RT-PCR primers were designed to this part also. (?) indicates sequences for which the length of the coding sequence was not listed on the NCBI website listing for that gene transcript.

2.11.2 Electrophoresis of PCR products

A 2% agarose gel (Sigma, A9539) was prepared in TBE buffer (5.4g Tris, 2.75g boric acid, 2ml 0.5M-EDTA pH 8 in 500ml water) and dissolved in a microwave oven. After allowing cooling, 0.003% (v/v) of a 10mg/ml ethidium bromide solution was added to the gel after it was poured into an electrophoresis apparatus (BioRad). Combs were placed in the gel to form wells and the gel was allowed to set.

5µl loading buffer (50% glycerol, 1mg/ml xylene cyanol, 1mg/ml bromophenol blue, 1mM EDTA) was added to 25µl PCR samples and 10µl was run on the gel at 80-90mV for approximately 2 hours. When the dye front was seen to have migrated the required distance, the gel was removed from the apparatus and examined on a UV-transilluminator and photographed.

2.11.3 Densitometry Analysis

Densitometry analysis of the PCR gels was carried out using Labworks Analysis Software (Version 3.0) to measure the maximum OD of the PCR bands in each gel. The results were imported into Microsoft Excel and bar charts were generated from this data.

2.12 RNA amplification, labelling and fragmentation of cRNA in preparation for hybridisation to Affymetrix array chips

Components required for this protocol were included in the Two-Cycle Target Labelling and Control Reagents (Affymetrix, P/N 900494) and MEGAscript® High Yield Transcription Kit, Ambion Inc, P/N 1334 with the exception of Ethanol (Sigma E7023).

The positive control Poly-A RNA is firstly diluted before spiking in with the sample RNA (see Section 2.12.1 for more information). It is important to note that the Poly-A spikes

were made up in non-stick RNase/ DNase free tubes (Ambion cat no. 12450(1.5ml)/ 12350(0.5ml)), which prevents the Poly-A spikes from sticking to the sides of the tubes and interfering with the final concentration of the positive controls.

2µl of the Poly-A spike control was mixed with 2µl of Affymetrix approved T7-Oligo dT primer (50µM) and 16 µl of RNase-free water. The tube was then flicked and centrifuged briefly. 2µl of this mix was added to 2µl of the 50ng/µl RNA sample (see Table 2.10.1, 3µl stock solutions of 50ng were made based on the volumes listed in Table 2.10.1 for 100ng of RNA) and 1µl of RNase-free water. The tubes were flicked and centrifuged briefly before being incubated for 6 minutes at 70°C. They were then incubated for 2 minutes on ice and centrifuged briefly.

As summarised in Fig. 2.12.1, the first cycle, first strand master mix was prepared by mixing 2µl of 5X first strand reaction mix, 1µl DTT (0.1M), 0.5µl RNase Inhibitor, 0.5µl dNTPs (10mM) and 1µl Superscript II for each sample required. It is worth noting that Affymetrix suggest that if there are more than 2 samples that it is prudent to include extra to compensate for potential pipetting inaccuracy or solution lost during the process. The 5µl of first strand master mix was added to each sample, the tube gently flicked, briefly centrifuged and placed immediately at 42°C for 1hour and 72°C for 10 minutes before being placed on ice for 2 minutes. The first cycle second strand master mix was prepared by adding 4.8µl of RNase-free water, 4µl of $MgCl_2$ (17.5mM), 0.4µl dNTP's (10mM), 0.6µl *E.Coli* DNA Polymerase I and 0.2µl of RNase H together for each sample required. This 10µl mix was then added to each tube which were then flicked and centrifuged briefly before being placed at 16°C for 2 hours and 75°C for 10 minutes and then ice for 2 minutes. The components for the first cycle IVT amplification (Ambion Megascript T7 kit) were assembled at room temperature. 5µl of each of the components ATP, CTP, UTP, GTP, enzyme mix and 10x reaction buffer were added together for each sample included before

being added to each sample. The tubes were gently flicked, centrifuged and placed at 37°C for 16hours.

The cRNA was then purified using the GeneChip Sample Cleanup module (Affymetrix, 900371) as recommended by the manufacturers instructions.

The quantity of the cRNA was subsequently checked by diluting 2µl of cRNA in 18µl of H_2O and reading the quantity by using the Nanodrop (see Section 2.9.1). 600ng is required for the second cycle of the protocol. (All CM and serum specimens had less than 600ng so all 9µl left was used for the next step, see Tables 3.7.1.1, 3.8.1.1 and 3.9.2.1 for results). 2µl of freshly diluted random primers (3µg/µl) was added to each sample and the tubes were flicked, centrifuged briefly and placed at 70°C for 10 minutes before being placed on ice for 2 minutes.

As summarised in Fig. 2.12.2, the second cycle first strand mix was prepared by adding 4µl of 5x first strand mix, 2µl of DTT (0.1mM), 1µl RNase Inhibitor, 1µl dNTP's (10mM) and 1µl Superscript II together for each sample included. This 9µl was added to each sample before they were flicked, centrifuged and placed at 42°C for 1hour and ice for 2 minutes. After this, 1µl of RNase H was added to each sample before they were flicked, spun, and placed at 37°C for 20 minutes, 95°C for 5 minutes and ice for 2 minutes.

4µl of a freshly prepared aliquot of T7 Oligo dT primer was added to each sample before they were flicked, centrifuged briefly and incubated at 70°C for 6 minutes and ice for 2 minutes. The second cycle second strand master mix was prepared by adding the following in a tube for each sample required; 88µl RNase-free water, 30µl 5x second strand reaction mix, 3µl of dNTPs (10mM) and 4µl *E.Coli* DNA polymerase I. This 125µl mastermix was added to each sample before being flicked, centrifuged briefly and incubated for 2 hours at 16°C. T4 DNA polymerase (2µl) was then added to each sample before incubating at 16°C for a further 10 minutes. After incubating the samples at 4°C for 2 minutes, they were

immediately purified using the GeneChip Sample Cleanup module (Affymetrix, 900371) following the manufacturers instructions.

All 12µl of cDNA were used for the second IVT step. The reagents required for this step were assembled at room temperature. The master-mix included 8µl RNase- free water, 4µl of IVT labelling buffer, 12µl IVT labelling NTP mix and 4µl labelling enzyme mix for each sample included. The 28µl volume was added to each sample before flicking, centrifuging briefly and incubating at 37°C for 16 hours.

The biotin-labelled cRNA was purified using the GeneChip Sample Cleanup module (Affymetrix, 900371), as recommended by the manufacturers, and quantified using a Nanodrop (a sample can also be taken at this point for analysis using a Bioanalyser nano chip). For quantification of cRNA when using total RNA as starting material, an adjusted cRNA yield needed to be calculated to reflect carryover of unlabeled total RNA. Using an estimate of 100% carryover, the formula below was used to determine adjusted cRNA yield:

adjusted cRNA yield = RNAm - (total RNAi) (y)

where RNAm = amount of cRNA measured after IVT (µg), total RNAi = starting amount of total RNA (µg)

y = fraction of cDNA reaction used in IVT (= 1 for CM samples)

The final step of the entire process was to fragment 20µg of the biotin-labelled cRNA by adding 8µl of fragmentation buffer to 20µg of cRNA and bringing the total volume of the reaction to 40µl, so that the concentration of the cRNA is 0.5µg/µl. This mix was incubated for 35 minutes at 94°C. A sample can be taken at this point for analysis on a Bioanalyser nano chip. From the fragmented cRNA 30µl (=15µg) was hybridised to the Affymetrix U133+2 chip. Figs. 2.12.1 and 2.12.2 present an overview of the 2 cycle amplification procedure discussed above.

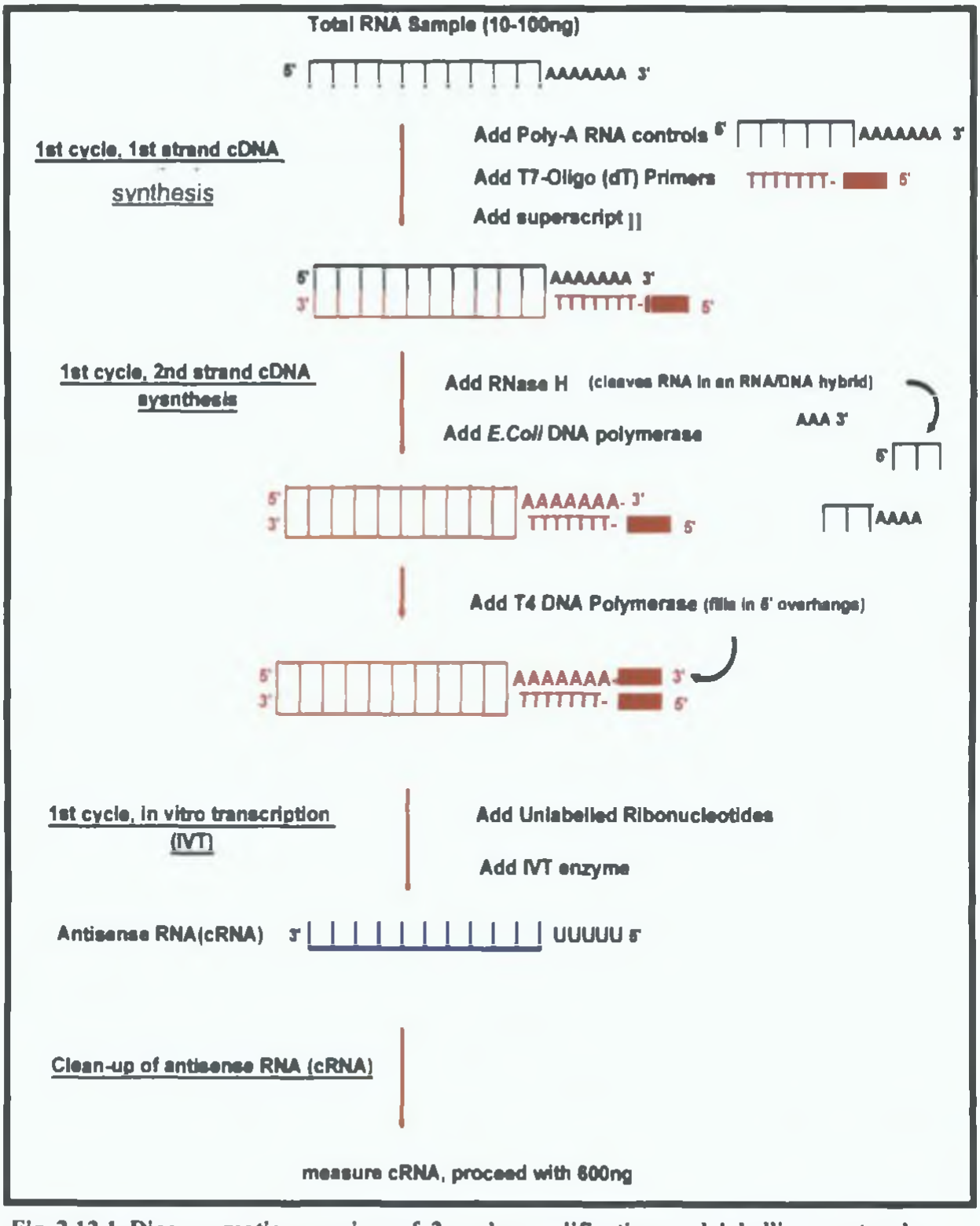


Fig 2.12.1 Diagrammatic overview of 2 cycle amplification and labelling protocol-cycle1

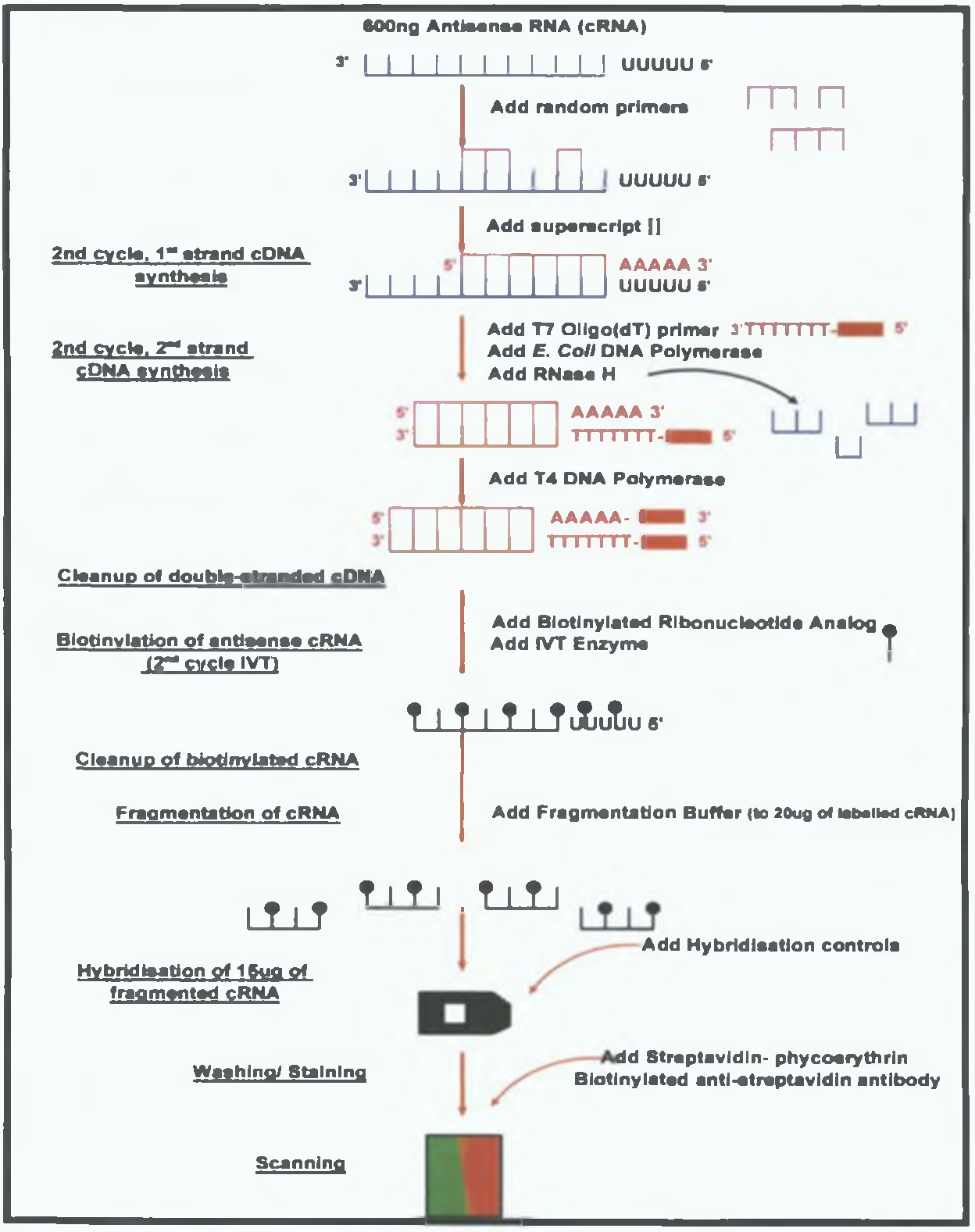


Figure 2.12.1 Diagrammatic overview of 2 cycle amplification and labelling protocol-cycle 2

2.12.1 Two-Cycle Target Labelling and Hybridisation Controls for Quality Control

Affymetrix supply a Eukaryotic Poly-A RNA Control Kit along with the Two-Cycle Target Labelling and Control Reagents. The kit is designed specifically to provide exogenous positive controls to monitor the entire GeneChip eukaryotic target labelling process. Each array contains probe sets for several *B. subtilis* genes that are absent in the samples analysed (*lys*, *phe*, *thr*, and *dap*). *Dap, lys, phe, thr*, and *trp* are *B. subtilis* genes that have been modified by the addition of poly-A tails, and then cloned into pBluescript vectors, which contain T3 promoter sequences. Amplification of the poly-A controls with T3 RNA polymerase yields sense RNAs, which can be spiked into RNA samples, carried through the sample preparation process, and evaluated like internal control genes. These Poly-A spike-in controls are pre-mixed at staggered concentrations to allow assessment of the overall success of the assay. The final concentrations of the controls, relative to the total RNA population, are: 1:100,000; 1:50:000; 1:25,000; 1:7,500, respectively. All of the Poly-A controls should be called Present with increasing Signal values in the order of *lys, phe, thr, dap.*

The hybridisation controls also provided with the Two-Cycle Target Labelling and Control Reagents include *bioB*, *bioC*, and *bioD* (from the biotin synthesis pathway of *E. coli*), and *cre* (from the recombinase gene of bacteriophage P1). The Affymetrix arrays contain probe sets for these non-eukaryotic transcripts, which serve as controls for hybridisation, washing, and staining procedures. These transcripts were labelled separately then added to the biotinylated target cRNA before hybridisation at 1.5 pM, 5 pM, 25 pM, and 100 pM final concentrations for *bioB, bioC, bioD*, and *cre*, respectively. A synthetic biotinylated control oligo (Oligo B2) is also added to the hybridisation solution to provide alignment signals for image analysis commonly known as "landing lights".

BioB is at the threshold level of assay sensitivity (1:100,000 complexity ratio) and should be called Present at least 50% of the time. *BioC, bioD*, and *cre* should always be called "Present" with increasing signal values, reflecting their relative concentrations.

The hybridisation controls can be used to indirectly assess RNA sample quality among replicates. When global scaling is performed, the overall intensity for each array is determined and is compared to a Target Intensity value in order to calculate the appropriate scaling factor. The overall intensity for a degraded RNA sample, or a sample that has not been properly amplified and labelled, will have a lower overall intensity when compared to a normal replicate sample. Thus, when the two arrays are globally scaled to the same Target Intensity, the scaling factor for the "bad" sample will be much higher than that for the "good" sample. Since the hybridisation controls are added to each replicate sample equally (and are independent of RNA sample quality), the intensities of the *bioB, bioC, bioD*, and *cre* probe sets will be approximately equal in all cases. As a result, the signal values (adjusted by scaling factor) for these control probe sets on the "bad" array will be adjusted higher relative to the signal values for the control probe sets on the "good" array.

2.13 Affymetrix GeneChips®

(All pictures in this section taken, with permission, from the Affymetrix website www.Affymetrix.com unless otherwise stated)

The microarray gene expression experiments which were performed in this body of work were carried out using Affymetrix® Human Genome U133A Plus 2 GeneChips®. The GeneChip® Human Genome U133 Plus 2.0 Array offers comprehensive analysis of genome-wide expression on a single array by analysing the expression level of over 47,000 transcripts and variants, including 38,500 well-characterised human gene transcripts. It is comprised of more than 54,000 probe sets and 1,300,000 distinct oligonucleotide features. These features are only 11 microns. Fig. 2.13.1.1(A) shows the size of the Human Genome Plus 2.0 Array chip.

GeneChip probe arrays are manufactured using technology that combines photolithography and combinatorial chemistry. Up to 1.3 million different oligonucleotide probes are synthesised on each array. The surface of the Affymetrix array is like a giant grid of squares that has been shrunk down to the size of a thumbnail. Each oligonucleotide is located in a specific area on the array called a probe cell. Each probe cell contains hundreds of thousands to millions of copies of a given oligonucleotide (see Fig.2.13.1.1 (B)).

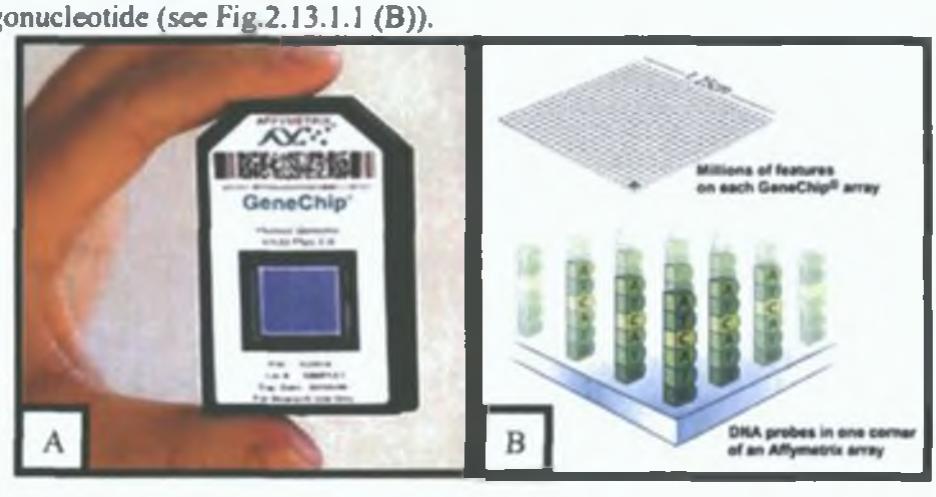


Figure 2.13.1.1 (A) Affymetrix GeneChip Human Genome Plus 2.0 Array, (B) Picture representation of feature design on arrays.

The probe arrays are manufactured in a series of cycles. Initially, a glass substrate is coated with linkers containing photo labile protecting groups. Then, a mask is applied that exposes selected portions of the probe array to ultraviolet light. Illumination removes the photo labile protecting groups enabling selective nucleoside phosphoramidite addition only at the previously exposed sites (see Fig. 2.13.1.2 for example). Next, a different mask is applied and the cycle of illumination and chemical coupling is performed again. By repeating this cycle, a specific set of oligonucleotide probes is synthesised with each probe type in a known location. The completed probe arrays are packaged into cartridges.

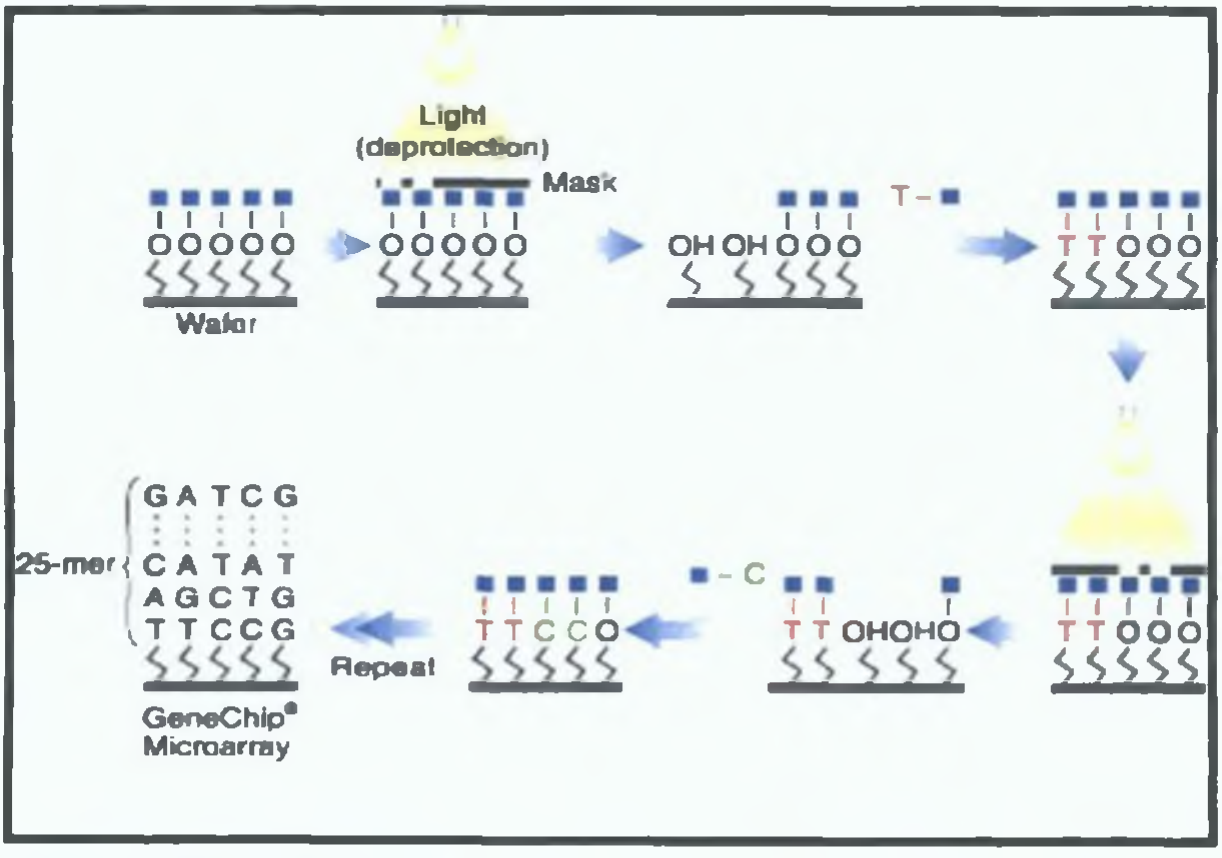


Figure 2.13.1.2 Photolithography process

2.13.1 Perfect/ Mismatch Probe Strategy

A core element of Affymetrix microarray design is the Perfect/Mismatch probe strategy. For each probe that is designed to be perfectly complimentary to a given target sequence, a partner probe is also generated that is identical except for a single base mismatch in its centre (see Fig. 2.13.1.1). These probe pairs, called the Perfect Match probe (PM) and the Mismatch probes (MM), allow the quantitation and subtraction of signals caused by non-specific cross-hybridisation. The differences in hybridisation signals between the partners, as well as their intensity ratios, serve as indicators of specific target abundance.

The chips have a "distributed probe set format", where the probe pairs of the same probe set are scattered into various places on the chip. This is to prevent a local image contamination to completely destroy the PM and MM information of a probe set.

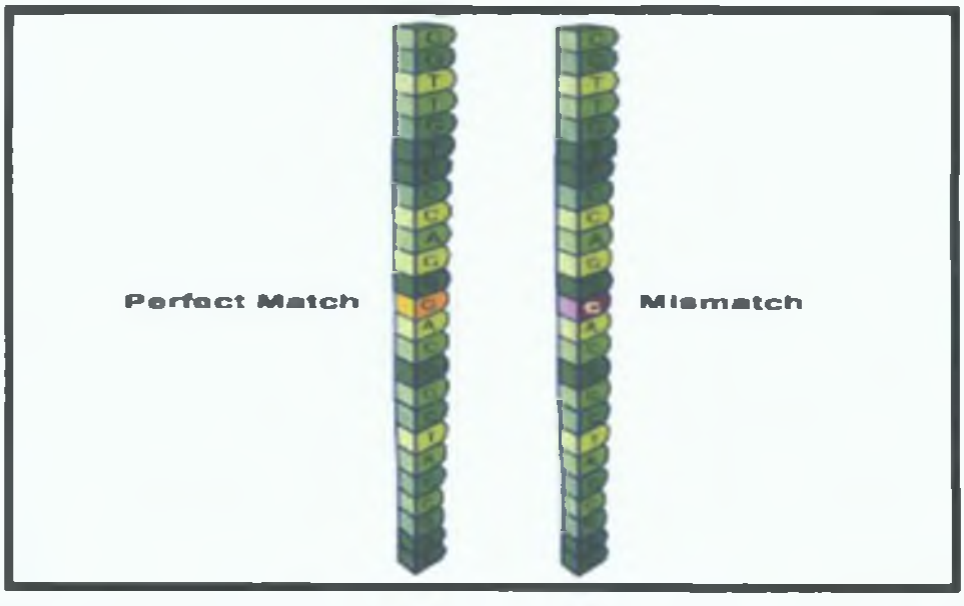


Figure 2.13.1.1 An example of a probe pair with perfect match/ mismatch oligonucleotides.

An example of a PM/ MM result for a probe set using GeneChip software is shown in Fig. 2.13.1.2. Each box represents the signal proportional to the bound target for each probe. The amount of light emitted at 570 nm is proportional to the bound target at each location on the probe array. Target abundance is calculated by subtracting the MM signal from the PM signal for each probe pair.

Figure 2.13.1.2 An example of scanned results for an individual probe set (generated using microarray data originating from our laboratory)

The PM/MM data for the same probe set (for a different sample) shown above in Fig. 2.13.1.2 was calculated using dChip software and is shown in Fig. 2.13.1.3. On the left, a grid displays the PM and MM data in blue and green curves, with the x-axis ordering probe sets from 1 to 20, and y-axis for probe intensities with the range (0, 5763).

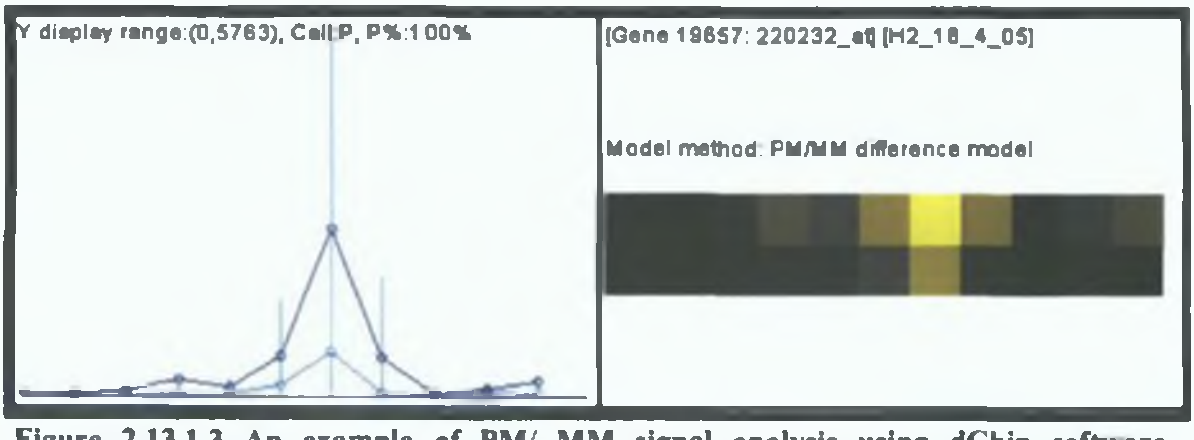


Figure 2.13.1.3 An example of PM/ MM signal analysis using dChip software (generated using microarray data originating from our laboratory).

2.13.2 Sequence retrieval

The sequences, from which the probe sets were derived, were selected from GenBank®, dbEST, and RefSeq. The sequence clusters were created from the UniGene database (Build 133, April 20, 2001) and then refined by analysis and comparison with a number of other publicly available databases, including the Washington University EST trace repository and the University of California, Santa Cruz Golden-Path human genome database (April 2001 release). Sequences from these databases were collected and clustered into groups of similar sequences.

The sequences were aligned to the draft of the human genome. They were subsequently orientated using consensus splice sites from the genome alignments, detected polyadenylation sites, coding sequence (CDS), and EST read direction annotations. Vector sequences were identified and removed along with low quality regions of EST sequences. Unigene (NCBI, Jan. 2003, Build 159) was then used to create initial clusters of cDNA sequences and sequence based subclustering was achieved using special software. Alternative transcript isoforms were split into separate clusters. At least 75% identity in all of the member sequences was required to create a consensus sequence.

The alignment process also extended the annotation information supplied by the databases pinpointing low quality sequences. These areas were usually trimmed for subsequent generation of high quality consensus sequences or, alternatively, Affymetrix employed quality ranking to select representative sequences, called exemplars, for probe design. Exemplar sequences are ESTs and mRNAs that are directly represented on the GeneChip array, whilst consensus sequences are built by the base calling algorithm, derived base-by-base from a cluster of public mRNA and/ or EST sequences. These sequences can be quite problematic. Fig. 2.13.3.1 shows an example of a consensus sequence built entirely on

ESTs. As ESTs have lower sequence accuracy, resulting probes may show as flawed when compared to the genome sequence and transcript record over time. mRNA sequences, which may appear later, are usually truncated at the 3' end so that EST- based consensus will overlap poorly with subsequent mRNA sequences that they correspond to *in vivo* (see top of Fig 2.13.2.1).

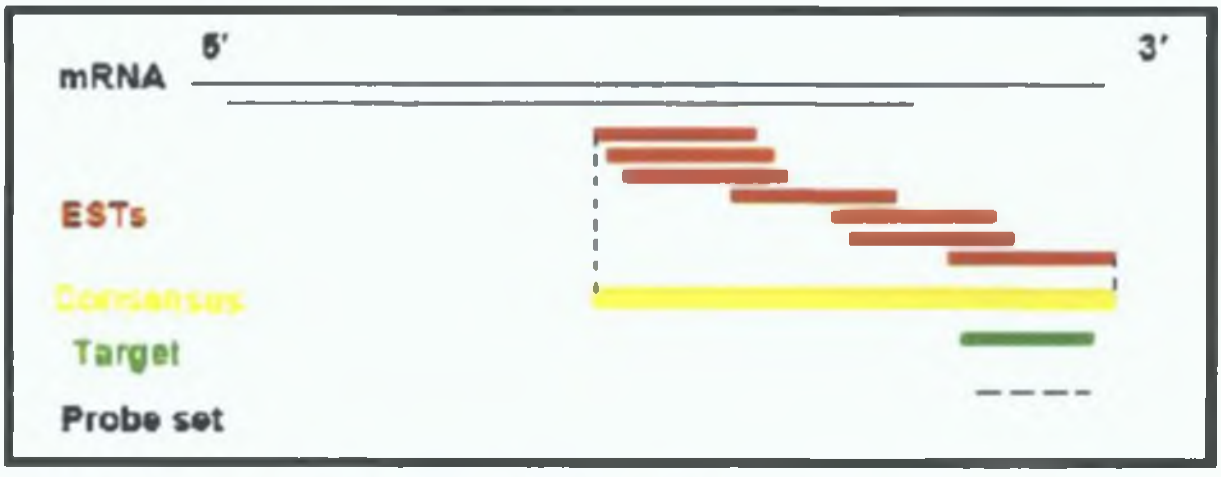


Figure 2.13.2.1 Example of consensus sequence build

Probe target regions were selected from the 600 bases closet to the 3' end of each transcript. The regions were categorised by having either 3' ends of RefSeq and complete coding sequence (CDS) mRNA sequences (full length end) or containing eight or more 3' EST reads terminating at the same position (strong evidence for polyadenylation) or 3' end of the assembly (consensus end).

2.13.3 Probe Set Design

The probe selection method used by Affymetrix for their U133 Plus 2.0 GeneChips takes into account probe uniqueness and the hybridisation characteristics of the probes, which allow probes to be selected, based on probe behaviour. Affymetrix use a multiple linear regression (MLR) model in the probe design that was derived from thermodynamic model of nucleic acid duplex formation. This model predicts probe binding affinity and linearity of signal changes in response to varying target concentrations. An advantage of this type of model-based probe selection system is that it provides a physical and mathematical foundation for systematic and large-scale probe selection. Also, an essential criterion of probe selection by Affymetrix for quantitative expression analysis is that hybridisation intensities of the selected probes must be linearly related to target concentrations

For each target, eleven probe pair sets have been selected, based on predicted probe characteristics including performance, uniqueness metrices and spacing rules. The probe sets were given different suffixes to describe their uniqueness and/ or their ability to bind different genes or splice variants.

- "_at" describes probes set that are unique to one gene
- "_a_at" describes probe sets that recognise multiple transcripts from the same gene
- "_s_at" describes probe sets with common probes among multiple transcripts from separate genes. The _s_at probe sets can represent shorter forms of alternatively polyadenylated transcripts, common regions in the 3' ends of multiple alternative splice forms, or highly similar transcripts. Approximately 90% of the _s probe sets represent splice variants. Some transcripts will also be represented by unique _at probe sets.
- "_x_at" designates probe sets where it was not possible to select either a unique

probe set or a probe set with identical probes among multiple transcripts. Rules for cross-hybridisation were dropped in order to design the _x probe sets. These probe sets share some probes identically with two or more sequences and therefore, these probe sets may cross-hybridise in an unpredictable manner.

Fig. 2.13.3.1 shows how the probe sets are designated "_at", "_s_at" and "_a_at". Not included in Fig 2.13.4.1 are probe sets with the suffix of "_x_at"

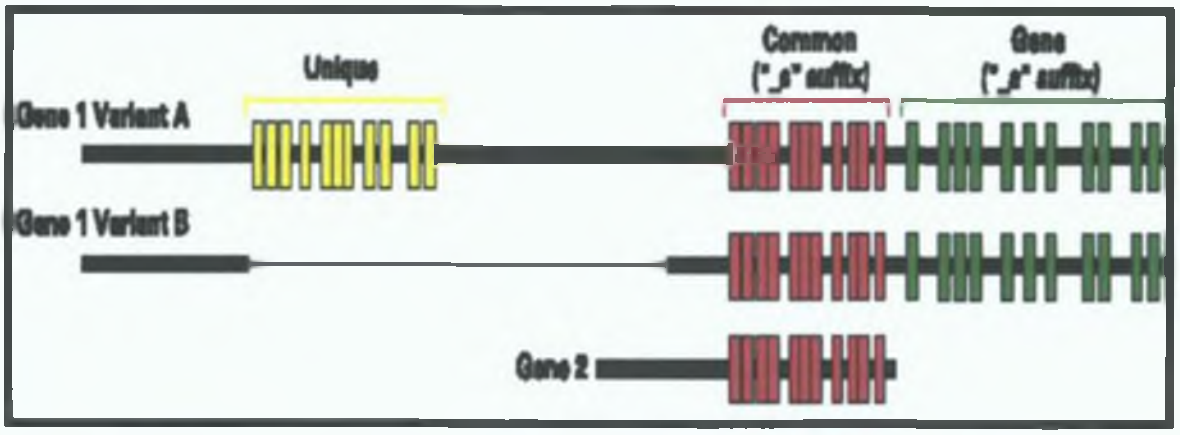


Figure 2.13.3.1 Probe set suffix assignment, yellow = "_at" unique probe set, pink = "_s_at" probe set, green = "_a_at" probe set.

2.13.4 Probe Set Grade Annotations

As well as defining the probe sets by their "uniqueness", Affymetrix also assign an annotation grade to each probe set. This annotation gives more information about the transcripts for which the probe sets are designed. The grades start at A, indicating that the probe sets match the mRNA sequence indicated as the assigned transcript. Other annotation grades include grade E and grade R annotations (grade R= probe sets that match only computationally predicted sequences) see Fig. 2.13.4.1 for example of grade A to C annotations.

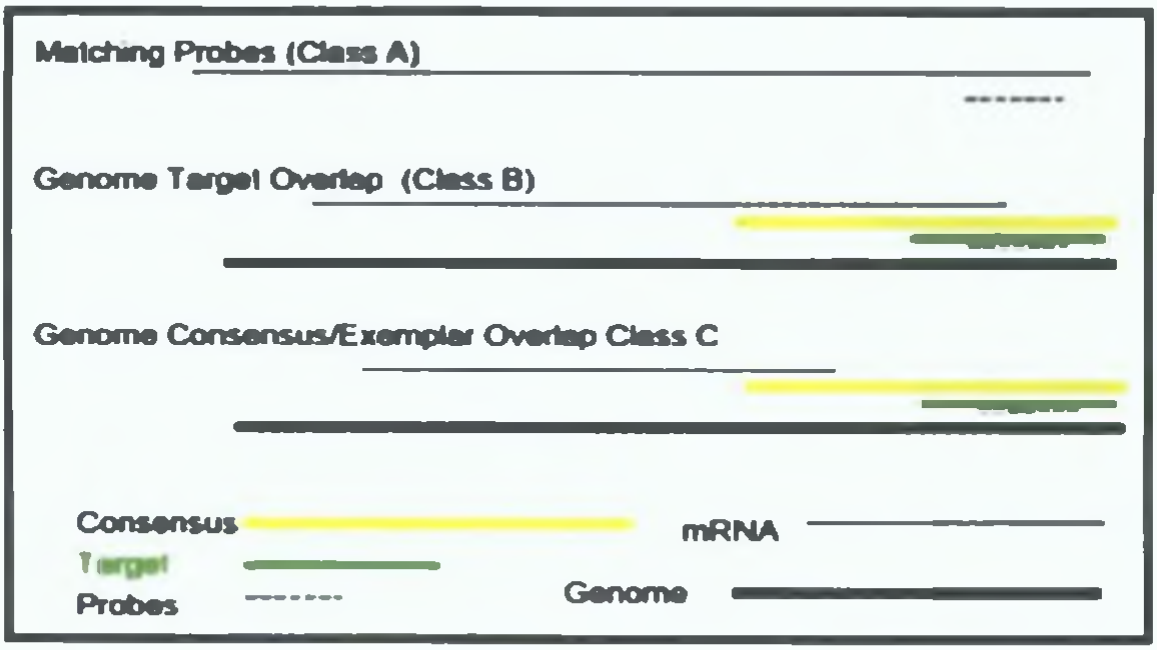


Figure 2.13.4.1 Annotation Grade assigned to probe sets, yellow line = consensus, green line = target, thick black line = genome and thin black line = mRNA.

Due to the different probe set suffixes and annotation grades, care must be taken when interpreting data that includes non-unique probe sets or probe sets without a class A grade annotation. It is possible to have a probe set that is unique to one gene, but that gene has been computationally predicted, so although the probe set suffix would be "_at", its annotation grade would be "R" and, therefore, not reliable.

2.13.5 Detection call

2.13.5.1 GeneChip software

Affymetrix have a system that analyses the raw data and assigns a "Presence" or "Absence" call to each gene transcript represented on the array. This analysis generates a Detection *p*-value, which is evaluated against user-definable cut-offs to determine the Detection call. This call indicates whether a transcript is reliably detected (Present) or not detected (Absent). Additionally, a signal value is calculated which assigns a relative measure of abundance to the transcript. Each probe pair in a probe set is considered as having a potential vote in determining whether the measured transcript is detected (Present) or not detected (Absent). The vote is described by a value called the Discrimination score [R]. The score is calculated for each probe pair and is compared to a pre-defined threshold *Tau*. Probe pairs with scores higher than *Tau* vote for the *presence* of the transcript. Probe pairs with scores lower than *Tau* vote for the *absence* of the transcript. The voting result is summarised as a *p*-value. The greater the number of Discrimination scores calculated for a given probe set that are above *Tau*, the smaller the *p*-value and the more likely the given transcript is truly Present in the sample. The *p*-value associated with this test reflects the confidence of the Detection call.

The Discrimination score is a basic property of a probe pair that describes its ability to detect its intended target. It measures the target-specific intensity difference of the probe pair (PM-MM) relative to its overall hybridisation intensity (PM+MM):

R = (PM - MM) / (PM + MM)

For example, if the PM is much larger than the MM, the Discrimination score for that probe pair will be close to 1.0. If the Discrimination scores are close to 1.0 for the majority of the probe pairs in a set, the calculated Detection *p*-value will be lower (i.e. more significant). A

lower *p*-value is a reliable indicator that the result is valid and that the probability of error in the calculation is small. Conversely, if the MM is larger than or equal to the PM, then the Discrimination score for that probe pair will be negative or zero.

The next step towards the calculation of a Detection *p*-value is the comparison of each Discrimination score to the user-definable threshold *Tau*. *Tau* is a small positive number that can be adjusted to increase or decrease sensitivity and/or specificity of the analysis (default value = 0.015). The One-Sided Wilcoxon's Signed Rank test is the statistical method employed to generate the Detection *p*-value. It assigns each probe pair a rank based on how far the probe pair Discrimination score is from *Tau*.

The user-modifiable Detection *p*-value cut-offs, Alpha 1 (α_1) and Alpha 2 (α_2), provide boundaries for defining Present, Marginal, or Absent calls. At the default settings, determined for probe sets with 16–20 probe pairs (defaults $\alpha_1 = 0.04$ and $\alpha_2 = 0.06$), any *p*-value that falls below α_1 is assigned a Present call; above α_2 is assigned an Absent call. Marginal calls are given to probe sets that have *p*-values between α_1 and α_2

The *p*-value cut-offs can be adjusted to increase or decrease sensitivity and specificity. Percent Present (%P) values depend on multiple factors including cell/tissue type, biological or environmental stimuli, probe array type, and overall quality of RNA. Replicate samples should have similar %P values.

2.13.5.2 dChip software

dChip can also accommodate the data from GeneChip software that detects Presence/ Absence calls using the methods described in Section 2.13.5.1 if it is included in a .txt file along with the .cel file. If these text files are not generated and included, as was the case in this study, dChip calculates the Presence/ Absence calls using a simplified version of the MAS4 algorithm that Affymetrix first used that is, the same decision matrix making calls is used but the background calculation is on the whole chip instead of 16 sectors.

In general the %Present calls using the dChip software is higher. In this thesis the %Present calls listed in the Affymetrix QA tables for each sample were taken from the GeneChip software, but as the data was normalised with dChip, the scatter plots/ Venn diagrams were generated using the dChip Present/ Absent calls. Although the %Present call was higher in the samples assessed by dChip, the increase was uniform across the samples.

2.13.6 Quality assessment of Affymetrix microarray chips

The quality of the data generated with Affymetrix microarray chips is assessed based on different criteria including the scaling factor, background and noise levels, GAPDH 3'/ 5' ratios and the %Present call.

The scaling factor is the multiplication factor applied to each signal value on an array. A scaling factor of 1.0 indicates that the average array intensity is equal to the target intensity. Scaling factors vary across different samples and so there are no set guidelines for any particular sample type. However, Affymetrix advise that if scaling factors differ by too much (i.e. ≥ 3 fold) within a set of experiments, that this may indicate wide variation in the .dat files and that the analysed data (in the .chp file) should be treated with caution.

A high background implies that impurities, such as cell debris and salts, are binding to the probe array in a non-specific manner, and that these substances are fluorescing at 570 nm (the detection wavelength). This non-specific binding causes a low signal-to-noise ratio (SNR), meaning that transcripts present at very low levels in the sample may be incorrectly called as Absent. High background creates an overall loss of sensitivity in the experiment. The noise is a measure of the pixel-to-pixel variation of probe cells on a GeneChip array.

2.13.7 Microarray Data Normalisation

The purpose of data normalisation is to minimise the effects of experimental and technical variation between microarray experiments so that meaningful biological comparisons can be drawn from the data sets and that real biological changes can be identified among multiple microarray experiments. Several approaches have been demonstrated to be effective and beneficial.

2.13.7.1 Genespring software

For the majority of experiments where a relatively small subset of transcripts is changing, a global method of scaling/normalisation is recommended. In such cases, since the majority of transcripts are not changing among samples, the overall intensities of the arrays should be similar. Differences in overall intensity are most likely due to assay variables including pipetting error, hybridisation, washing, and staining efficiencies, which are all independent of relative transcript concentration. Applying aglobal method of normalisation corrects for these variables. For global scaling, it is imperative that the same Target Intensity value is applied to all arrays being compared.

For some experiments where a relatively large subset of transcripts is affected, the "Selected Probe Sets" method of scaling/normalisation is recommended. The global approach does not make sense in this situation since the overall intensities among arrays are no longer comparable. In this case differences in overall intensity are due to biological and/or environmental conditions. Applying a global method skews the relative transcript concentrations. One option is to apply the "Selected Probe Sets" method using the 100 Normalisation Control probe sets, which are available for the major catalogue arrays. For replicates and comparisons involving a relatively small number of changes, the scaling/normalisation factors (calculated by the global method) should be comparable

among arrays. Larger discrepancies among scaling/normalisation factors (e.g., three-fold or greater) may indicate significant assay variability or sample degradation, leading to "noisier" data. Scaling/normalisation factors calculated by the "Selected Probe Sets" method should also be equivalent for arrays being compared. Larger discrepancies between scaling/normalisation factors may indicate significant assay or biological variability or degradation of the transcripts used for scaling/normalisation, which leads to noisier data.

2.13.7.2 dChip software

DNA-Chip Analyzer (dChip) is a software package implementing model-based expression analysis of oligonucleotide arrays and several high-level analysis procedures (Li and Wong 2001(a) and (b)). This model-based approach allows probe-level analysis on multiple arrays. By pooling information across multiple arrays, it is possible to assess standard errors for the expression indexes. This approach also allows automatic probe selection in the analysis stage to reduce errors due to cross-hybridising probes and image contamination.

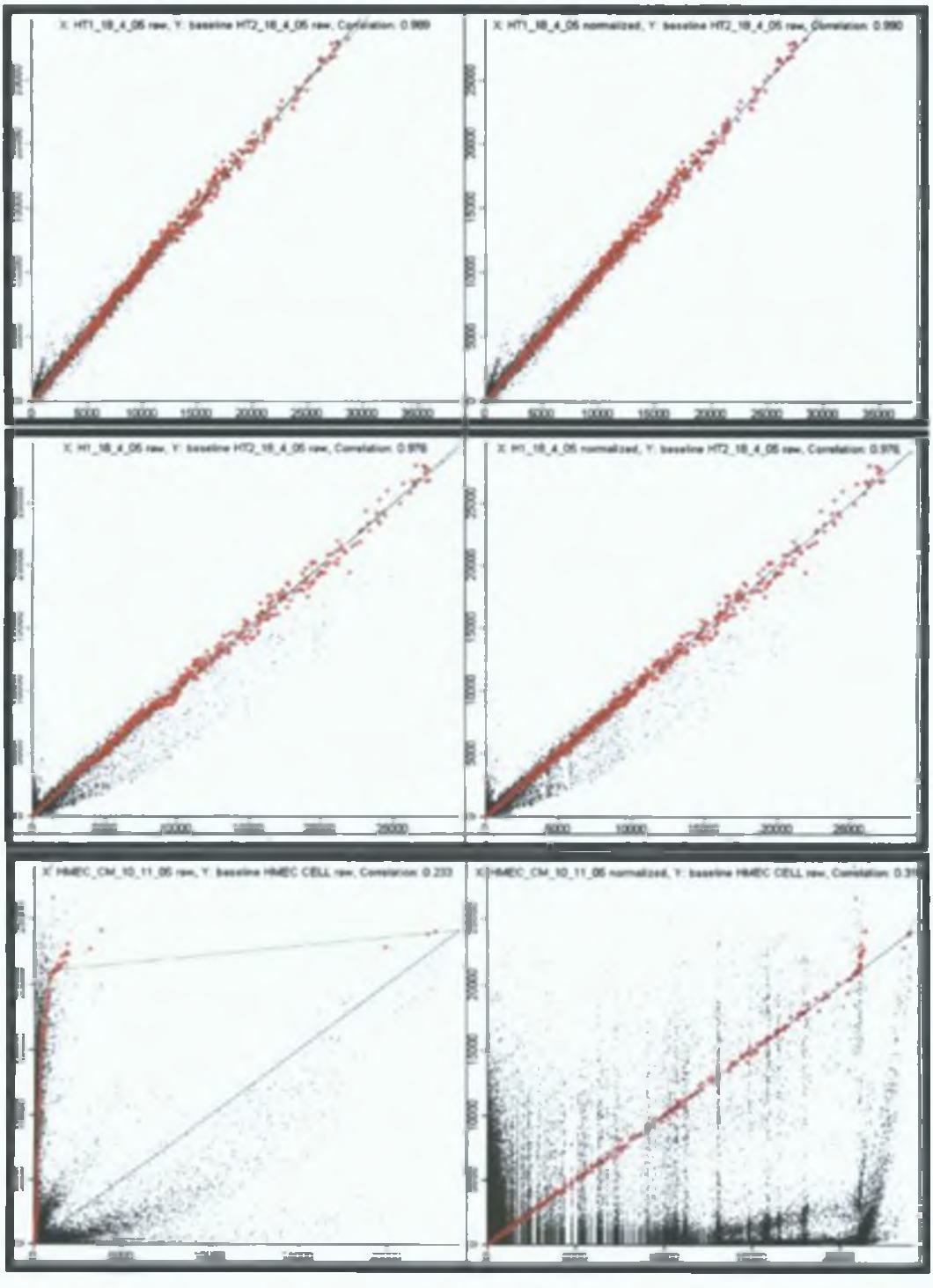
Since scanned images may have different overall brightness, generally normalisation is needed to adjust the brightness of the arrays to comparable level. dChip software needs to normalise arrays at PM and MM probe level before computing model-based expression levels. This is because the MAS algorithm analyses one array at a time (thus the Signal or Average Difference of different arrays can be scaled or normalised after computing them); while model-based method assumes the arrays are already at comparable brightness.

By default for dChip analysis, an array with median overall intensity is chosen as the baseline array against which other arrays are normalised at the probe intensity level. The software allows a different array to be chosen as the baseline, if by checking array images or outlier image it is found that the default baseline array has problems (due to

contamination or bad hybridisation).

dChip uses invariant set normalisation to normalise the arrays. This is carried out by choosing a subset of PM probes with small within-subset rank difference in the two arrays, to serve as the basis for fitting a normalisation curve. The fitted curve is the running median curve in the scatterplot of probe intensities of two arrays (with the baseline array on Y-axis and the array to be normalised on X-axis). When fitting the running median curve at the two ends, 5% of the "invariant" points are used to fix data at one end to make the high-end normalisation relationship more smooth and robust. The final running median curve is a piece-wise linear curve.

The normalisation transformation is then carried out for all the points (probes) in the array on X-axis (Y-axis is the baseline array and is not changed). The normalised probe values beyond the range of expression values [0, 65535] are truncated at the boundaries. The total effect of normalisation is a rotation and straightening of the scatter plot, so it better centres on the diagonal line y = x (see Fig. 2.13.7.2.1 for example of data before and after normalisation). It has been observed that when an experiment contains various tissue types, normalising all arrays together may reduce the similarity among the same tissue type at the expression index level. The likely cause is the wide scattering when plotting data for two tissues against each other and there are not enough points centering on the diagonal line to form a convincing normalisation curve (especially at high-end), thus the normalisation between different tissues may not work well and may introduce additional variation. If, by looking at such normalisation plots, the scattering is large between different tissues or conditions (e.g. correlation <0.5), tissue type-/ sample- specific normalisation can be carried out separately and the data pooled at a later stage.



(a)

(b)

(c)

Figure 2.13.7.2.1 Examples of normalisation curves obtained using dChip. (a) Replicate CM samples, (b) Parent vs. drug selected variant CM samples and (c) CM vs. cell RNA samples. Each dot represents PM or MM probe values in the two arrays (for a, b and c respectively). The blue line is the diagonal line y=x, the red circles are the probes selected in the "Invariant set", and the green curve is the running median normalisation curve based on the "Invariant set".

Usually the deviation of the blue line and the green curve indicates the need for normalisation (that is, one array is brighter than the other). Fig. 2.13.7.2.1 (a) is an example of two replica samples before and after normalisation (correlation score of 0.989 before normalisation and 0.99 after normalisation). Fig 2.13.7.2.1 (b) shows plots before and after normalisation of two different samples (correlation score 0.976 before and after normalisation). Fig. 2.13.7.2.1 (c) shows the plots before and after normalisation of two different sample types, in this case a CM sample and a cell sample. The correlation scores of 0.23 and 0.31 before and after normalisation, respectively, shows that these two samples cannot be normalised together; rather, they need to be normalised in separate groups and pooling of the data, at a later stage. This step also requires further scaling of values and application of a new MBEI if samples are to be compared against each other for changes in expression.

2.13.8 Microarray data visualisation/ manipulation

2.13.8.1 Genespring

The Genespring software allows visualisation, organise and manipulation of gene expression data. GeneSpring provides a host of tools to ask detailed questions about complex data sets. Sixteen transformations are available for creating powerful and flexible normalisation scenarios. Normalisation steps can be applied in virtually any order and include operations such as dye swapping experiments and median polishing. Scenarios can be saved and applied in other experiments. GeneSpring offers visually intuitive filtering tools for both entry-level and advanced users. All visual filtering windows generate graphs of results in real-time. These filters allow researchers to exclude particular conditions, set minimum and maximum values and choose specific gene lists to filter. The advanced

filtering window allows creation complex Boolean expressions to identify gene transcripts with a highly specific expression pattern. It also provides various analysis tools, such as t-tests, 2-way ANOVA tests and 1-way post-hoc tests for reliably identifying differentially expressed gene transcripts.

GeneSpring also has class prediction tools that can identify gene transcripts capable of discriminating between one or more experimental parameters or sample phenotypes. Groups of gene transcripts identified by expression profiling can be further characterised by performing sequence searches for potential regulatory elements. GeneSpring provides sophisticated clustering methods to uncover patterns of gene expression data and the relationships between these patterns. Researchers can use one, or a combination of, clustering option(s) to characterise their data: gene trees (hierarchical clustering), experiment trees, self-organising maps, k-means, Principal Components Analysis (PCA) and QT clustering. (QT clustering is an unsupervised technique that allows specification of both the minimum size and maximum correlation coefficient of each cluster). Principal Components Analysis (PCA) allows reduction of the complexity of the data by discovering a number of principal components that define most of the data variability.

The Genespring software was used in this study to plot the data from the replicate samples onto an X, Y, and Z 3-D scatter plot (or X, Y scatter plot if only 2 replicate samples available). Each probe set is represented by a dot and its placement on the scatter plot corresponds to the expression level of the probe set in each of the three samples (see Fig. 3.7.3.1, Section 3.7.3 for example). These "dots" representing individual probe sets can be coloured according to assigned criteria allowing the data to be quickly visually interpreted. The first step carried out to further enhance the detail of the scatter plot was to create an associated Venn diagram. The exact numbers of probe sets called Present (P) or Marginal (M) (see section 2.13.6 for explanation of detection call) in each sample is overlapped with

the other replicate samples. Any probe sets not called P or M are coloured grey (these would be assigned as Absent (A) in all three arrays) and the number of probe sets in this category is written outside the Venn diagram.

Hierarchical clustering was also performed on the samples in this study using a Pearson correlation to interpret how similar the samples are to each other.

2.13.8.2 dChip software

As well as normalisation of data dChip can be used to compare gene expression profiles of different sample groups to identify gene transcripts that are reliably differentially expressed between the two groups. When a group contains multiple samples, the group mean and its standard error is computed by pooling arrays considering measurement accuracy. Different filter criteria can be employed when comparing two groups. They are fold change, absolute difference between the two group means and p-value. The default p-value threshold of 0.05 filters gene transcripts that differ in group means with a two-tailed p-value<0.05. Other comparison criterion requires that the percentage of samples called Present in the samples used in both groups be larger than a threshold and finally there is another optional criterion that specifies the p-value for a paired t-test. This option assumes the first baseline sample is to be compared with the first experiment sample and the second baseline sample is to be compared with the second experiment *etc.* This paired t-test has larger power that the unpaired t-test. The results for the sample comparisons are exported into tab delimited files and can be easily viewed using Microsoft Excel.

Since tens of thousands of gene transcripts are compared, many gene transcripts can be false positives. However, gene transcripts are not all independent and gene transcripts in the same pathway could have similar t-statistics or p-values. In dChip, the empirical false discovery rate (FDR) can be estimated by permutation. The option "Analysis/Compare

samples/Combine comparison/Permute samples to assess False Discovery Rate" can be chosen to assess how many gene transcripts would be obtained by the same comparison criteria when permuting group labels randomly. It is usually good to try different comparison criteria to finally obtain a list of gene transcripts with median FDR <5% or 10%.

The gene lists can also be used for combined comparisons, for example if you compare sample groups A and B to obtain a gene list 1, and compare sample groups C and D to get a gene list 2, there is a provision in dChip to find the gene transcripts that are common on both lists or gene transcripts that are in one list but not the other.

When a list of gene transcripts is obtained by comparing the samples, dChip has a "classify gene" option which classifies the gene transcripts in the list into different groups according to GeneOntology or other annotational terms. DChip has gene information files (obtained from Affymetrix) which enables annotations to be attached to certain probe sets, the files include 1200 Gene Ontology terms, 3 Protein Domain terms, 160 Pathway terms and 257 Chromosome terms. The gene ontology terms range from reproduction to positive regulation of lyase activity. P- values are calculated for the groups to identify any significantly overrepresented groups, for example if a certain number of gene transcripts were randomly selected, out of all the annotated gene transcripts, what is the significance if a large number of them have the same gene ontology?.

2.13.8.3 Netaffx

The NetAffx- Analysis Centre is a comprehensive resource of integrated array contents and functional annotations available through the Affymetrix website. The flexible query capabilities provided help to retrieve biological information for specific probe sets. The software is freely available to users *via* registration through the Affymetrix Netaffx

website:

http://www.affymetrix.com/analysis/index.affx

Affymetrix also have a Gene Ontology (GO) tool. Gene Ontology is widely accepted as the standard for vocabulary describing the biological process, molecular function, and cellular component for gene transcripts (Ashburner *et al.* 2000). The Gene Ontology Mining Tool maps GeneChip probe sets to these hierarchical vocabularies. In addition to providing the GO terms for annotated gene transcripts, it provides graphical, interactive views of probe set representation within the biological process, molecular function, or cellular component hierarchies. The graphs allow the user to visualise input probe lists in the context of biological information and to visually determine the relationships among probe sets, based on their locations in the GO graph, which aids in the biological interpretation of a complex set of results.

As the graph is hierarchical, users can view their results within a context of broad or detailed GO categories. By clicking on a specific GO term in the output graph, a list of probe sets with annotations at or downstream to this term will be retrieved. This functionality allows the partitioning of probe sets based on the molecular function, biological process or cellular component of gene transcripts.

The original or reference category can be thought of as the parent category. The categories branching to the right directly from the reference category are the child categories. For example, in Figure 2.13.11, the parent category 8371 obsolete contains 16 gene transcripts and has 5 associated children categories: 6832 small molecule transport, 9315 drug resistance, 9406 virulence, 7012 actin, and 7048 oncogenesis. GO categories become more detailed as you move from parent to child, or from left to right.

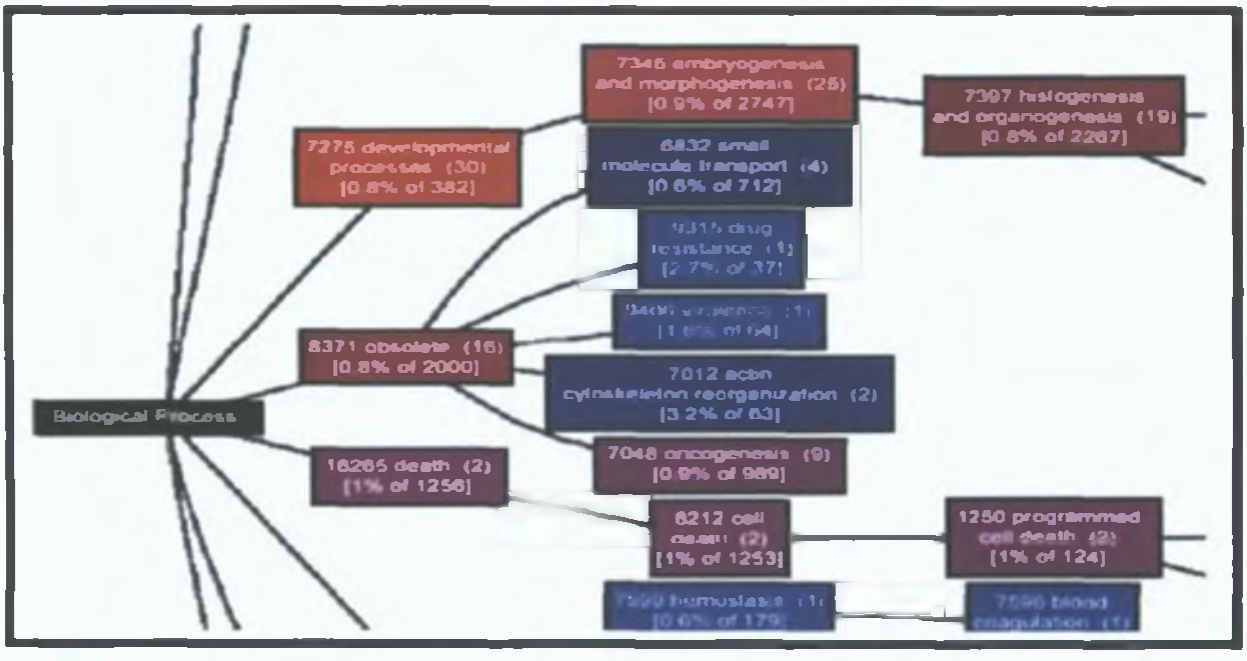


Figure 2.13.8.2.1 Example screenshot of Netaffx GO maps. The parent category in the biological Process map, 8371 obsolete, contains 16 gene transcripts and has 5 associated children categories: 6832 small molecule transport, 9315 drug resistance, 9406 virulence, 7012 actin, and 7048 oncogenesis. GO categories become more detailed as you move from parent to child, or from left to right. (from Affymetrix Netaffx Analysis Centre User's Guide, Affymetrix P/N 701757 Rev. 1)

2.13.8.4 Onto-Express

Intelligent Systems and Bioinformatics Laboratory, Computer Science Department, Wayne State University developed Onto-Express (OE) as a novel tool able to automatically translate lists of gene transcripts from microarray experiments of differentially regulated gene transcripts into functional profiles characterising the impact of the condition studied. The software is freely available to academic users from http://vortex.cs.wayne.edu:8080/ontoexpress/servlet/UserInfo. OE constructs functional profiles (using Gene Ontology terms) for the categories: biochemical function, biological process, cellular role, cellular component, molecular function and chromosome location. Statistical significance values are calculated for each category.

Essentially, the software works on the same principles as netaffx software (see section 2.13.9.3) but instead of generating a map of parent-child category relationships, the Onto-Express software generates detailed pie-charts, with one category per segment. It uses the detailed child categories to generate the pie-charts so, depending on the input list; it can potentially become quite detailed. One possible limitation of the software is that if a gene has two or more known molecular functions, it will be listed twice, or more, so that an input list of for example 100 probe sets, may generate a pie chart where the individual numbers of gene transcripts in each segment add up to many more than 100 gene transcripts. In general though, both Netaffx and Onto- Express can help to visualise a list of gene transcripts/ probe sets in a biologically-relevant context.

2.13.9 Sequence analysis

The UTRdb is a non redundant database of 5' and 3' UTR sequences generated by a computer program through the scanning of EMBL/ GenBank database entries. A software program called CleanUP removes redundant sequences. These redundant sequences are removed if they are the shorter of two matching sequences that have sequence identity and overlapping percentage above 95 and 90% respectively.

Also included in the UTRdb entries is information regarding the number of exons spanning the UTR gene region, whether the UTR region is complete or incomplete and the presence of repeat elements. In addition to this, is information regarding whether or not functional activity has been experimentally proven. The information is cross-linked with the Accession Number of the primary EMBL entry as well as between 5'UTR and 3'UTR sequences corresponding to the same region.

Information regarding a description of the biological role of the corresponding element, the

pattern consensus structure and the related bibliography are found in the connected UTRsite.

The part of the UTR database used for work presented in this thesis is the UTRscan program. The UTRscan program scans sequence data submitted to it for functional elements defined in the UTRsite collection. The program is internet based and is free to use after registering with the database. Sequences are submitted in Fasta format and the results of the scan are emailed back to the submitter detailing any regulatory elements found and the exact position and sequence of the elements are listed also.

Descriptions for regulatory elements (as per database) found in sequences submitted to UTRscan during the course of this work are listed and described here.

2.13.9.1 15-Lipoxygenase Differentiation Control Element, (15-LOX DICE)

The 15-LOX DICE regulatory sequence can be found in both 5' and 3' UTR regions. Erythroid 15-lipoxygenase (15-LOX) synthesised in the early stages of erytropoiesis catalyses the degradation of lipids and is an important factor responsible for the degradation of mitochondria during reticulocyte maturation. The 3'UTR of 15-LOX mRNA contains tandem repeat sequences, referred to as "DICE" for "Differentiation Control Element" which specifically bind to regulatory proteins, formerly called LOX-BP. This binding specifically inhibits 15-LOX mRNA translation. In rabbit 15-LOX mRNA, 10 tandem copies of the repeat sequence have been found, but only 4 (3 tandem) in human. Experimentally, a functional DICE consisting of two repeat elements has been described. The regulatory binding proteins have been identified as hnRNP K and hnRNP E1.

Many Notch pathway target gene transcripts in Drosophila, including most members of the basic helix-loop-helix repressor family and the Bearded family, contain a series of conserved sequence motifs in their 3' UTRs, including the Brd box, the K box, and the GY box.

2.13.9.2 Brd-Box (Brd)

The Brd box (pattern AGCTTTA) is present in one or more copies in many 3'UTRs and mediates negative post-transcriptional regulation by affecting transcript stability and translational efficiency. Regulation by the Brd box is spatially and temporally ubiquitous, and likely involves the formation of RNA-RNA duplexes with complementary sequences found at the 5'end of certain micro RNAs (miRNAs). Brd box-complementary miRNAs are also found in other nematodes, suggestive of functional conservation.

2.13.9.3 K-Box (KB)

The K box (pattern cTGTGATa) is present in one or more copies in many 3'UTRs and mediates negative post-transcriptional regulation, mainly effected by decreased transcript levels. Regulation by the K box is spatially and temporally ubiquitous, and likely involves the formation of RNA-RNA duplexes with complementary sequences found at the 5'end of many micro RNAs (miRNAs). K box-complementary miRNAs are found in other metazoa, suggestive of functional conservation.

2.13.9.4 GY-Box (GY)

The GY box (GTCTTCC) is present in one or more copies in several of these 3' UTRs, but does not appear to have strong cis-regulatory activity.

However, its function likely involves the formation of RNA duplexes with a complementary sequence found in the 3' UTRs of proneural basic helix-loop-helix gene

transcripts (the proneural box, ATGGAAGACAAT) and with complementary sequences found at the 5' ends of certain micro RNAs (miRNAs).

2.13.9.5 Alcohol dehydrogenase 3'UTR downregulation control element (ADH_DRE)

A highly conserved eight base sequence (AAGGCTGA) within the 3' untranslated region of the Drosophila Alcohol dehydrogenase (Adh) mRNA, identified by phylogenetic analysis, was found to produce a two-fold downregulation of Adh gene expression. The experimental assay was carried out by in vitro deletion mutagenesis followed by P element mediated germline transformation. Downregulation was found to be the result of a temporally and spatially generated decrease in the Adh transcript possibly resulting by either a reduction of transcription rate or most likely by a reduction of the mRNA stability.

2.13.9.6 Cytoplasmic polyadenylation element (CPE)

Cytoplasmic polyadenylation is an evolutionarily conserved mechanism regulating translational activation of a set of quiescent maternal messenger RNAs (mRNAs) during early development. Cytoplasmic poly (A) elongation occurs in a wide range of species, ranging from clam to mouse. The relevance of this process during late oogenesis and early embryogenesis has been shown by studies in the mouse, Xenopus, and Drosophila. In mouse and Xenopus, the only vertebrates so far examined, the critical regulatory sequences, referred to as Cytoplasmic Polyadenylation Elements (CPEs), are AU-rich and located in the 3' untranslated region (3'-UTR) near the canonical nuclear polyadenylation element (AAUAAA), which also is required for proper poly (A) addition. The CPE has the general structure of UUUUUUAU. However, the CPE is not identical in all mRNAs and its position varies relative to AAUAAA (generally within 100 nucleotides). The minimal CPE capable to stimulate elongation of a poly (A) tail appears to be UUUUAU, and recent experiments show also the existence of substantial context and position effects on CPE

function.

2.13.9.7 Internal Ribosome Entry Site (IRES)

Internal mRNA ribosome binding is a mechanism of translation initiation alternative to the conventional 5'-cap dependent ribosome scanning mechanism. The internal initiation mechanism, first observed in picornavirus, is also found in some cellular mRNAs, such as human immunoglobulin heavy chain binding protein (BiP) mRNA, Antennapedia mRNA of drosophila and FGF-2 mRNA. This mechanism may be advantageous for specific translation of mRNAs during cell cycle when underphosphorilated eIF-4F prevents conventional translation initiation. The internal initiation mechanism is dependent on a cis-acting element, defined as internal ribosome entry site (IRES), located in the 5'UTR region of the mRNA.

3 RESULTS

The aims of Section 3.1, 3.2 and 3.3 were:

1. To establish if it is possible to routinely extract and amplify extracellular mRNA from the conditioned media of cultured human cancer cells.
2. To compare different methods of extracting the RNA and determine which method, if any provides both reliable and reproducible results.
3. To establish if the mRNA is present in full length transcripts

3.1 Identifying suitable time point for CM collection and analysis

In order to establish an appropriate point for sampling CM, cells were cultured over 96 hours with CM samples removed from the media every 24 hours. See Ssction 2.7.1 for details of the experimental design of the study.

After samples were collected and stored at -80°C until required, RNA was extracted from the samples and reverse- transcribed (see section 2.10). PCR was then carried out for β-actin. Fig 3.1.1 shows the expression of β-actin in the conditioned media over the 96 hours. β-actin mRNA was detected at all time points sampled indicating that any time point over the 96 hours could be selected. The next step was to look at the cells under the microscope over the 96 hours to check the viability of the cells and make sure that the mRNA detected in the CM was not due to lysis of cells resulting from an over-confluent population.

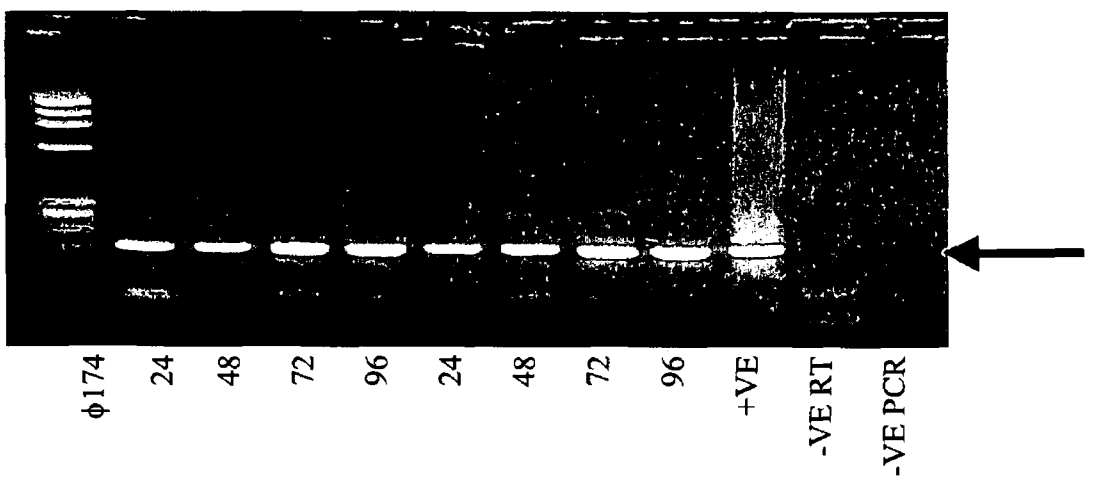


Figure 3.1.1 Expression of β-actin at time points over 96 hours in RPMI ML CM. β-actin gene transcript (arrow) was detected in RNA isolated from all CM samples after passing through a 0.22 μm filter. Amplified product was detected in the positive control, RPMI TAX cell RNA, and was undetected in the negative control samples (-VE RT = RT reaction with H_20 instead of RNA as control; -VE PCR = PCR reaction with H_20 instead of cDNA as control).

At each time point analysed, the cells were viewed under the microscope before every CM sample was taken to check for large amounts of cells/ cell debris floating in the media. As an example, DLKP cells after 48 hours in culture are shown in Fig 3.1.2. These cells were photographed to show that any mRNA isolated from the CM was not from lysis of dead cells due to over- confluent cell populations. Figure 3.1.2 shows that after 48 hours the cells are not confluent and look healthy and viable with little/ no dead cells floating in the media. At 72 but, more so at 96 hours, the confluency of the cells was more than 80% and much more cell debris could be seen floating in the medium.

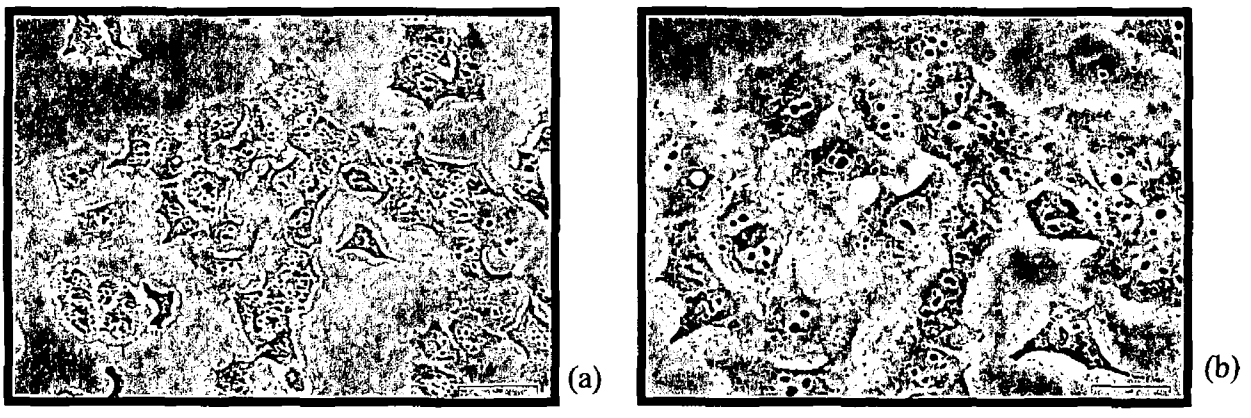


Figure 3.1.2 DLKP cells in culture for 48 hours (a) 20X and (b) 40X. The size bars in the bottom right-hand side of the photographs correspond to 100µm (a) and 50µm (b).

3.2 Determining a reproducible minimum detection level

β-actin was amplifiable at all time points, i.e. 24- 96 hours using RNA extracted with Tri Reagent (see Section 3.1). In order to ensure that other less abundant transcripts were also amplifiable by RT-PCR, we analysed the CM RNA for other transcripts including CK-19, HnRNP B1 and MDR1. The reason for this choice was that both CK-19 and HnRNPB1 gene transcripts were found in the serum of lung cancer patients in a study by Fleischhacker *et al* (2001) and MDR1 is a P-glycoprotein that has a role in multiple drug resistance. As RPMI 2650 Taxol (RPMI TX) are known to overexpress MRP1 mRNA with no expression

in RPMI 2650 Melphalan (RPMI ML), this transcript was also chosen (Liang *et al* 2001).

3.2.1 RNA Extracted using Promega SV Total Isolation Kit

Flasks of RPMI TAX and RPMI ML cells were set up in duplicate. 1ml volumes of CM was collected at 24 hour intervals over 96 hours and were filtered through 0.22μm filters as described in Section 3.1. RNA extraction was carried out on a 100μl volume of filtered CM, using the Promega SV Total Isolation kit (see Section 2.8.2.1 for details).

RT-PCRs were then performed on the duplicate samples to check for expression of CK-19. Fig 3.2.1.1 (a) and (b) show that the PCR results obtained using this RNA extracted from 100μl of CM using the Promega SV Total Isolation kit when trying to amplify the CK-19 gene transcripts were not reproducible. Duplicate samples showed little/ no concordance with each other. The different sampling time points used had no effect on the reproducibility.

CK-19

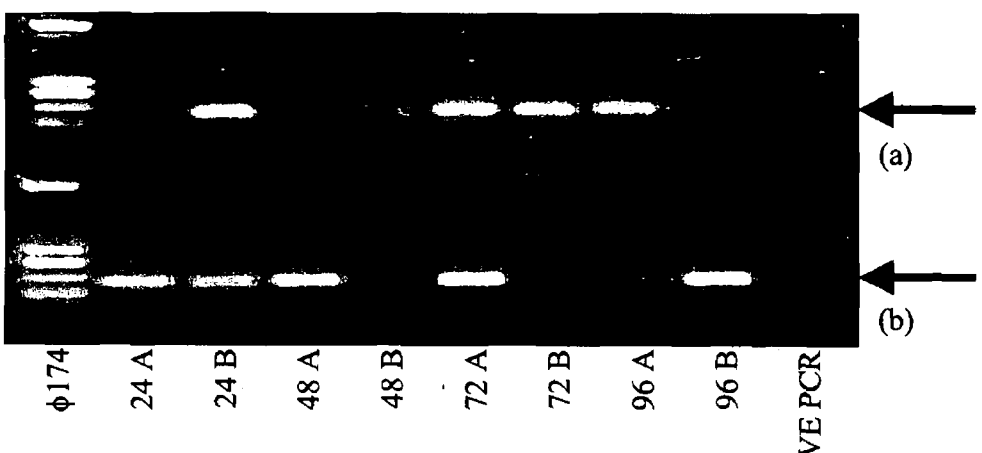


Figure 3.2.1.1 Expression of CK-19 in RPMI TAX (a) and RPMI ML (b) over 96 hours. CK–19 gene transcript (arrow) was not detected in all duplicate RNA samples (see 48A and B samples in Fig 3.2.1.1 (a) and (b) above). Amplified product was undetected in the negative control sample (-VE PCR = PCR reaction with H_20 instead of cDNA as control).

3.2.2 Comparison of RNA Extracted Using Tri Reagent with RNA extracted using the Promega SV Total Isolation system

As the duplicate samples analysed in Section 3.2.1 did not show similar results, it was decided to concentrate the mRNA available for RT-PCR by eluting extracted RNA into a smaller volume of RNase- free water. As 100µl was the minimum eluant possible for the Promega system, the extraction procedure could not be scaled- up to accommodate this without a significant increase in cost by having to buy in extra buffers for the kits, it was decided to try using Tri Reagent to extract the RNA.

Requirements for Tri Reagent extraction are that for every 250µl of solution sample, 750µl of Tri Reagent is added. As the RNA is precipitated using isopropanol, washed in ethanol, and then the resulting pellet is resuspended in any desired amount of RNase- free water, it meant that this protocol could be readily scaled-up to obtain greater yields of RNA.

Two CM sample sizes were chosen, 100µl (for comparison with Promega kit results) and 250µl (to allow RNA to be extracted from a larger amount of CM). The 100µl volume enabled the equivalent of 9µl of CM to be used in each RT-PCR reaction whilst the 250µl volume enabled the equivalent of 50µl of CM to be used in each RT-PCR, i.e. over five times more CM/ RT-PCR reaction (see Section 2.7.2 for experimental design of study). RNA extractions were carried out according to Section 2.8.2. The RNA was eluted using RNase-free water with the Promega system as it was supplied with the kit. UHP water was treated with DEPC (see Section 2.1.1) to remove/ destroy RNases and this was used to resuspend the RNA pellets precipitated using the Tri Reagent protocol.

The RT-PCR results using RNA extracted with Tri Reagent, in Figs. 3.2.2.1-2 and 3.2.2.4-5 seem to indicate a general increase in reproducibility compared to the RT-PCR results obtained using RNA extracted with the Promega system (see Fig. 3.2.1.1) especially with

RNA extracted from RPMI TAX CM. Although β-actin transcripts were amplifiable for all samples (Fig. 3.2.2.1), when RNA extracted from the smaller volume of CM was used in the RT reaction (9μl instead of 50μl), PCR results for both CK-19 and HnRNP B1 were conflicting in the RPMI ML duplicate samples (see Figs 3.2.2.2 (b) and 3.2.2.4 (b), 48.100 samples). This indicates the possibility that RNA extracted from the smaller volume of CM may be at the limit of detection for RT-PCR to detect it.

MDR1 transcripts were amplifiable in the RPMI TAX CM samples and not the RPMI ML CM samples indicating that the detection of gene transcripts in the CM of cultured cells may be a cell- specific process (Fig. 3.2.2.5).

β-actin

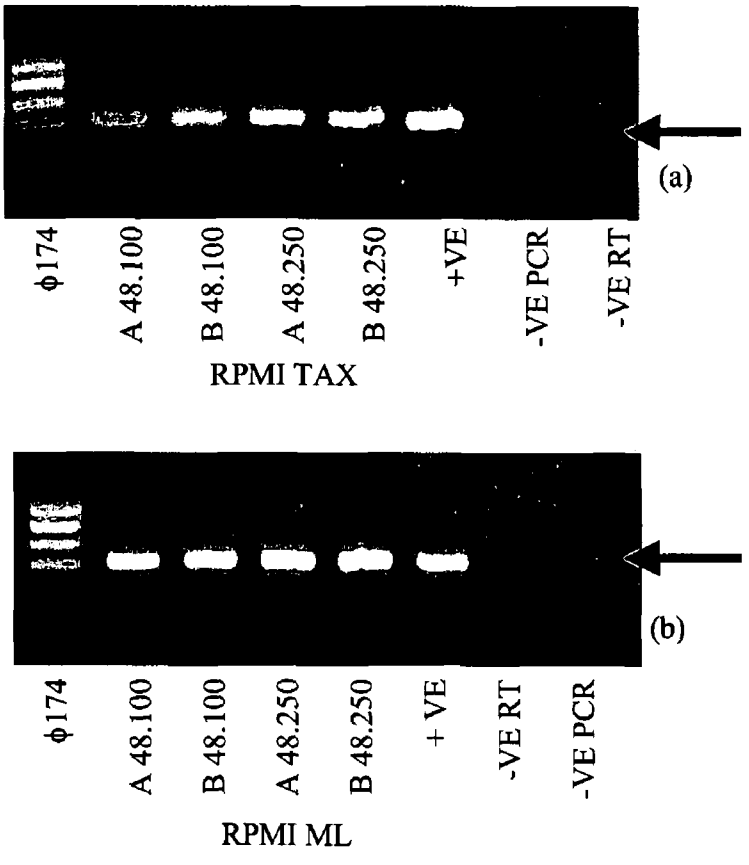


Figure 3.2.2.1 Amplification of β-actin (196bp) in RPMI TAX (a) and RPMI ML (b). 48.100 = RNA extracted from 100µl of CM conditioned for 48 hours (comparable to results obtained using Promega Total Isolation SV kit ≈9µl CM/ RT-PCR). 48.250 = RNA extracted from 250µl of medium conditioned for 48 hours (≈50µl CM/ RT-PCR). β-actin gene transcript (arrow) was detected in RNA isolated from all CM samples from RPMI TAX (a) and RPMI ML (b). Amplified product was detected in the positive control, RPMI cell RNA, and was undetected in the negative control samples (-ve RT = RT reaction with H_20 instead of RNA as control; -ve PCR = PCR reaction with H_20 instead of cDNA as control).

CK-19

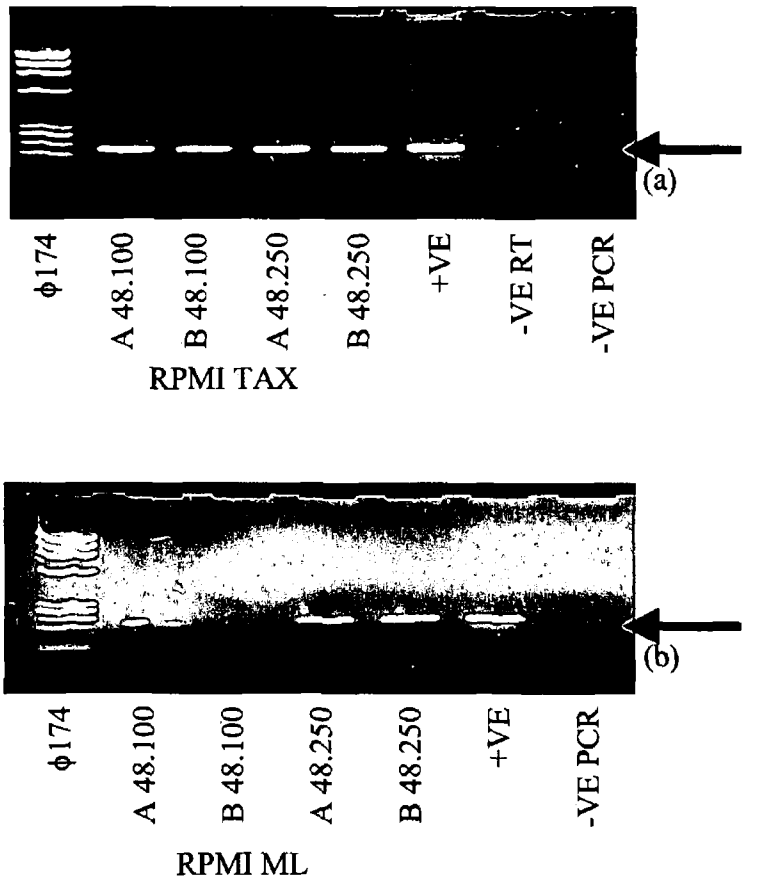


Figure 3.2.2.2 Amplification of CK-19 in RPMI TAX (a) and RPMI ML (b). 48.100 = RNA extracted from 100µl of CM conditioned for 48 hours (comparable to results obtained using Promega Total Isolation SV kit ≈9µl CM/ RT-PCR). 48.250 = RNA extracted from 250µl of CM conditioned for 48 hours (≈50µl CM/ RT-PCR). CK-19 gene transcript (arrow) was detected in RNA isolated from all CM samples from RPMI TAX CM (a) and in all RPMI ML CM samples except sample Ml B 48.100 (b). Amplified product was detected in the positive control, RPMI cell RNA, and was undetected in the negative control samples (-ve RT = RT reaction with H_20 instead of RNA as control; -ve PCR = PCR reaction with H_20 instead of cDNA as control).

HnRNP B1

Initially the PCR primers used to amplify the HnRNPB1 transcript were chosen from Fleischhacker *et al* (2001), but the primers were not RNA specific and PCR bands attributable to DNA contamination were detected (see Fig. 3.2.2.4). This showed a possible weakness with the Tri Reagent protocol in so far as DNA was present in the RNA suspension. To overcome this, the forward primer was redesigned so that it crossed an exon-intron boundary allowing only amplification of mRNA transcripts (see Fig. 3.2.2.5). All future PCR primer pairs were designed with this in mind (see Section 3.8.11.1 for more information).

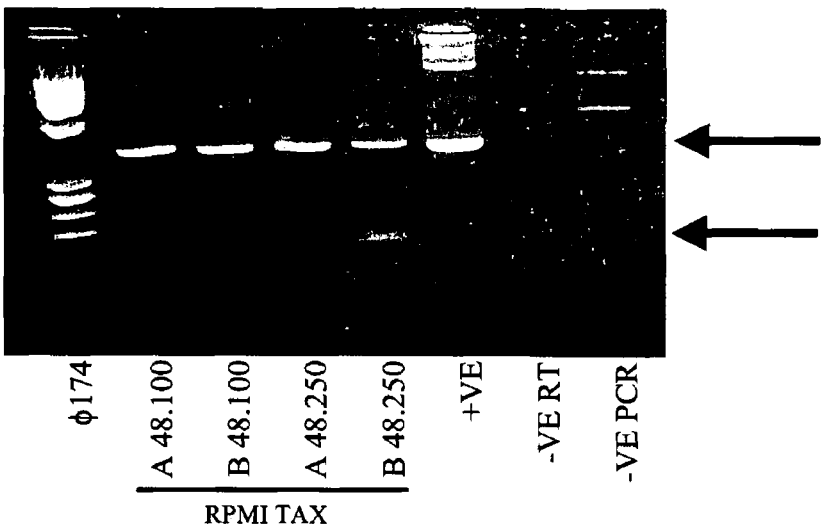


Figure 3.2.2.3 RT-PCR of RPMI TAX CM for HnRNP B1 (166bp). 48.100 = RNA extracted from 100µl of CM conditioned for 48 hours (comparable to results obtained using Promega Total Isolation SV kit ≈9µl CM/ RT-PCR). 48.250 = RNA extracted from 250µl of CM conditioned for 48 hours (≈50µl CM/ RT-PCR). HnRNP B1 mRNA gene transcript (black arrow) was detected in one CM sample (B. 48.250), whilst PCR products attributable to DNA contamination (red arrow, 420bp) were amplified in all samples. Amplified product was undetected in the negative control samples (-ve RT = RT reaction with H_20 instead of RNA as control; -ve PCR = PCR reaction with H_20 instead of cDNA as control).

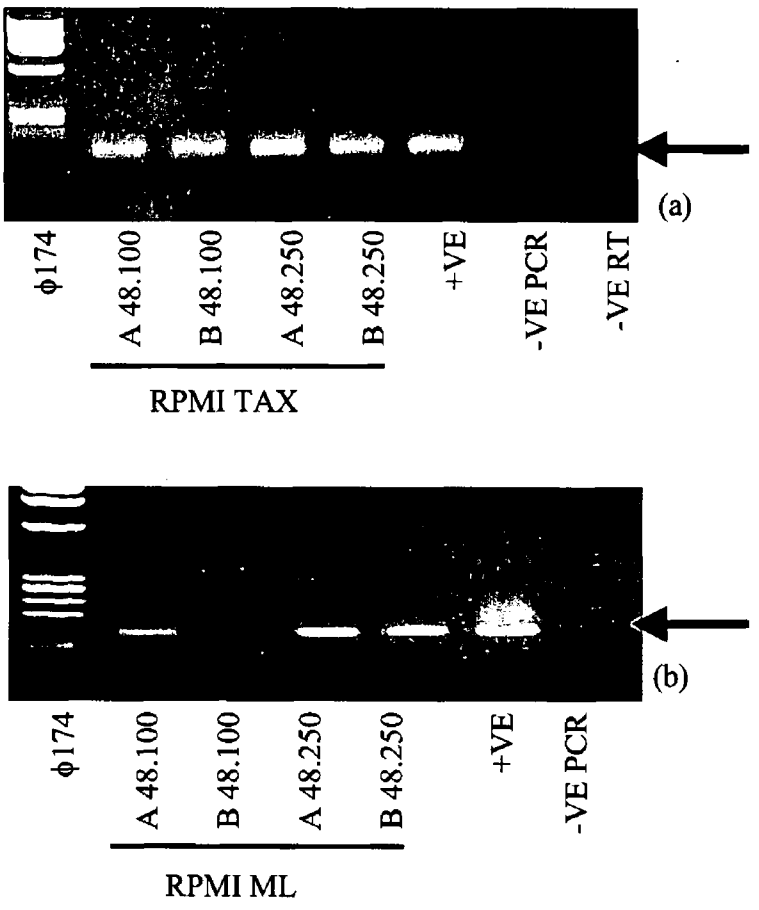


Figure 3.2.2.4 Amplification of HnRNP B1 in RPMI TAX (a) and RPMI ML (b). 48.100 = RNA extracted from 100µl of CM conditioned for 48 hours (comparable to results obtained using Promega Total Isolation SV kit ≈9µl CM/ RT-PCR). 48.250 = RNA extracted from 250µl of CM conditioned for 48 hours (≈50µl CM/ RT-PCR). HnRNP B1 gene transcript (arrow) was detected in RNA isolated from all CM samples from RPMI TAX CM (a) and RPMI ML CM except sample Ml B 48.100 (b). Amplified product was detected in the positive control, RPMI cell RNA, and was undetected in the negative control samples (-ve RT = RT reaction with H_20 instead of RNA as control; -ve PCR = PCR reaction with H_20 instead of cDNA as control).

MDR 1

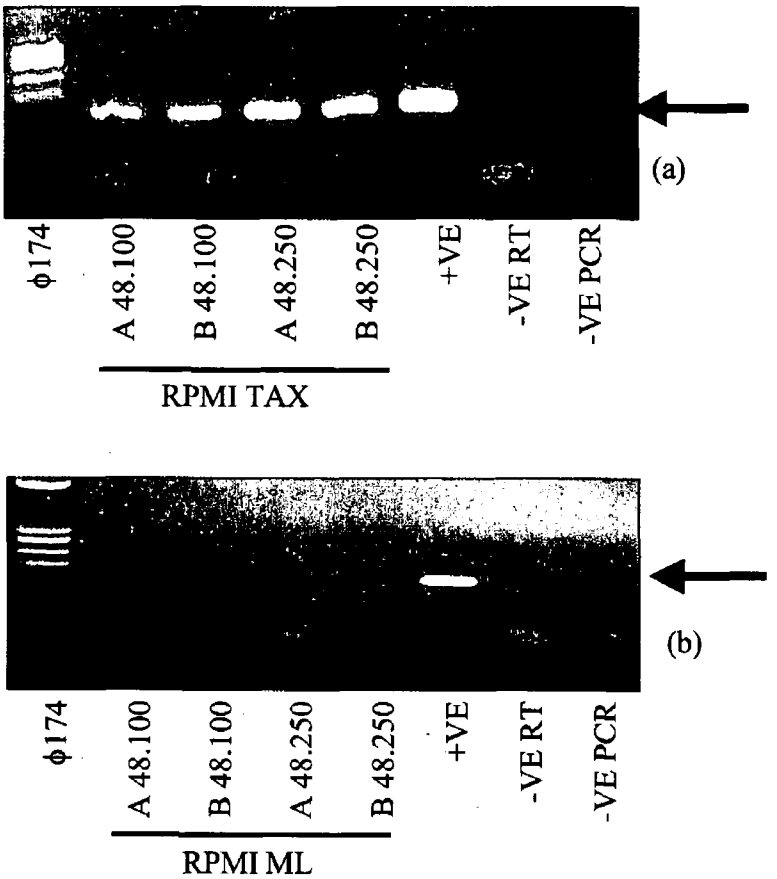


Figure 3.2.2.5 Amplification of MDR1 in RPMI TAX (a) and RPMI ML (b). 48.100 = RNA extracted from 100µl of CM conditioned for 48 hours (comparable to results obtained using Promega Total Isolation SV kit ≈9µl CM/ RT-PCR). 48.250 = RNA extracted from 250µl of CM conditioned for 48 hours (≈50µl CM/ RT-PCR). MDR 1 gene transcript (arrow) was detected in RNA isolated from all CM samples from RPMI TAX CM (a) and RPMI ML CM (b). Amplified product was detected in the positive control, RPMI cell RNA, and was undetected in the negative control samples (-ve RT = RT reaction with H_20 instead of RNA as control; -ve PCR = PCR reaction with H_20 instead of cDNA as control).

3.3 Amplification of β-actin gene using 5' and 3' specific primers

Previous studies of human serum/ plasma have suggested that the RNA detected in cell-free samples is probably present as short fragments (El- Hefnawy *et al* 2004). Morozkin *et al* (2004) suggested that extracellular RNA isolated from Hela and A431 growth medium was represented by 100-200 nucleotide long fragments. To investigate if full–length transcripts may be detected extracellularly, we designed primers that would anneal to different parts of the β-actin mRNA transcript (See Fig 3.3.1 for locations). The size of the amplified regions ranged from 142bp to 383bp. One of the sets of primers annealed to the 5' untranslated region (UTR) of the β-actin transcript, and, as the mRNA is reverse transcribed using only Oligo dT as the reverse primer, any amplification of this region would indicate that for β-actin at least, full-length mRNA transcripts are present. Fig 3.3.2 shows that all four regions were successfully amplified.

Figure 3.3.1 Location of amplified regions along β-actin gene transcript. 5' untranslated region is denoted by broken line (.....), the blue line indicates a region of 171bp that includes the 5'UTR region, the pink line represents a 196bp region that starts at approximately the 5' UTR region (sequence analysed by Fleischhaker *et al* 2001). The yellow and green lines represent regions of 383bp and 143bp, respectively. The reverse primer for these two amplified regions is the same.

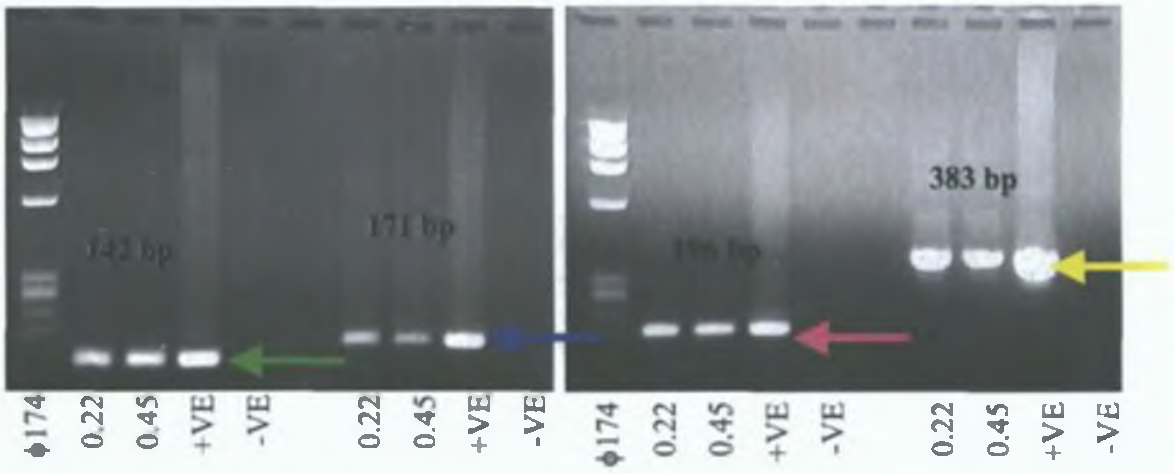


Figure 3.3.2 Amplified regions of β-actin gene using DLKP CM RNA. β-actin gene transcripts (the coloured arrows here correspond to the target regions shown in Fig 3.3.1) were detected in RNA isolated from CM after passing through a either a 0.22μm or 0.45μm filter. Amplified product was detected in the positive control, RPMI TAX cell RNA, and was undetected in the negative control samples (; -ve PCR = PCR reaction with H_20 instead of cDNA as control). The length of each amplified product is written above the PCR bands in black.

The RT-PCR results shown in Fig 3.3.2 indicate for β-actin, at least, the presence of full-length transcripts in conditioned media that has been filtered through either a 0.22μm or 0.45μm filter.

Having established that extracellular mRNA could be reliably and reproducibly amplified using a modified Tri Reagent protocol for RNA extraction, the aim of Section 3.4 was:

1. To investigate if any known tumour-related mRNAs already studied in out laboratory are expressed in a panel of human cancer cell line conditioned media samples
2. If tumour- related mRNAs are expressed, at what levels are they expressed?
3. To investigate if tumour- related mRNAs are expressed in all of the conditioned media samples corresponding to cells expressing the gene transcript.
4. To establish that the extracellular mRNA amplifiable from the conditioned media was not attributable in some way to the media that the cells are conditioned in.

3.4 Analysis of gene transcripts in RNA isolated from cell lines and conditioned media

Once it was established that we could successfully extract and amplify mRNA from CM using a reproducible protocol, developed during the course of this study and that full-length mRNA transcripts were present in the extracellular environment of cancer cells, a more comprehensive study was performed, to include other known tumour markers previously studied in cell lines in our laboratory. Two different filter sizes were used to filter the CM. Ng *et* al (2002) showed that filter size can have an effect on the amount of RNA extractable from samples, so here the CM was filtered with 0.22μm and 0.45μm filters, (both of which would exclude cells from the CM) to establish if there was any effect on the gene transcripts amplifiable. Seven target gene transcripts were chosen and primers were chosen that amplified different regions of the transcripts, i.e. some were targetedto the 5' end of the transcript (HnRNP B1, GST π); to the middle of the transcript (MRP 1, TOPO II, CK-19) and to the 3'end of the transcripts (Bcl-2, MDR 1). Fig 2.11.1.1 (Section 2.11.1) details the positions of the amplified regions of the mRNA transcripts selected for the study.

All transcripts underwent 45 PCR cycles at an annealing temperature of 54°C. RT-PCR was carried out on the corresponding cell RNA for the same transcripts, with co-amplification of β-actin as an internal control. 30 PCR cycles were found to be sufficient for cell line samples with annealing temperature of 54°C. Cell RNA was also used as a positive control for all CM PCR reactions.

Although β-actin was routinely amplified in all CM, attempts to co-amplify β-actin inhibited detection of other gene products of interest. For this reason, β-actin was not co-amplified with the cDNA, when analysing CM. To ensure that any negative result obtained for a tumour- specific target gene transcript was not a false negative i.e. as a result of

problems in the RT or PCR step, for every set of PCRs carried out using fresh cDNA, one aliquot was used to amplify the 5' UTR region of β-actin mRNA.

The β-actin transcript co-amplified with the target mRNA of interest in the cell RNA samples amplified a region in the 3' end of the mRNA transcript. The reason for not using the same set of primers for both CM and cell cDNAwas simply that the size of the PCR product for the 5'UTR targeted primers (171bp) was similar to the PCR product size for some of the target gene transcripts in the study, so to ensure clear separation of the PCR products when run on an agarose gel, β-actin primers with a larger amplicon size (383bp) were chosen for co-amplification with the cell RNA.

β-actin, CK-19 and HnRNP B1 gene transcripts were amplifiable for all cell lines and for their corresponding CM samples filtered through either a 0.22μm or 0.45μm filter (Figs. 3.4., 3.4.2 and 3.4.3 respectively).

Fig. 3.4.4 shows that GST π mRNA gene transcripts were detected in all cell line RNA except MCF-7 and BT474 cell lines. Detection of this mRNA in CM filtered through both size filters was found only in the CM samples that corresponded to the cell lines positive for expression of this transcript with the exception of the 0.45μm filtered CM taken from MCF-7 cultured cells. A faint band corresponding to GST π mRNA was present in this sample. This result was reproducibly found. Detection of Topoisomerase II and MRP1 transcripts were found in all cell line RNA, however differential expression of the transcript was detected in the corresponding CM samples (Fig. 3.4.5 and 3.4.6 respectively). Topoisomerase II mRNA gene transcripts were detected in 7/8 0.45μm filtered CM and in only 3/8 0.22μm filtered CM samples. MRP1 mRNA gene transcripts were detected in 5/8 of the 0.45μm filtered CM samples and in only 2/8 of the 0.22μm filtered CM samples. This indicates the possibility that these mRNA transcripts may be bound to proteins/ vesicles inhibited by filtration of the CM with a 0.22μm filter.

Bcl-2 was expressed (*albeit* at relatively low levels) in the entire 8 cell lines analysed (see Fig. 3.4.7 (a)). However, this gene transcript was apparently not secreted by these cells *i.e.* with the exception of a very weak band resulting following amplification of 0.45 μm filtered RPMI-TAX CM, bcl-2 mRNA was undetected in all CM.

MDR1 mRNA transcripts were detected in 3/8 of the cell lines analysed, with faint expression in a further 2/8 samples (Fig. 3.4.8). Detection of the mRNA in the corresponding CM samples however, was only reproducibly found in RPMI TAX CM samples.

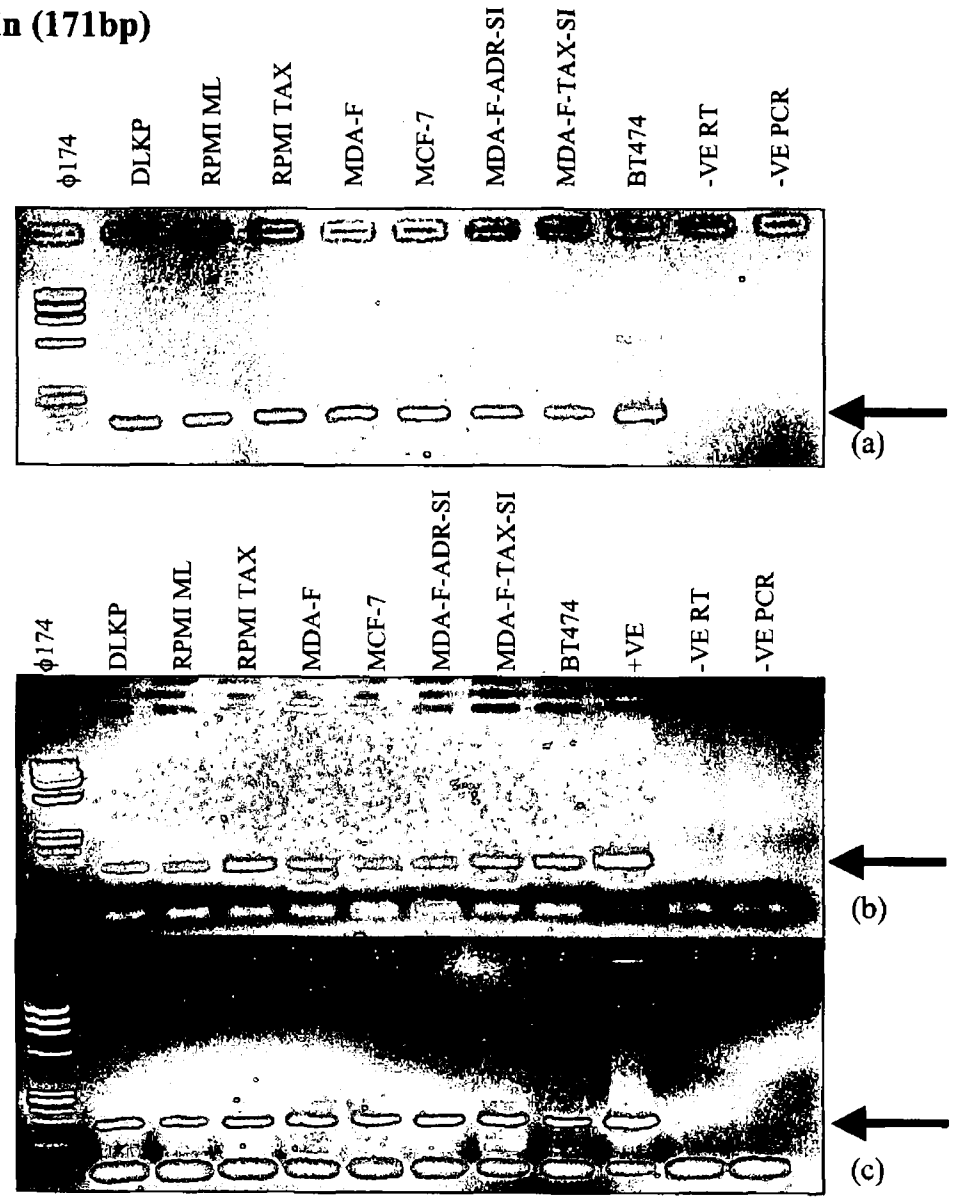


Figure 3.4.1 Expression of β-actin (171bp) in cells and filtered CM. β-actin gene transcript (arrow) was detected in RNA isolated from all cell lines analysed (a) and in corresponding CM after passing through a (b) 0.22 μm or (c) 0.45 μm filter. Amplified product was detected in the positive control and was undetected in the negative control samples (-ve RT = RT reaction with H_20 instead of RNA as control; -ve PCR = PCR reaction with H_20 instead of cDNA as control).

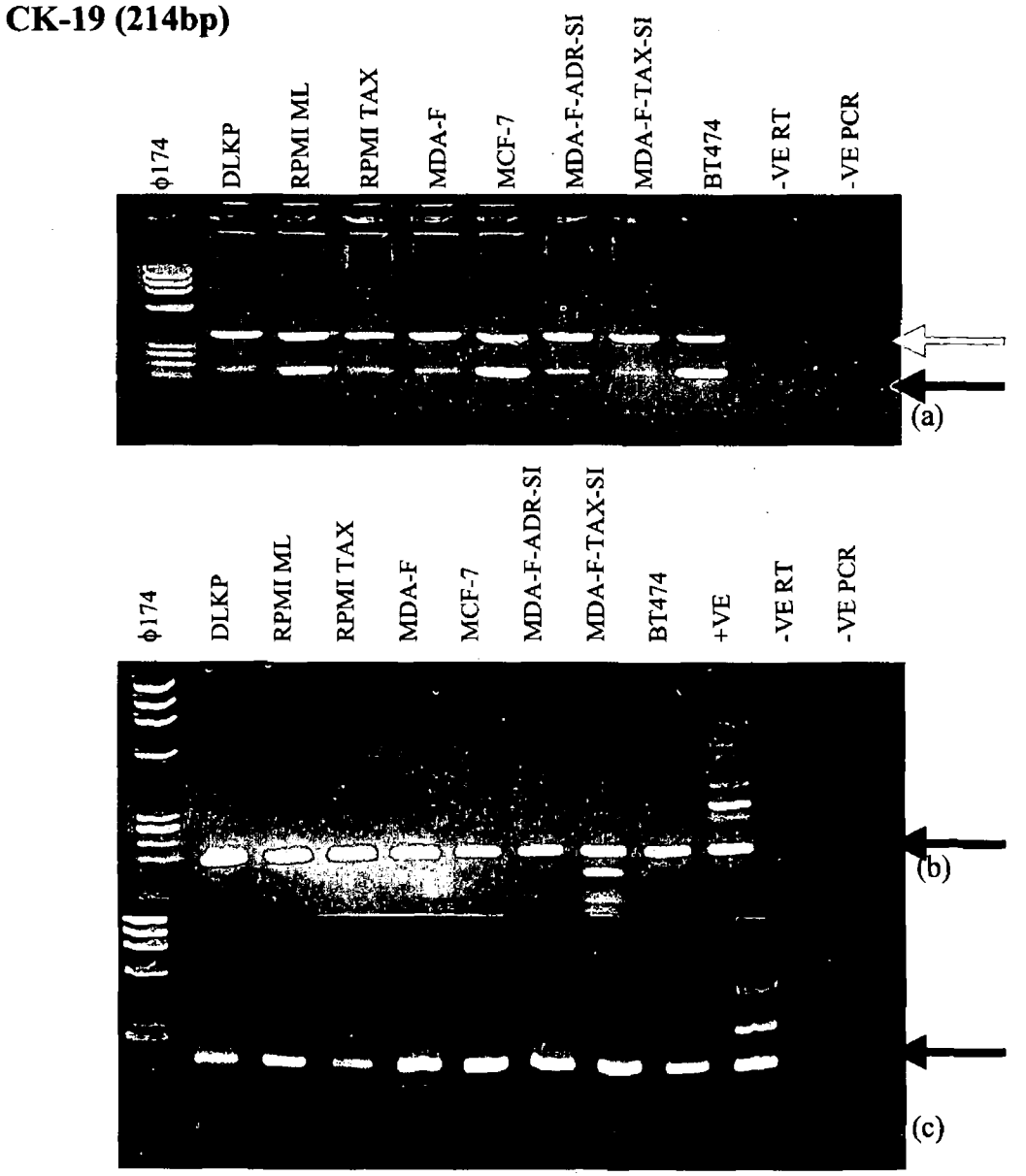


Figure 3.4.2 Expression of CK-19 in cells and filtered CM. CK-19 (black arrow) and β-actin (yellow arrow) gene transcripts were detected in RNA isolated from all cell lines analysed (a) and CK-19 gene transcripts in corresponding CM after passing through a (b) 0.22 μm or (c) 0.45 μm filter. Amplified product was detected in the positive control and was undetected in the negative control samples (-ve RT = RT reaction with H_20 instead of RNA as control; -ve PCR = PCR reaction with H_20 instead of cDNA as control).

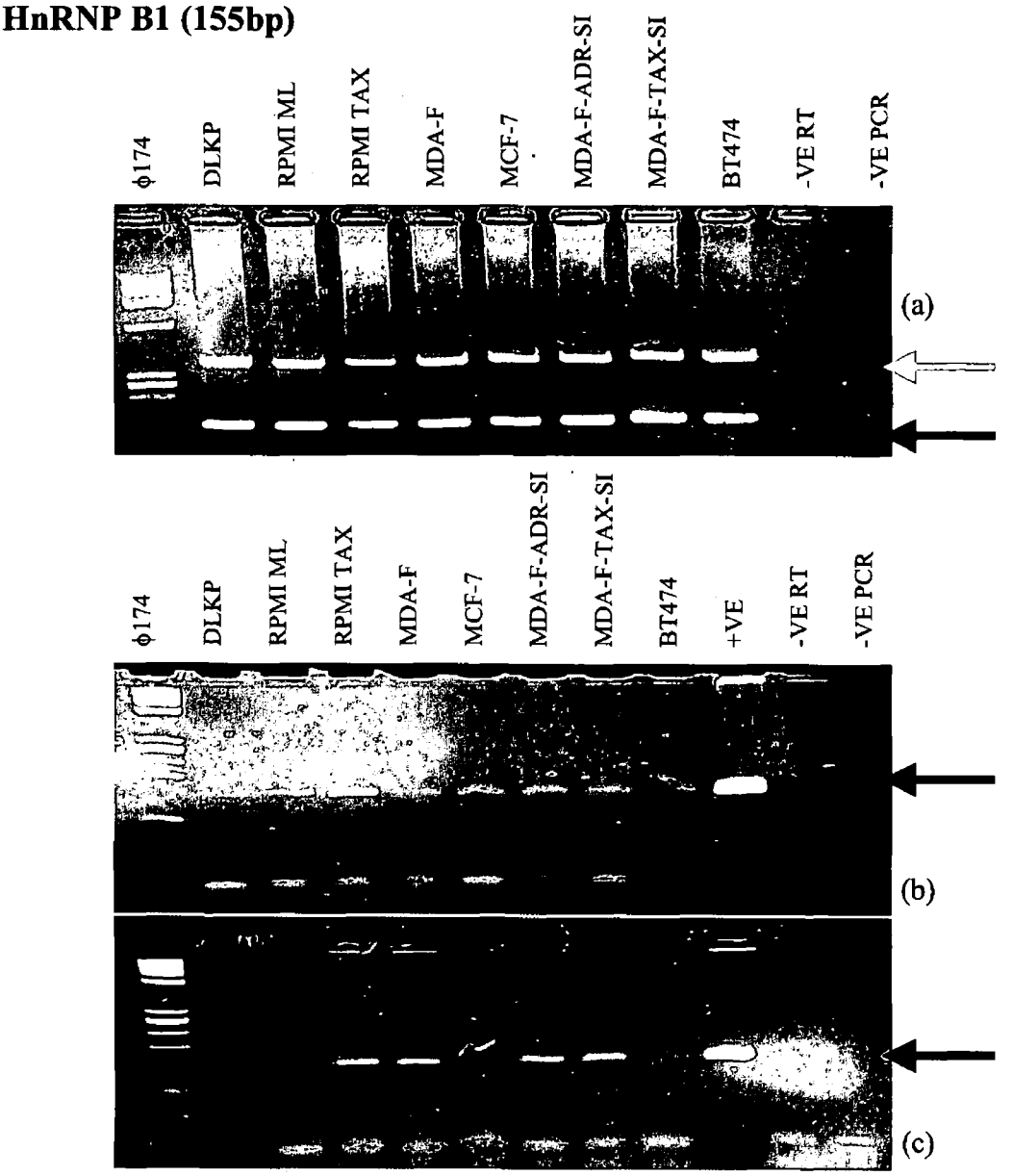


Figure 3.4.3 Expression of HnRNP B1 in cells and filtered CM. HnRNP B1 (black arrow) and β-actin (yellow arrow) gene transcripts were detected in RNA isolated from all cell lines analysed (a) and HnRNP B1 gene transcripts in corresponding CM after passing through a (b) 0.22 μm or (c) 0.45 μm filter. Amplified product was detected in the positive control and was undetected in the negative control samples (-ve RT = RT reaction with H_20 instead of RNA as control; -ve PCR = PCR reaction with H_20 instead of cDNA as control).

GST π (270bp)

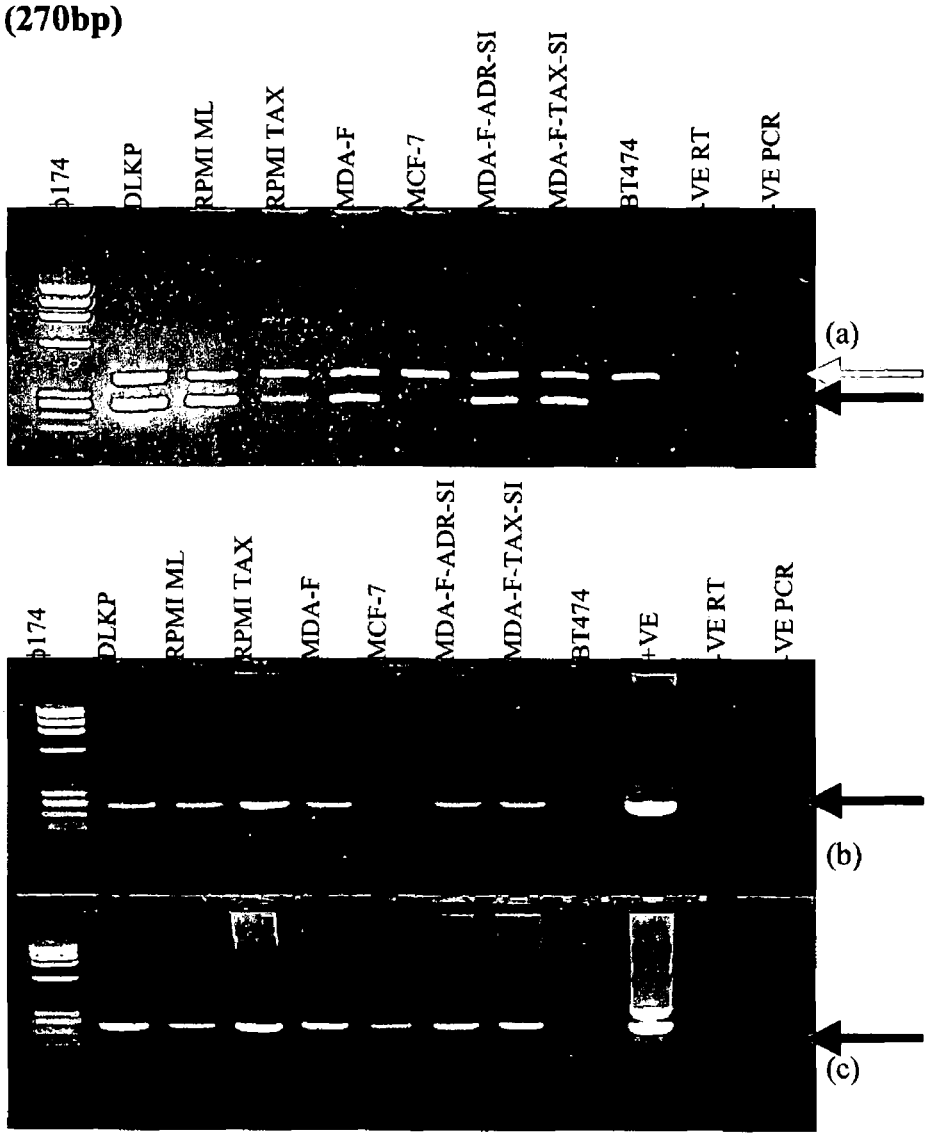


Figure 3.4.4 Expression of GST π in cells and CM. GST π mRNA (black arrow) was expressed by most cell lines and CM. In the case of MCF-7, no product was detected in cell or CM samples; for BT474, GST π was undetected in cell RNA and 0.22 μm filtered CM (b), but a weak band was detected in the 0.45 μm filtrate (c). β-actin was co-amplified with GST π when analysing RNA from cell lines (yellow arrow). Amplified product was detected in the positive control (+VE) (-RT = RT reaction with H_20 instead of RNA as control; -PCR = PCR reaction with H_20 instead of cDNA as control).

TOPO II (216bp)

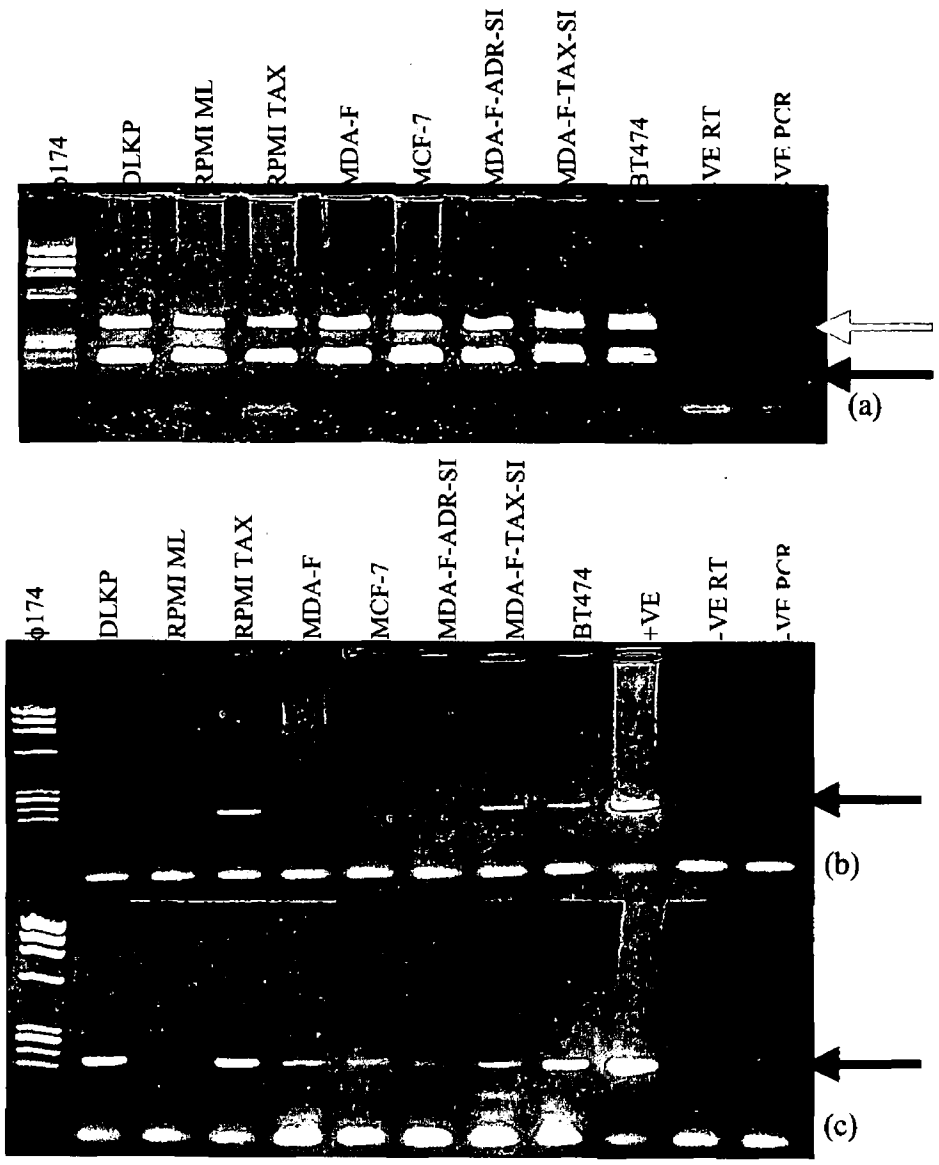


Figure 3.4.5 Expression of TOPO II in cells and CM. Topoisomerase II mRNA (black arrow) was expressed by all cell lines, detected in all 0.45 μm filtered CM (c), but in only 3 0.22μm filtered CM specimens (b). β-actin was successfully co-amplified in cell RNA (yellow arrow (a)). Amplified product was detected in the positive control and was undetected in the negative control samples (-ve RT = RT reaction with H_20 instead of RNA as control; -PCR = PCR reaction with H_20 instead of cDNA as control).

MRP1 (203bp)

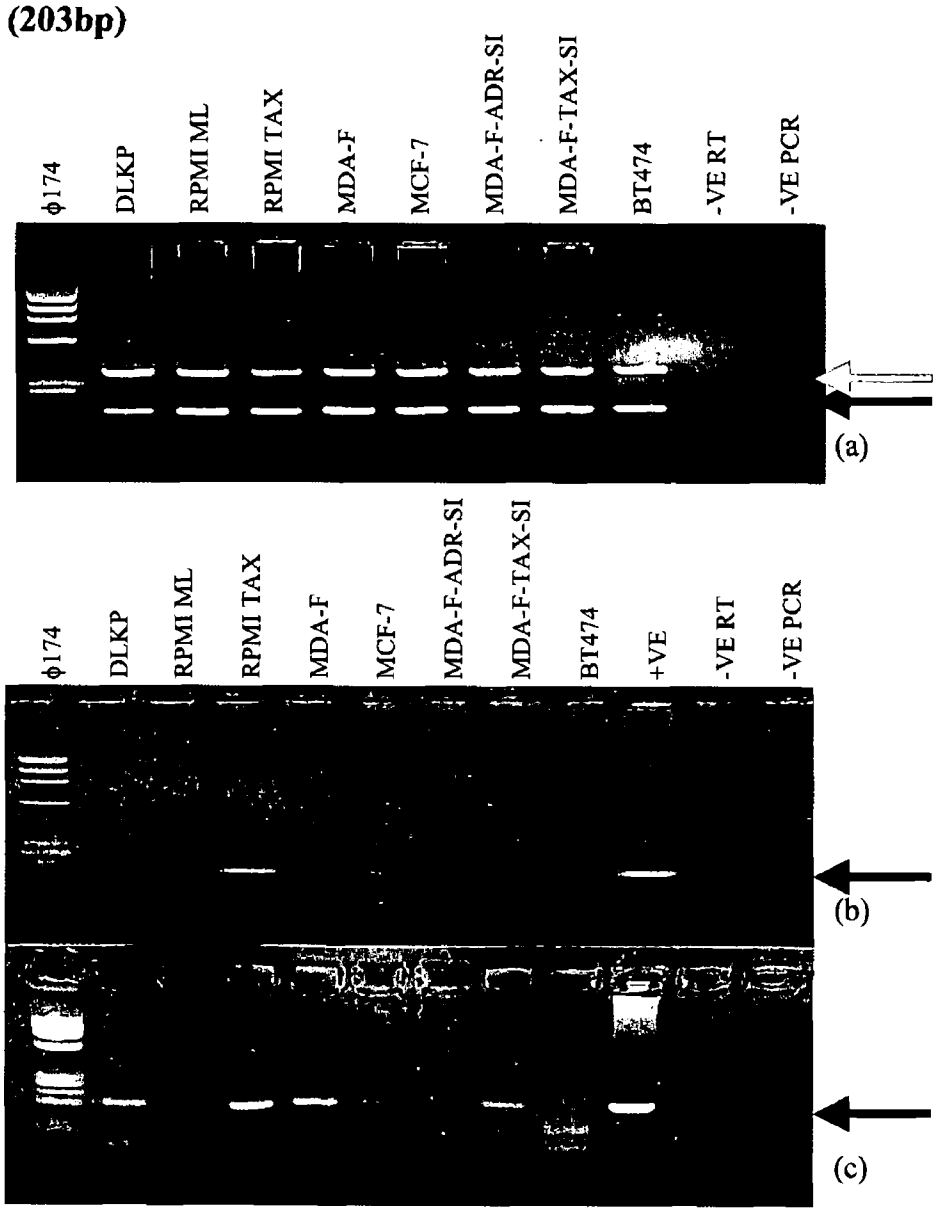


Figure 3.4.6 Expression of MRP1 in cells and CM. MRP1 (black arrow) was expressed by all cell lines, detected in only 2 of the 0.22 μm filtered CM (b), but in 5 of the 0.45 μm CM specimens (c). β-actin was co-amplified with MRP1 when analysing RNA from cell lines (yellow arrow (a)). Amplified product was detected in the positive control and was undetected in the negative control samples (-ve RT = RT reaction with H_20 instead of RNA as control; -PCR = PCR reaction with H_20 instead of cDNA as control).

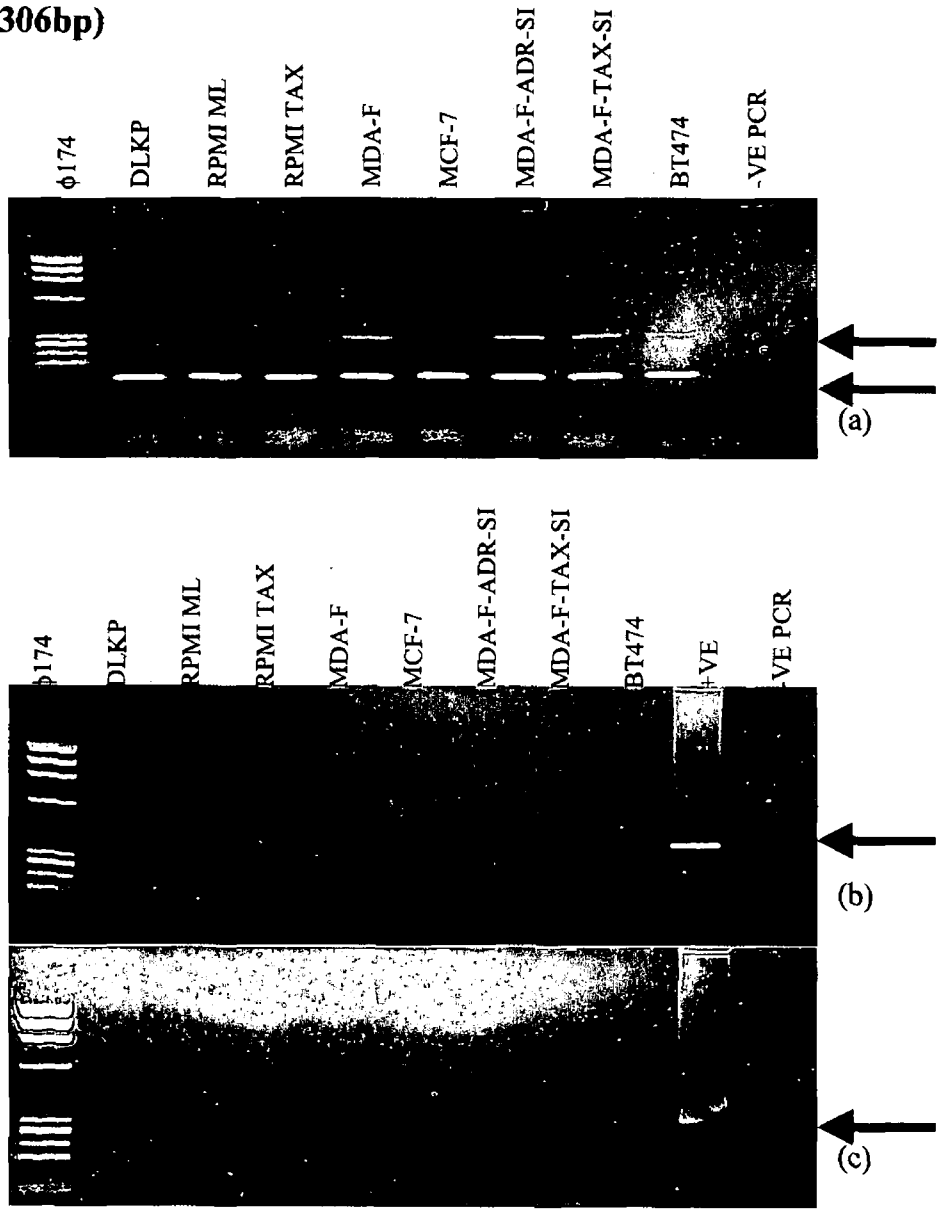


Figure 3.4.7 Expression of bcl-2 in cells and CM. bcl-2 (black arrow) was expressed, to some extent, by all cell lines analysed (a), but with the exception of RPMI-TAX 0.45 μm filtrate (c), was undetected in CM ((a) and (b)). β-actin was co-amplified with bcl-2 when analysing RNA from cell lines (green arrow (a)). Amplified product was detected in the positive control (+VE) and was undetected in the negative control samples (-ve RT = RT reaction with H_20 instead of RNA as control).

MDR1 (157bp)

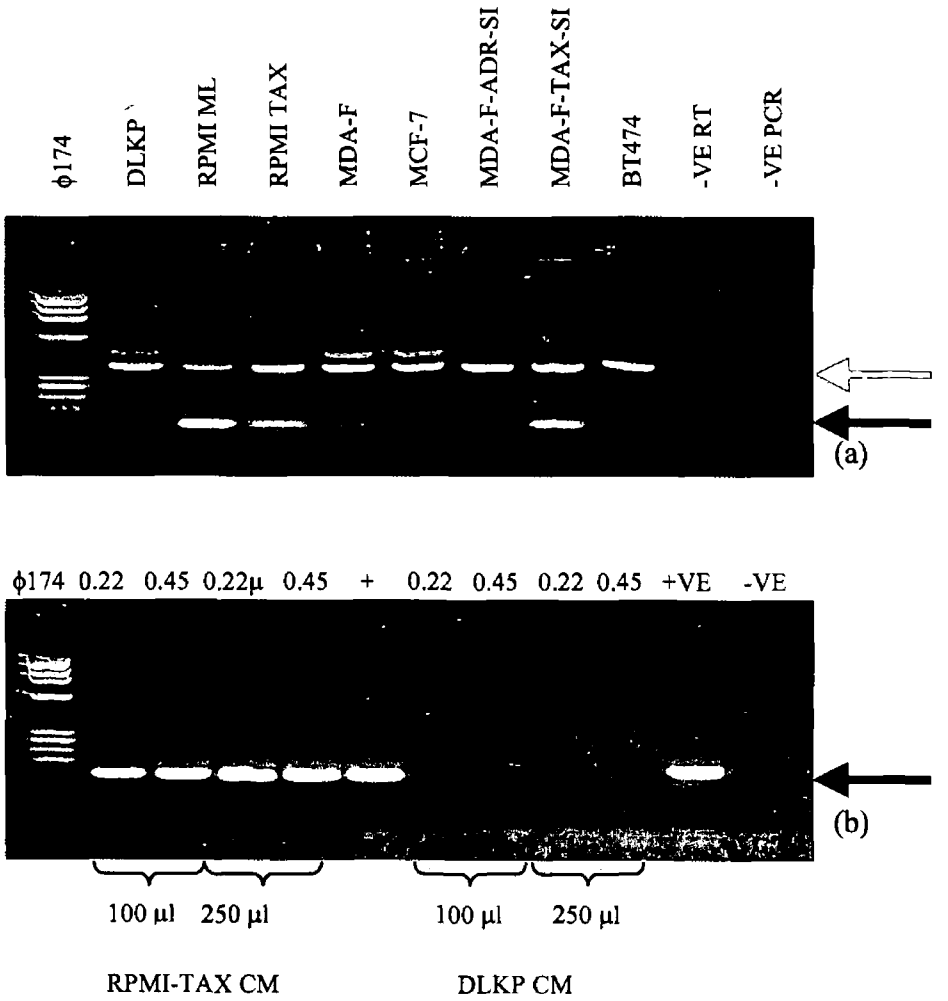


Figure 3.4.8 Expression of MDR 1 in cells and CM. MDR1 mRNA was expressed in RPMI ML, RPMI TAX, MDA-F-TAX-SI, and to a lesser extent MDA-F and BT474 cell mRNA (a). β-actin was co-amplified with MDR1 when analysing RNA from cell lines (yellow arrow (a)). Amplified product was detected in the positive control and was undetected in the negative control samples (-ve = RT reaction with H_20 instead of RNA as control; -PCR = PCR reaction with H_20 instead of cDNA as control).

3.4.1 RT-PCR of media unconditioned by any cells

As mRNA gene transcripts have been detected in serum specimens from normal, healthy humans, it is likely that mRNA transcripts are also detectable in foetal calf serum. To eliminate the possibility that the mRNA found to be amplifiable in CM could be attributed to the foetal calf serum used in the media to culture the cells, the Tri Reagent extraction protocol was carried out on ATCC media (5% FCS) that was unconditioned i.e. unexposed to any cells and filtered with a 0.45μm filter (see Section 2.8.2.2). RT-PCR was carried out on the sample for β-actin as it was present in all the CM samples, therefore if any human mRNA was present in the media unconditioned by any cells, the likelihood is that β-actin would be amplified. 45 cycles of PCR were used. Fig 3.4.1.1 shows that no β-actin gene transcripts were detected in the media that was unexposed to cells indicating that human β-actin mRNA previously detected in the CM samples from cultured human cancer cell lines was not attributable to the FCS used in the preparation of the cell culture media/ or by contamination of the media by any other source.

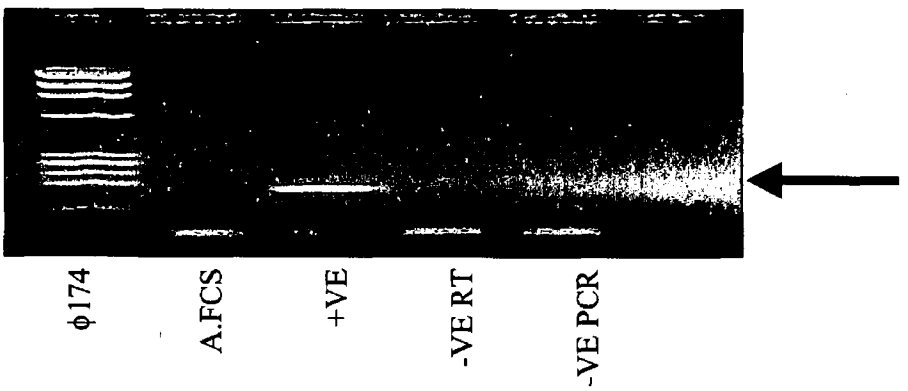


Figure 3.4.1.1 RT-PCR of media unconditioned by cells for β-actin (171bp). β-actin gene transcript (arrow) was detected in the positive control, DLKP cell RNA, sample only. Amplified product was undetected in the unconditioned media sample and negative control samples (-ve RT = RT reaction with H_20 instead of RNA as control; -ve PCR = PCR reaction with H_20 instead of cDNA as control). A.FCS = ATCC+ 5%FCS media.

The aim of Section 3.5 was to test biological sample concentration columns that are used in our laboratory to concentrate protein solutions, to see if they could be employed to concentrate the quantity of RNA in the conditioned media to levels required for labelling and hybridisation to Affymetrix array chips.

3.5 Concentration of CM with Millipore concentrators

It was decided that because we had proven that it was possible to amplify cancer related target gene transcripts in CM samples (Section 3.4) that we should investigate the possibility of concentrating the RNA in CM samples by using Millipore concentrators.

The concentrators can be used to concentrate biological samples containing protein/ nucleic acids based on a process called ultra-filtration. RNA would then be extracted from the concentrated sample to see if there was an increase in the yield of RNA. The aim was to achieve a high enough yield of RNA to be able to use it in microarray studies. The minimum RNA required, at the time, for amplification, labelling and hybridisation to an Affymetrix Plus chip was 5μg.

DLKP cells were used for this study analysis. The experimental designs for the concentration protocols are detailed in Section 2.7.4- 6. RNA was extracted from all of the samples using the Tri Reagent protocol described in Section 2.8.2.2 and resuspended in 15μl DEPC treated water. RNA was also extracted from 25mls of CM using a scaled- up Tri Reagent protocol where the volumes of all solutions were increased 10 fold and the protocol followed as per Section 2.8.2.2, so for example 75mls of Tri Reagent was mixed with the 25 mls of CM.

The RNA samples from all of the protocols mentioned were compared using the Nanodrop to measure the yield (see Section 2.9.1), and the Bioanalyser to look at the fragment sizes in the samples (see Section 2.9.2). RT-PCR was also carried out using selected samples to detect β-actin mRNA transcripts in the different samples.

3.5.1 RNA yield before and after concentration.

The yield of RNA in all of the concentrated samples (C) is higher than the yield in the

samples before concentration (B) and the yield of RNA in the samples before concentration (B) is higher than in the flow- through samples (FT) (see Fig. 3.5.1.1).

The total RNA yield of the CM passed through a 0.22μm filter (as measured by the Nanodrop), was 897ng, 1.13μg and 443ng before concentration (0.22B), with concentration (0.22C) and in the flow- through (0.22FT) respectively. The CM passed through the 0.45μm filter yielded slightly higher amounts of RNA in all three samples; 987ng 1.26μg and 502ng in the RNA samples before concentration (0.45B), with concentration (0.45C) and in the flow- through (0.45FT) respectively. This indicates that the different filter sizes the CM was filtered through had a slight affect on the yield of RNA and also that RNA was able to pass through the column. Whilst there was an increase in the RNA yield in both concentrated samples (0.22/ 0.45C) compared to the RNA yield of the samples before concentration (0.22/ 0.45B), the increase was not in line with the magnitude of concentration of the sample i.e. 14.5mls of CM was concentrated to ~500μls, but the RNA yields, as determined by the Nanodrop, showed an increase in RNA yield of less than 1.5 fold. The RNA yield of the samples from the second concentration protocol (0.45B (4X), 0.45C (4X) and 0.45FT (4X)) are approximately 4 times more concentrated than the RNA yield of the samples obtained in the first concentration protocol (0.45B, 0.45C and 0.45FT) with the exception of 0.45B (4X) (see Fig. 3.5.1.1). This was a direct result of re-suspending the 4 RNA pellets, resulting from the Tri Reagent RNA extraction of 4 x 250μl CM aliquots, into one 15μl volume of water in the second concentration protocol as opposed to one RNA pellet being resuspended in 15μl water in the first concentration protocol. The addition of RNasin had no effect on the yield of RNA in the samples obtained with the third concentration protocol (see Fig. 3.5.1.1, samples 0.45C (4X) +/- RNasin).

A total of 2.7μg of RNA was extracted from 25mls of CM using 75mls of Tri Reagent (sample 25.45 in Fig. 3.5.1.1). This was achieved by scaling up the volumes of all solutions

required for the Tri Reagent RNA extraction protocol and employing the use of an ultracentrifuge at 4°C instead of a bench top centrifuge. This low yield however, is more likely due to the problems encountered during the extraction protocol. The ultracentrifuge tubes have a round bottom, so when the RNA was precipitated using isopropanol and centrifuged, a uniform pellet was not formed. To try to get a uniform pellet, all of the isopropanol was removed, with the exception of approximately 1ml in each of the four tubes, and the tubes were vortexed to lift the RNA pellets. The isopropanol- pellet mixes were then poured into 1.5ml eppendorf tubes so that they could be centrifuged again to obtain a single pellet. This resulted in a large loss of RNA due to the transferral process.

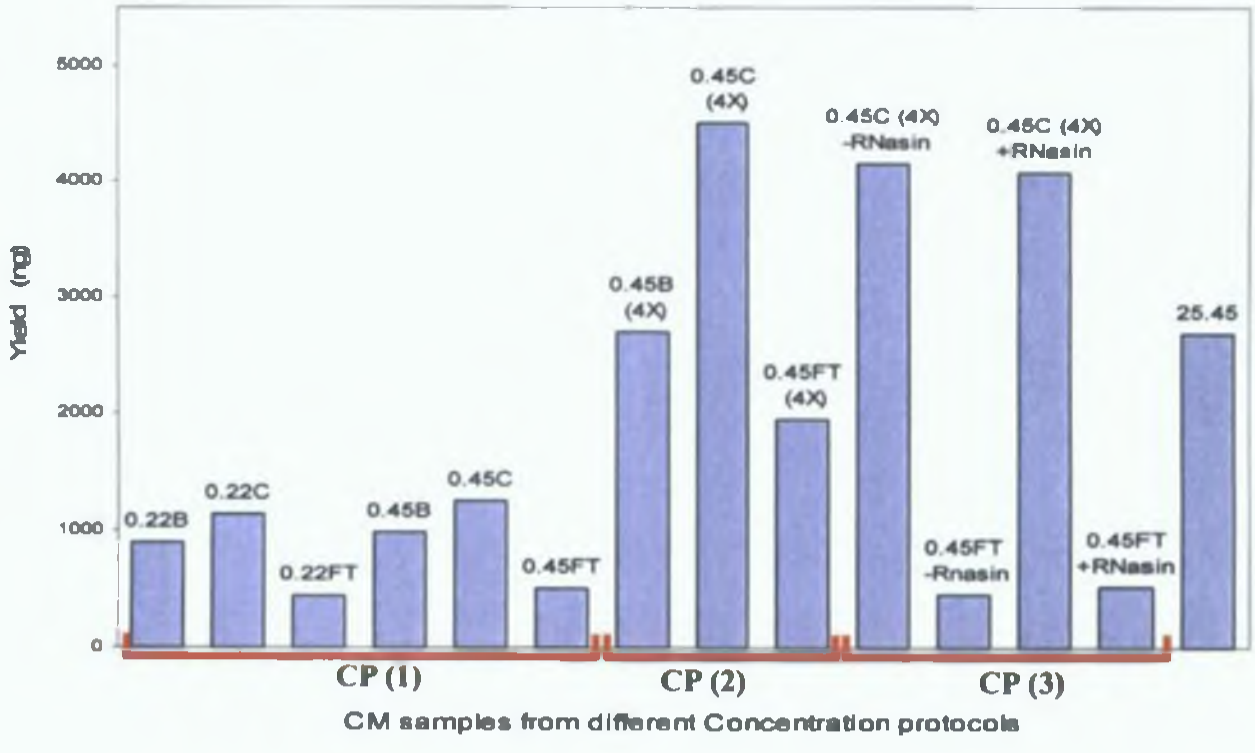


Figure 3.5.1.1 Yield of RNA before and after concentration of CM. The samples obtained using the different protocols are plotted on the X-axis (sample name above each column), with the total RNA yield (ng) plotted on the Y-axis. CP1= Concentration Protocol 1, CP2= Concentration Protocol 2, CP3= Concentration Protocol 3.

3.5.2 Bioanalyser traces of RNA before and after concentration.

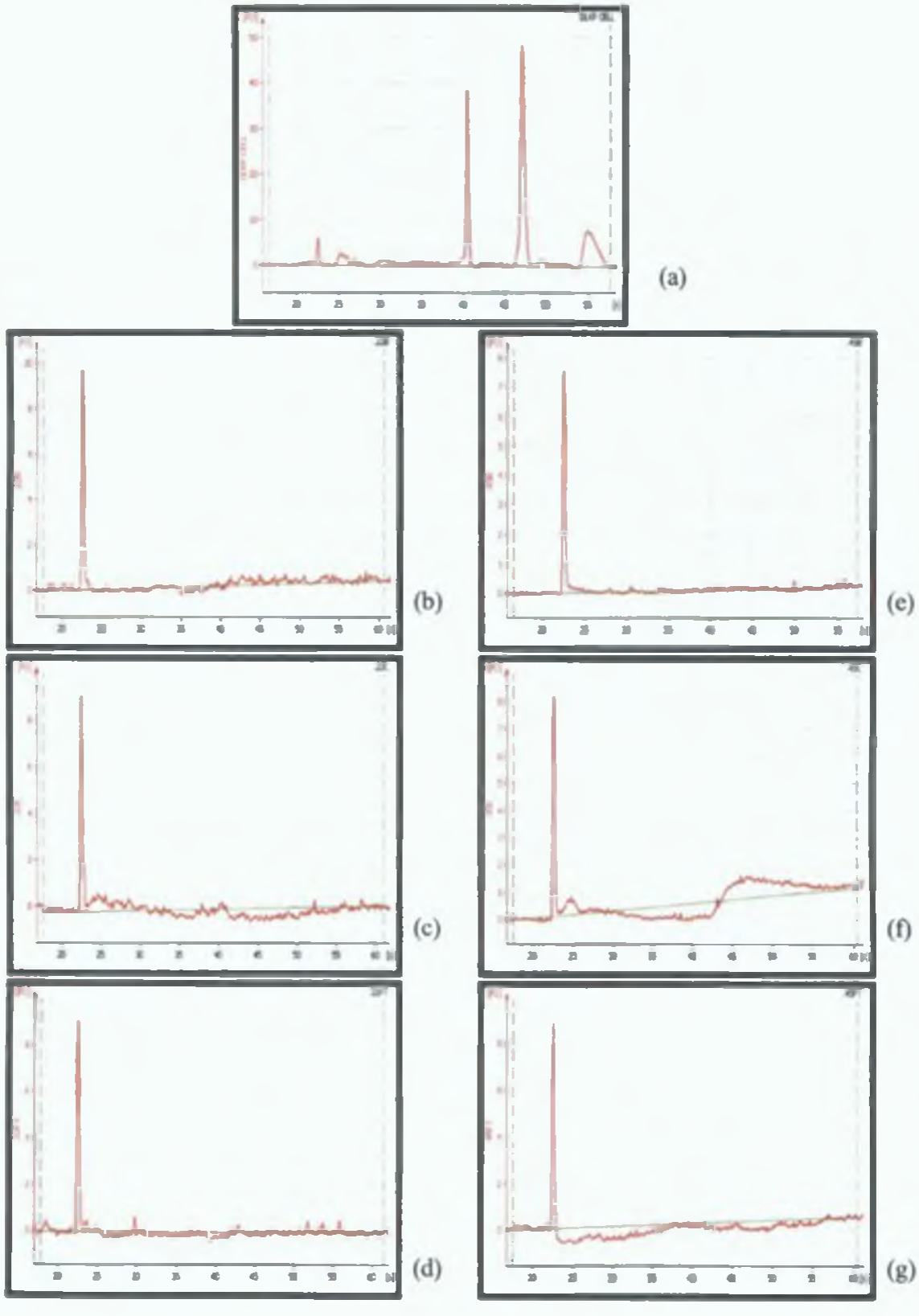
As the Nanodrop can determine the concentration of RNA in the sample but not the integrity of the RNA, aliquots of each sample were run on an Agilent Bioanalyser RNA nano chip to see if fragments of RNA were visible on the chips.

The bioanalyser tracing shown in Fig 3.5.2.1 (a) shows a typical bioanalyser tracing obtained using good quality cell RNA. Integrity of this RNA is measured by 18S and 28S ribosomal bands that appear at 40 and 47 seconds (time is plotted on the X-axis) correlating to 2000 and 4000 nucleotides in length respectively.

None of the RNA samples obtained from CM, before concentration, with concentration or flow-through show the presence of 18s and 28s ribosomal bands regardless of protocol, centrifuge speeds or time taken to process. This result was not unexpected as it would mean that ribosomal RNA is secreted into the CM.

With the exception of the 0.45FT (4X) sample, all RNA samples, were four RNA pellets were resuspended into one 15µl volume of water (0.45B (4X), 0.45C (4X), 0.45C (4X) – RNasin, 0.45C (4X) +RNasin) and the RNA sample resulting from the scale-up of the Tri Reagent protocol (25.45), show evidence of RNA fragments in the samples ranging in sizes up to 1000nts as denoted by the peaks at 25 seconds to 40 seconds (Figs. 3.6.5.2.1 (h), (i), (k), (l) and (m) respectively). The addition of RNasin also had little effect on the profile of RNA extracted from the concentrated CM. (Fig 3.5.2.1 (l) with RNasin, 3.5.2.1 (k) without RNasin).

It was decided at this stage to carry out RT- PCR using some of the samples shown in Fig 3.5.2.1 to test for the presence of β-actin mRNA gene transcripts using oligo dT as the reverse transcription primer and PCR primers that amplify a region of the β-actin gene transcript close to the start of the 5' coding region.


(a)
(b)
(c)
(d)
(e)
(f)
(g)

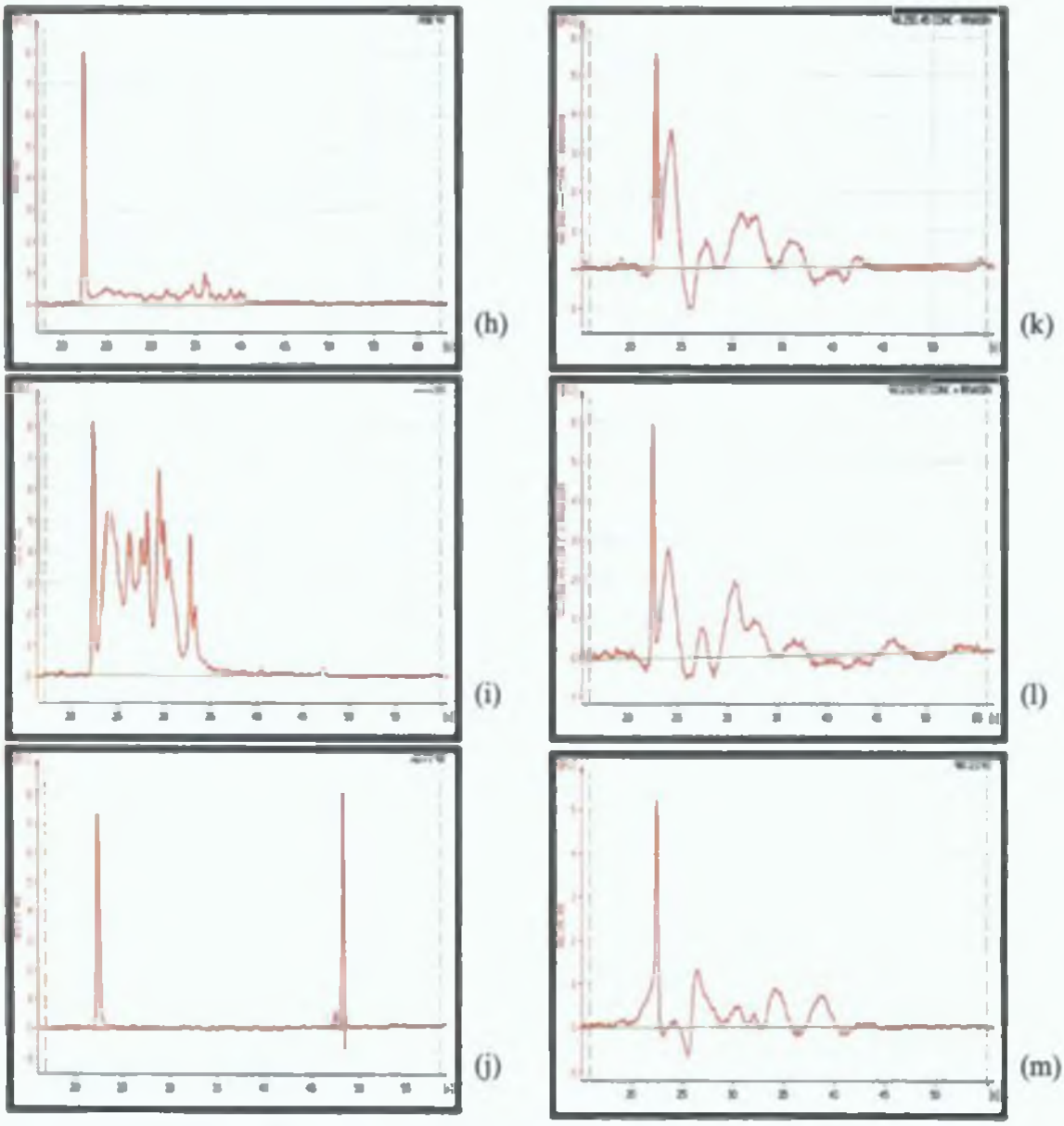


Figure 3.5.2.1 Bioanalyser traces of CM RNA before and after concentration, a = DLKP cell RNA (+ve control), b = 0.22B, c = 0.22C, d = 0.22FT, e = 0.45B, f =0 .45C, g = 0.45FT, h = 0.45B (4X), i = 0.45C (4X), j = 0.45FT (4X), k = 0.45C (4X) –RNasin, l = 0.45C (4X) +RNasin, m = 25.45. The time in seconds is plotted on the X-axis with the fluorescence plotted on the Y-axis. Smaller fragments move more quickly through the gel and are detected first, so peaks on the left hand side of the electrophoregram represent smaller fragments and the height of the fluorescent peaks indicate the abundance of the fragments.

3.5.3 RT-PCR of CM RNA samples before and after concentration for β-actin

β-actin gene transcripts were amplifiable in all CM samples before concentration, after concentration or in the flow through samples with the exception of the flow through sample corresponding to the CM that was passed through a 0.22μm filter (see Fig 3.5.3.1).

In summary, the biological sample concentration columns did increase the yield of RNA extractable from the CM, as evident by the nanodrop, bioanalyser and RT-PCR , but the yield was not enough for labelling and hybridisation to Affymetrix microarray chips as 5μg of total RNA was required at the time.

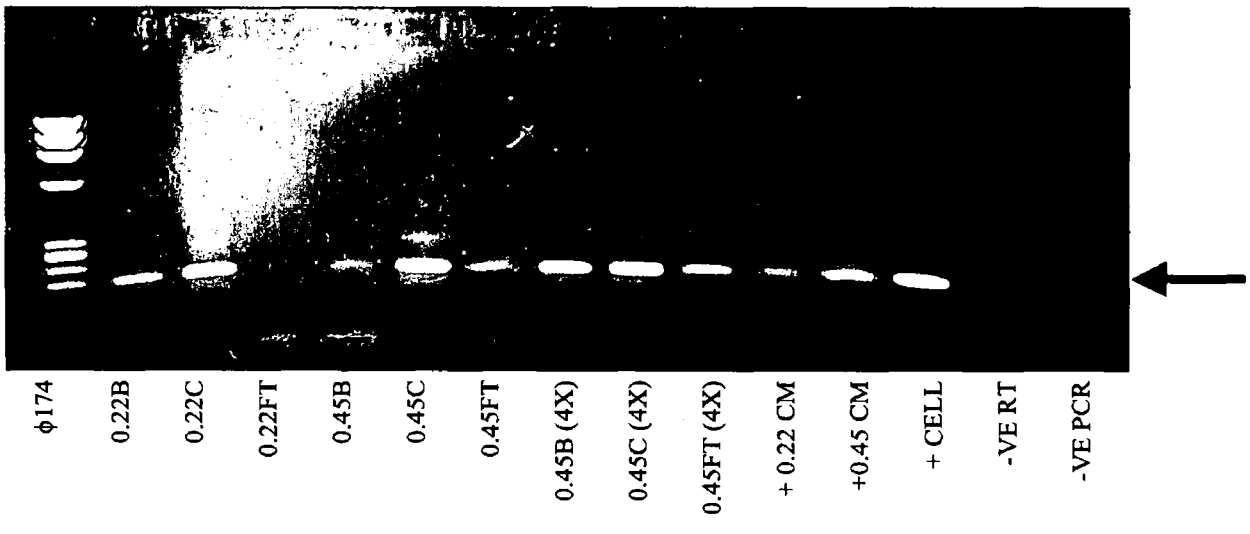


Figure 3.5.3.1 RT- PCR for β-actin (196bp) in CM before and after concentration. β-actin gene transcripts (arrow) were detected in all samples except the flow-through sample that had been filtered with a 0.22μm filter before it was concentrated. Amplified product was detected in the positive control samples (+ 0.22 CM, + 0.45CM and +CELL) and was undetected in the negative control samples (-ve RT = RT reaction with H_20 instead of RNA as control; -ve PCR = PCR reaction with H_20 instead of cDNA as control). The 0.22 CM and 0.45 CM positive control samples were DLKP CM samples previously used in RT-PCR and found to be positive for β-actin, and the cell RNA was from DLKP cell RNA.

The aim of Section 3.6 is to investigate if quantification of RNA from CM extracted using Tri Reagent was overestimated when quantified using the Nanodrop.

3.6 Interference in quantification of small amounts of RNA

The previous study (Section 3.5) indicated that there were some discrepancies with the readings obtained using the Nanodrop with both the bioanalyser traces and the RT-PCR results. Ginzinger (2002) advised caution when using nucleic acids purified with phenol-/ chloroform- based methods for Q-PCR experiments as trace amounts of phenol can contribute to reduced reverse-transcription efficiencies and over estimation of RNA quantification using a UV spectrophotometer. Even though reverse transcription of the CM samples did not seem to be affected, we decided to test if the Tri Reagent method of RNA extraction was contributing to an overestimation of the RNA in the CM samples when read using the Nanodrop (see Section 2.9.1)

To do this, firstly an aliquot of Tri Reagent was diluted 1:100 with both phosphate buffered saline (PBS) and ultra high processed water (UHP) then quantified using the Nanodrop (see Fig. 3.6.1).

UHP (in place of a solution containing RNA) was then processed using the Tri Reagent RNA modified extraction protocol (Section 2.8.2.2) or processed using an RNeasy RNA extraction kit according to the manufacturers protocol. The Tri Reagent extraction protocol involved adding 250µl of UHP to 750µl of Tri Reagent and processing the sample according to the modified protocol with or without an overnight isopropanol precipitation step. This was compared to samples of UHP and PBS that were either firstly processed with Tri Reagent then cleaned up using an RNeasy clean-up protocol or processed only using the RNeasy extraction protocol. The RNeasy RNA extraction procedure is a column- based extraction protocol which relies on lysing RNA using a lysis buffer and adding an ethanol based solution to enable the RNA to bind to a column. The column is then washed with an ethanol based solution and finally eluted using RNase- free water. No phenol is required for

this protocol, so comparisons between the phenol and non-phenol based methods were achievable.

The Tri Reagent/ RNeasy clean-up involves processing the samples using Tri Reagent according to Section 2.8.2.2, then instead of adding isopropanol to precipitate the RNA from the colourless upper aqueous layer, one volume of 70% Ethanol was added and the gently mixed solution was put on to an RNeasy column. The samples were processed according to the manufacturers clean-up protocol.

The Nanodrop reading obtained for diluted Tri Reagent shown in Fig. 3.6.1(a) shows that there is a major peak at 270nm and a minor peak at 260nm. As RNA is quantified by the amount of UV absorbed at 260nm, the Nanodrop is falsely stating that the sample contains ~1.6μg/ μl of RNA.

When an RNA solution is replaced with UHP and the Tri Reagent RNA extraction protocol followed (see Section 2.8.2.2), the Nanodrop falsely quantifies that there is >100ng of RNA in the UHP sample (Figure 3.6.1(b)). There are two peaks at 260 and 270nm which are similar to the peaks obtained when diluted Tri Reagent was read using the Nanodrop (see Fig. 3.6.1(a)). The use of ice- cold or room temperature isopropanol in the protocol does not reduce the interference in the Nanodrop reading.

When an RNA solution is replaced with UHP and the Tri Reagent RNA extraction protocol followed as above with the exception that the isopropanol precipitation step is not carried out overnight, the Nanodrop falsely quantifies that there is >60ng of RNA in the UHP sample (Figure 3.6.1(c)). There is again a peak at 270nm which is similar to the peaks obtained when diluted Tri Reagent was read using the Nanodrop (see Fig. 3.6.1(a)). The use of ice- cold or room temperature isopropanol does not reduce the interference in the Nanodrop reading. The inclusion of an overnight isopropanol precipitation at -20°C step

does seem to increase the false reading (Fig.3.6.1(b)- Nanodrop reading of >100ng compared to >60ng when no over-night isopropanol precipitation step is included (Fig 3.6.1(c)).

The UV readings obtained using UHP/ PBS processed with Tri Reagent and cleaned up using a Qiagen RNeasy column according to the manufacturer's protocol are shown in Fig 3.6.1(d). The sample was eluted using 30µl of RNase free water. A peak at 270nm is still visible in both samples indicating that clean- up of the sample using the RNeasy column was not enough to eradicate the interference in sample quantification.

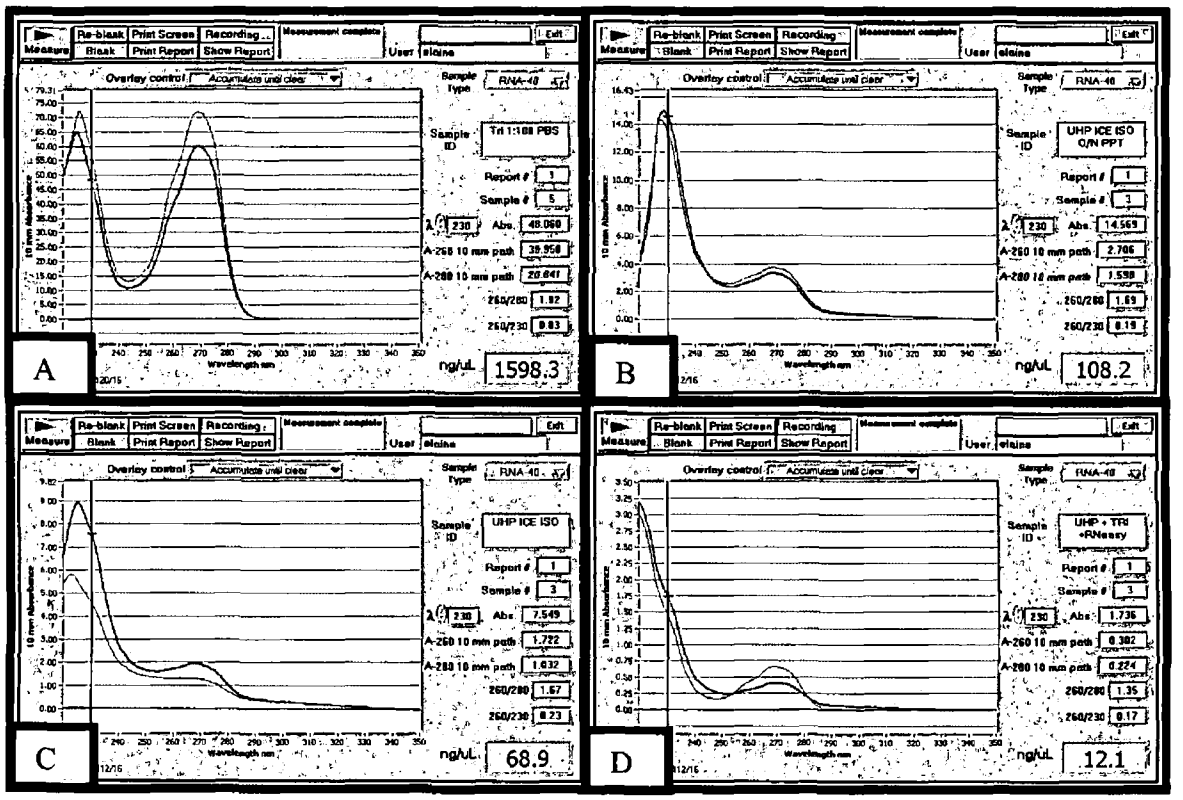


Figure 3.6.1. Nanodrop screenshots of quantified samples. Wavelength is plotted on the X-axis and UV absorbance on the Y-axis. (a) Tri Reagent diluted with PBS/ UHP. Tri Reagent was diluted 1:100 with either PBS or UHP and 1µl was read using the Nanodrop (b) Tri Reagent extraction of UHP overnight (O/N) precipitation. Either ice- cold isopropanol (black line) or room temperature isopropanol (red line) was used in the RNA extraction protocol. 1µl from a 15µl sample was tested (c) Tri Reagent extraction of UHP with no O/N precipitation step. Either ice- cold isopropanol (green

line) or room temperature isopropanol (pink line) was used in the RNA extraction protocol. 1µl from a 15µl sample was tested (d) Tri Reagent extraction of UHP with using RNeasy cleanup protocol. Either UHP (pink line) or PBS (green line) was used instead of RNA and processed using a modified Tri Reagent protocol that includes an RNeasy cleanup step. 1µl from the 30µl eluted from the column was tested.

Fig. 3.6.2 shows that unlike the UHP/ PBS samples that were processed only with Tri Reagent (pink and black lines) or processed with Tri Reagent and cleaned up using an RNeasy column (Fig 3.6.1(d)), the UHP/ PBS samples that were processed with the RNeasy kit RNA extraction procedure alone (light green and brown lines in Fig. 3.6.2), which does not require phenolshow no absorbance at 260nm.

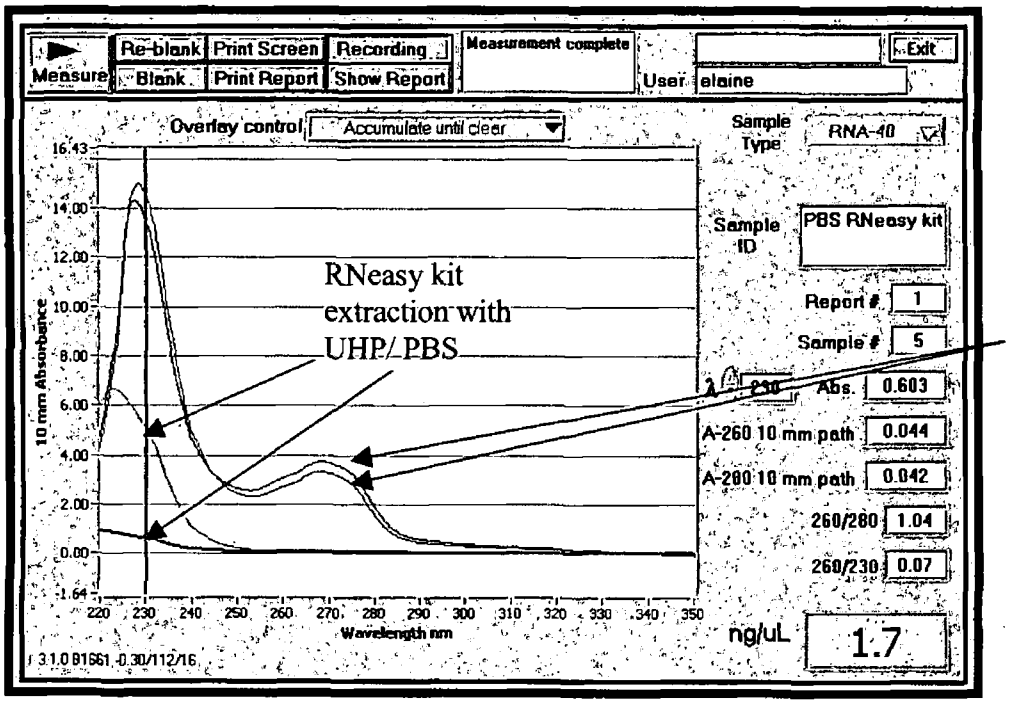


Figure 3.6.2 RNeasy extraction of UHP. Wavelength is plotted on the X-axis and UV absorbance on the Y-axis. Samples already shown in Fig 3.6.3 are re- read and compared to UHP (light blue line) or PBS (brown line) that was processed using the RNeasy RNA extraction procedure only, with no Tri Reagent

The extraction of RNA from CM using Tri Reagent shows evidence of phenol contamination as indicated by two peaks at 260 and a larger one at 270nm (Fig 3.6.3). However RNA extracted from cells (Fig 3.6.1) shows no indication of a contaminating peak at 270nm despite also being extracted using Tri Reagent. This seems that phenol carryover leading to overestimation of samples containing RNA only affects/ interferes with the quantification of RNA when the RNA solution cannot be diluted before quantification as in the case of RNA extracted from CM.

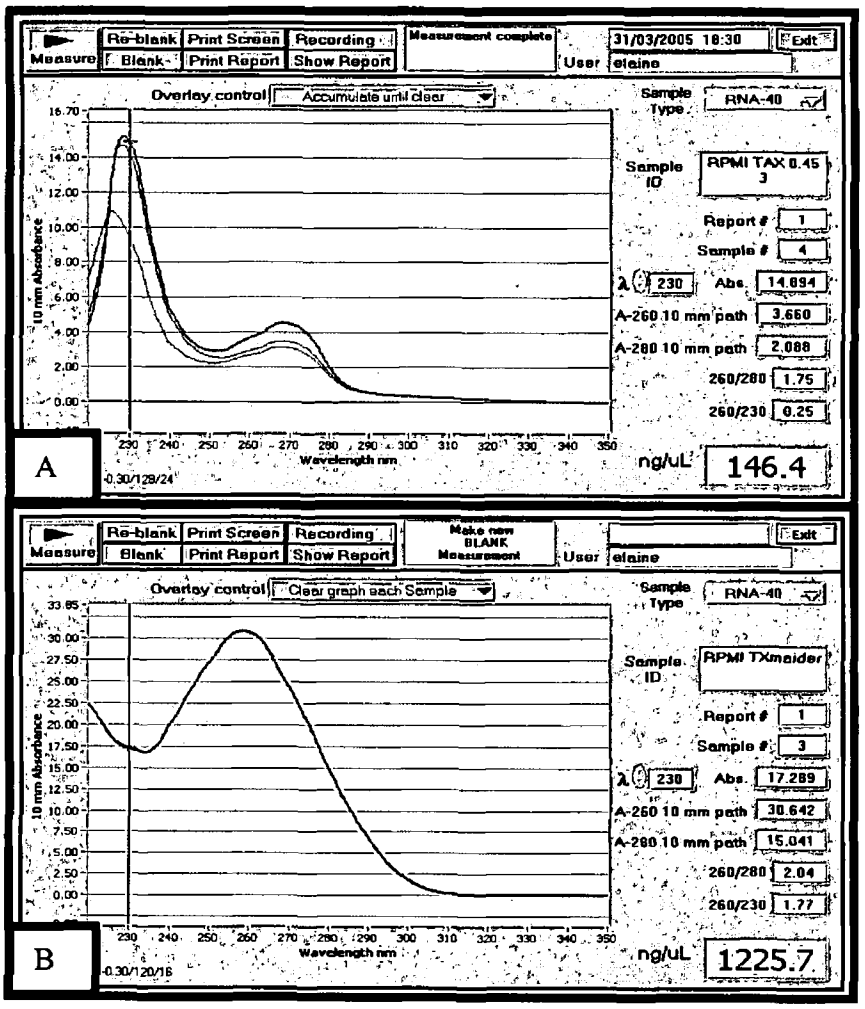


Figure 3.6.3 Nanodrop screenshot of quantification of 1µl of (A) CM RNA and (B) cell RNA extracted using Tri Reagent (Section 2.8.1 and 2.8.2.2) RNA. The wavelength is plotted on the X- axis and the UV Absorbance on the Y-axis.

The aim of Section 3.7 was to assess if it was possible to amplify enough cRNA from CM samples to be hybridised to an Affymetrix human genome Plus 2.0 microarray chip.

3.7 Pilot Microarray Study of DLKP CM

Although quantification of RNA extracted from CM is overestimated when using the Nanodrop (see Section 3.6), there is no doubt to the presence of mRNA in these samples because of all the RT-PCR work carried out on these samples (see Sections 3.1- 3.4).

Affymetrix recently released a 2-cycle amplification protocol that enables the amplification and labelling of RNA when as little as 10-100ng of starting material is available. We decided to test if CM RNA could be successfully amplified and labelled to yield the 15μg of biotinylated cRNA required for the protocol. As there is no other reliable means of quantifying the CM RNA at the time, 100ng from each sample, as quantified by the Nanodrop, were used as starting material. As the protocol can handle 10ng- 100ng, it was hoped that we would have adequate RNA in the samples to fall into this range. This would mean that the need to concentrate the RNA with biological sample concentrators would not be necessary.

DLKP cells were set up and CM collected and processed according to Section 2.7.7. Initially only one sample was tested (DLKP 1), but when we found it was possible to profile CM RNA using CM extracted from this sample, a further 2 biological repeat samples were prepared (DLKP 2 and DLKP 3). RNA was extracted from the CM samples according to Section 2.8.2.2. 100ng of the extracted RNA, as measured by the Nanodrop was then linearly amplified using the methods described in Section 2.12. 15μg of this cRNA for each of the three replicate CM samples was then successfully hybridised to an Affymetrix U133A Plus 2.0 chip.

3.7.1 Nanodrop/ bioanalyser traces for DLKP CM

The Nanodrop readings shown in Fig 3.7.1.1 show that whilst the profile of the DLKP 3 CM sample before amplification (Fig. 3.7.1.1 (a)) shows the presence of two distinct peaks at 260 and 270nm indicating the presence of phenol/ Tri Reagent carryover, the profile of cRNA after two rounds of amplification (Fig 3.7.1.1 (b)) shows no evidence of any phenol contamination/ Tri Reagent carryover as evident by the presence of just one single peak at 260nm.

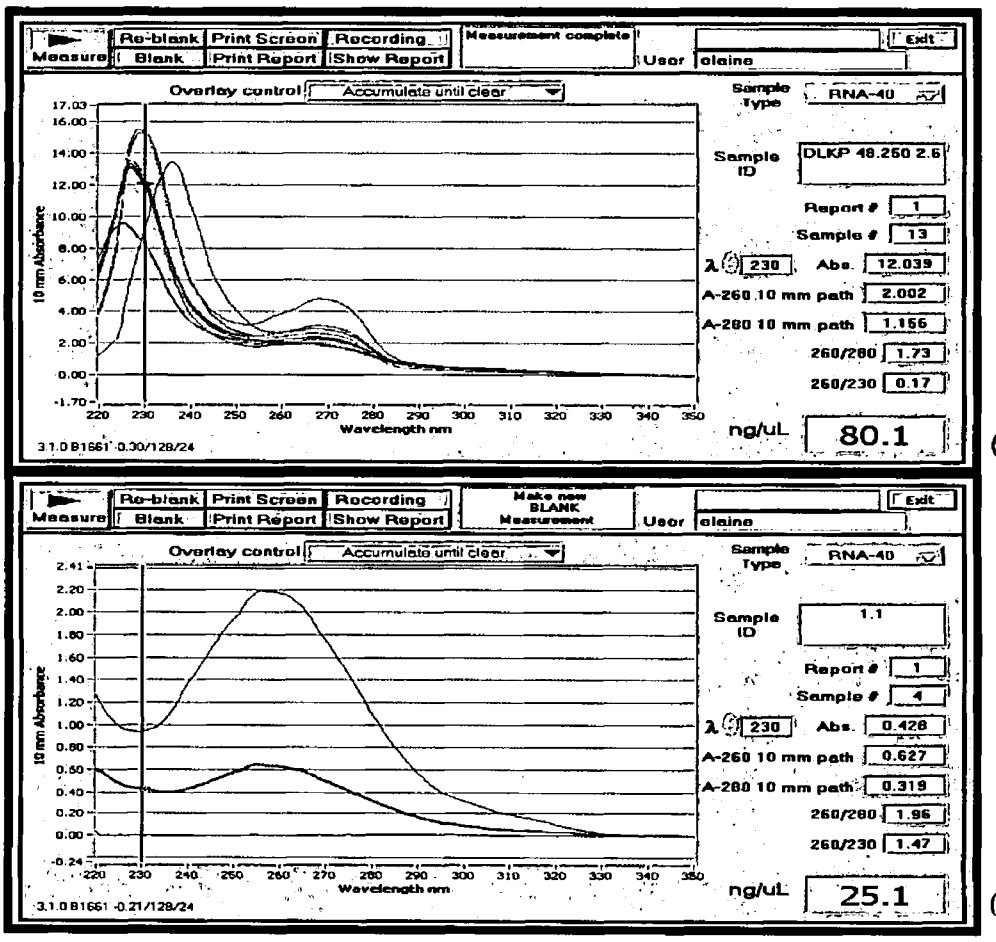


Figure 3.7.1.1 An example of Nanodrop readings obtained for DLKP CM RNA before amplification ((a), multiple samples) and after two rounds of amplification (b) (1.1= DLKP 3 sample, green line). Wavelength is measured on the X-axis, absorbance on the Y-axis. RNA absorbs UV at 260nm.

The quantity of cRNA generated after both cycles of linear amplification is listed in Table

3.7.1.1. The results show that after two rounds of linear amplification, the mRNA in the three DLKP samples have been amplified enough to yield the 15µg of labelled cRNA required for hybridisation to the Affymetrix Plus 2.0 microarray chips.

Following fragmentation of the cRNA in preparation for hybridisation to the Affymetrix microarray chips, an aliquot of the samples before amplification, after amplification and after fragmentation were run on an Agilent Bioanalyser chip (see Section 2.9.2).

Sample	After 1st IVT	After 2nd IVT
DLKP 1	335.7ng	26.65µg
DLKP 2	143.6ng	21.56µg
DLKP 3	243ng	23.75µg

Table 3.7.1.1 Quantity of RNA during 2 cycle amplification protocol. DLKP 1/ 2/ 3 indicate the CM samples that were linearly amplified, IVT= in vitro transcription.

The results of the bioanalyser tracings shown in Fig 3.7.1.2 indicate that even with a sample of RNA that showed no fragments visible in a bioanalyser RNA nano trace (Fig 3.7.1.2 (a)), the two cycle amplification protocol was able to successfully amplify 23.75µg of biotinylated cRNA (Fig. 3.7.1.2 (b)). This cRNA was then fragmented into nucleotides of approximately 200bps long (Fig 3.7.1.2 (c)) and hybridised to the Affymetrix Plus 2.0 chip. Initially, the DLKP 2 CM sample did not yield the 15µg of RNA required for hybridisation to the Affymetrix chips so the amplification and labelling process was repeated for this sample. The second attempt was successful in yielding the required 15µg cRNA.

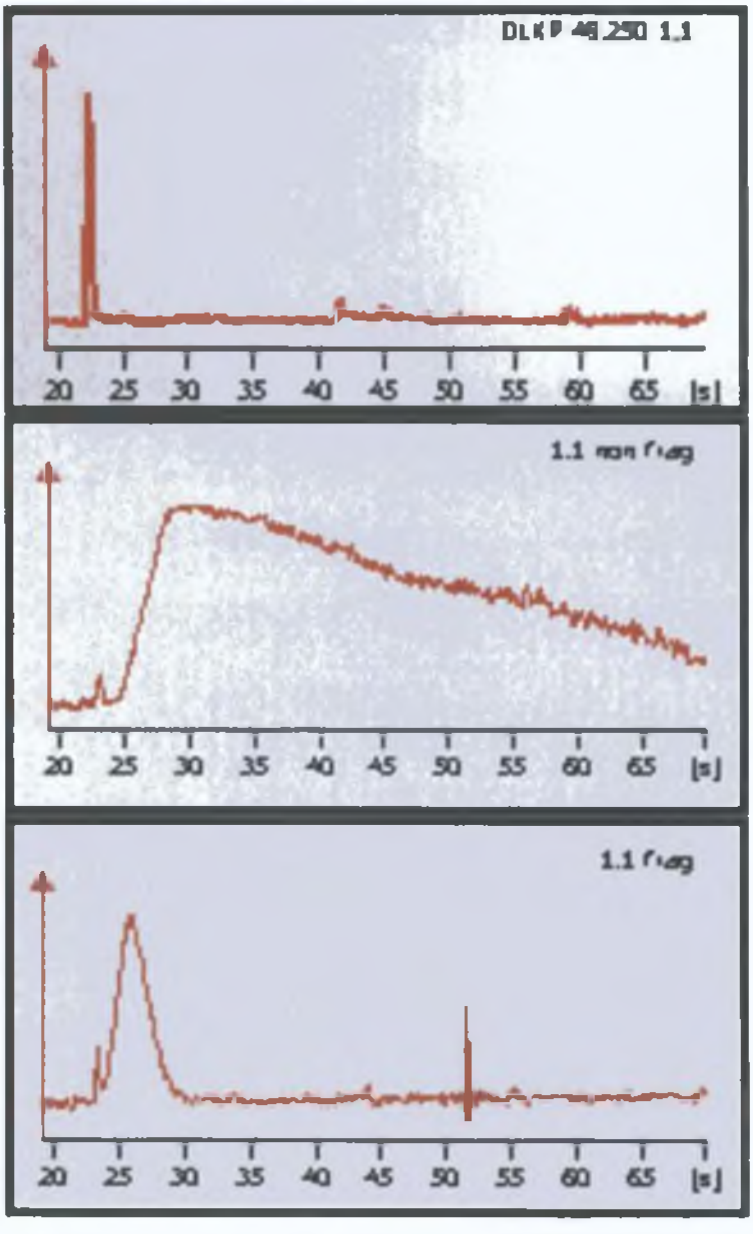


Figure 3.7.1.2 Bioanalyser traces of DLKP CM before and after IVT. The time in seconds is plotted on the X-axis with the fluorescence plotted on the Y-axis. Smaller fragments move more quickly through the gel and so are detected first, so peaks on the left hand side of the electrophoregram represent smaller fragments and the height of the fluorescent peaks indicate the abundance of the fragments. (a) RNA extracted from DLKP 3 CM (b) DLKP 3 cRNA after two rounds of amplification and labelling (c) DLKP 3 cRNA, fragmented to 200 nucleotides.

3.7.2 Microarray Q.A results for DLKP CM

There was an average Present call of 23.1% as determined by Affymetrix calculations. The scaling factors of the DLKP CM samples (Table 3.7.2.1), although higher than the scaling factor usually obtained when dealing with cell derived RNA (1-3), they are similar to each other and therefore are acceptable.

The six probes sets that target β-actin and GAPDH mRNA (3 each) were called Present in all three samples. The GAPDH 3'/5' and 3'/M ratios were higher than expected (compared to ratios no higher than 3 with cell RNA). But as the RNA we are using to amplify and hybridise to the arrays is essentially extracellular mRNA and this RNA has never been hybridised to microarray chips in publications before, we have nothing to compare it to other than the profiles resulting from the "perfect RNA" extracted from cells. The noise and background QA parameters were within acceptable ranges.

Parameter	DLKP 1	DLKP 2	DLKP 3
Scaling Factor	8.858	10.268	8.269
Noise	1.17	1.210	1.34
Background	37.83	36.77	40.15
% Present	23.2	22.6	23.5
3'/ 5' GAPDH ratio	11.66	6.31	7.15
3'/M GAPDH ratio	3.15	1.47	1.27

Table 3.7.2.1 QA results of DLKP CM Affymetrix U133A Plus 2.0 chips. (Explanation for the QA parameters measured with the microarray chips to determine the quality of the chips can be found in Section 2.13.7.)

The data for the three DLKP CM samples were normalised with dChip (Section 2.13.8.2) and imported into Genespring (Section 2.13.9.1) for visual inspection.

3.7.3 Scatter plot and Venn diagrams of gene transcripts expressed in DLKP CM RNA

Approximately 20,000 probe sets are called P in each of the replicate samples, of this 9,528 probe sets are called P in all three arrays (See Venn diagram Fig. 3.7.3.1). 22,099 probe sets are called A in all three DLKP CM samples. Over 6,000 probes sets called P in each sample is also called present by at least 2 of the three replicate samples. The remainder of the probe sets called P in each of the replicate samples are called Absent in the other two replicate arrays (approximately 4,500 probe sets per replicate sample.

The Venn diagram in Fig. 3.7.3.2 indicates that there is a distinction between the D1 sample and D2 and D3 samples, D1 has approximately 1,100 more probe sets with expression levels over 100 than the D2 and D3 samples. The D2 and D3 samples have about 200 probe sets each that are expressed at levels over 100 in only those samples compared to 1,307 probe sets expressed at levels over 100 in the D1 and not in the D2 or D3 samples . 2,139 probe sets are commonly expressed at levels over 100 by all three replicate samples.

The scatter plot shown in Fig. 3.7.3.2 demonstrates that the majority of gene transcripts with levels of expression over 100 fluorescent units are expressed at similar levels by all three samples (indicated by the white dots). The gene transcripts with low levels of expression (under 100 expression units, coloured by grey dots), are not uniformly expressed in the three replicate samples i.e. they do not group around a central line indicating similar expression levels, instead they form a bell- shape curve signifying wide variation. This indicates that samples with a low level of expression <100 units, may simply be attributable to noise and so care must be taken when assessing changes in gene expression at such low levels.

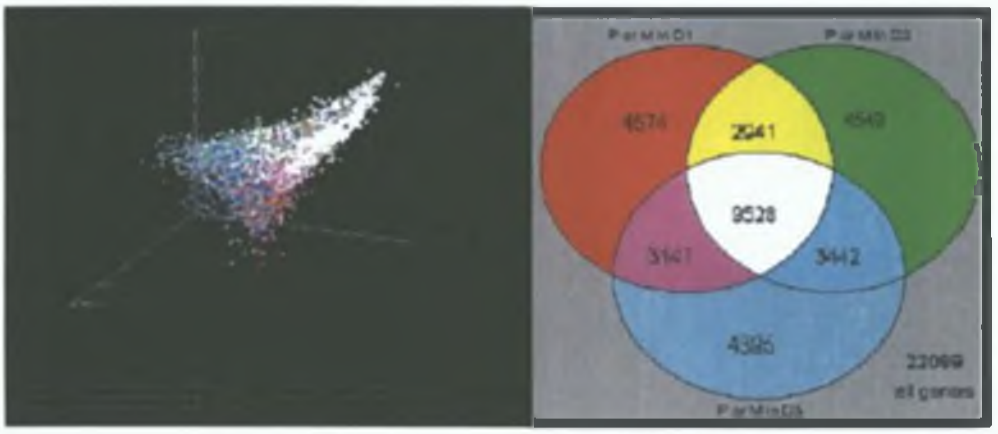


Figure 3.7.3.1 Scatter plot and Venn diagram of probe sets called P or M in DLKP CM samples. In the scatter plot (left-hand side) the expression levels of the probe sets in the different samples are plotted, DLKP 1 CM on the X-axis, DLKP 2 on the Y-axis and DLKP 3 on the Z axis. The colour of each dot, which represents a probe set, correlates to its grouping on the Venn diagram (right- hand side). In the Venn diagram, D1= DLKP 1, D2= DLKP 2, D3= DLKP 3

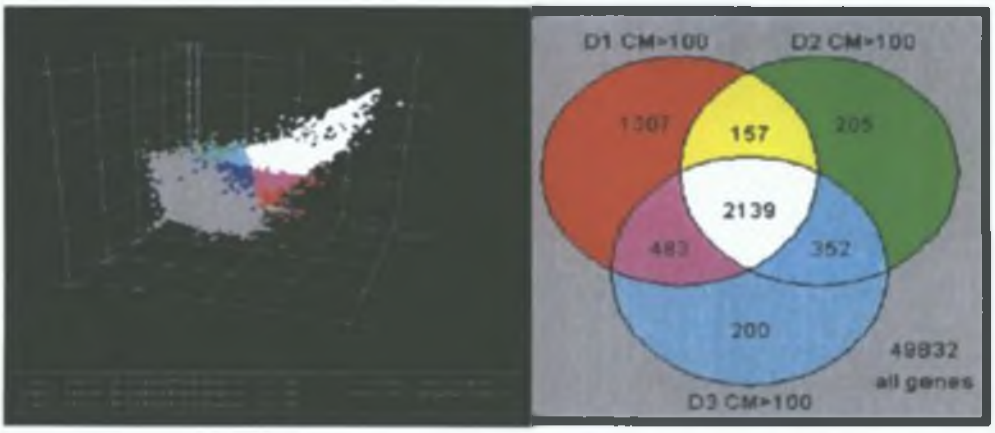


Figure 3.7.3.2 Scatter plot and Venn diagram of probe sets with expression over 100 in DLKP CM samples In the scatter plot (left-hand side) the expression levels of the probe sets in the different samples are plotted, DLKP 1 CM on the X-axis, DLKP 2 on the Y-axis and DLKP 3 on the Z axis. The colour of each dot, which represents a probe set, correlates to its grouping on the Venn diagram (right- hand side). In the Venn diagram, D1= DLKP 1, D2= DLKP 2, D3= DLKP 3

The gene expression profiles for the DLKP CM samples were also compared to the profiles

obtained with DLKP cell RNA. The DLKP cell RNA samples RNA extraction, amplification, labelling and hybridisation were prepared by Helena Joyce and Jason McMorrow. It must be noted that the DLKP CM RNA in this study was not conditioned by these cells and also that these cells underwent just one round of amplification so any comparisons between the two profiles must be made with these notes in mind.

3.7.4 Scatter plot and Venn diagrams of gene transcripts expressed in DLKP cell RNA

22, 381 probe sets are called P in all three DLKP cell samples compared to just 9,528 probe sets commonly called Present in the three replicate DLKP CM samples (see Fig. 3.7.4.1 for details). 19,056 probe sets are called A in the three cell samples. Approximately 2,500 probe sets are called P or M in the each of the three cell samples that are called Absent in the other two replicate samples and a further 3,200- 3,700 probe sets are called P in each replicate sample that are also called P in at least one other replicate sample.

Visually it is clear comparing the scatter plots in Fig. 3.7.4.1 to the 3-D scatter plot obtained for DLKP CM samples (Figs. 3.7.3.1/ 2) that the expression pattern of probe sets called P or M in the cell RNA samples is much more uniform and with tighter clustering around a central line. A scatter plot with a linear 45° angle on an X Y Z scatter plot would indicate that the expression levels of the probe sets in the three samples were identical, so the tighter the expression pattern in the 3-D scatter plot to a 45° angle, the more uniform the expression.

This is also reflected in the fact that of all the probe sets called P or M in the DLKP cell RNA samples, 63% of the probe sets (22,381) are commonly called P or M in all three samples (Fig. 3.7.4.1) compared to only 30% of the probe sets called P or M in the DLKP CM samples (9,528/ 32, 576 probe sets, Fig. 3.7.3.1).

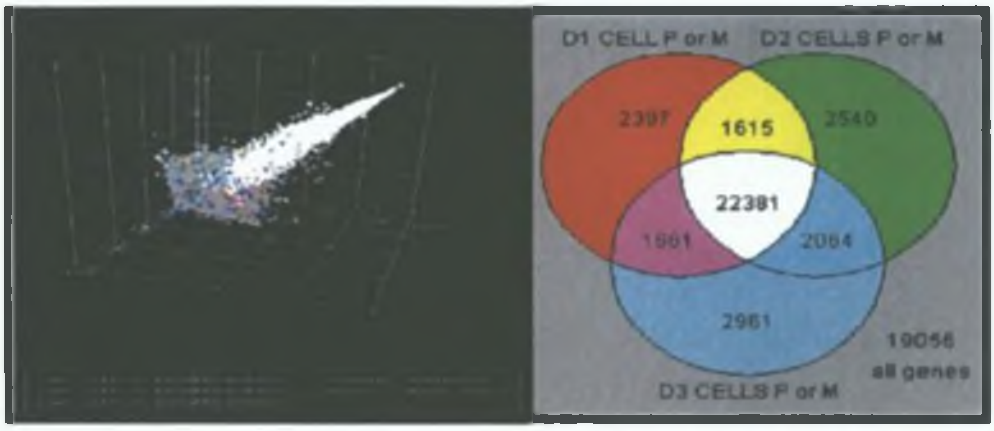


Figure 3.7.4.1 Scatter plot and Venn diagram of probes called P or M in DLKP cell RNA. In the scatter plot (left- hand side), DLKP 1 cell RNA is plotted on the X-axis, DLKP 2 cell RNA on the Y-axis and DLKP 3 on the Z-axis. The colour of the dots in the scatter plot correlates to their grouping according to the Venn diagram (right-hand side). In the Venn diagram, D1= DLKP 1 cell RNA, D2 = DLKP 2 cell RNA, D3= DLKP 3 cell RNA.

3.7.5 Common gene transcripts called P or M in 3 DLKP cell arrays and 3 DLKP CM arrays

Of the 9,528 probe sets called P in all of the CM samples, 80% of them (7, 645 probe sets) were also called P in the three DLKP cell samples. 20% of the probe sets called P in all three CM samples however were not called P in all three cell samples.

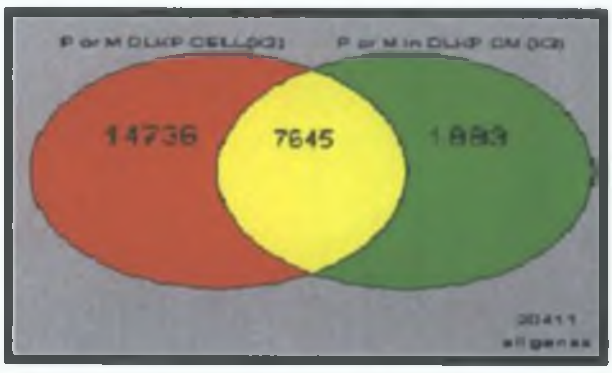


Figure 3.7.5.1 Gene transcripts called P or M in 3/3 DLKP cell arrays vs. Gene transcripts called P or M in 3/3 DLKP CM arrays. The number of probe sets called P or M in all DLKP cell RNA samples (left circle) are overlapped with the number of probe sets called P or M in all DLKP CM RNA samples (right circle).

It was decided to apply an expression cut – off of 100 to the CM samples only because Fig. 3.7.3.2 showed that these probe sets displayed more uniform expression to see if the concordance of overlap between the two different sample types could be increased.

3.7.6 Common gene transcripts called P or M in 3 DLKP cell arrays and 3 DLKP CM arrays when expression cut- off of 100 is applied to CM only

Of the 2,139 probe sets that are expressed at over 100 fluorescent units in the CM samples, Fig. 3.7.6.1 shows that 70% of these are also called P or M in 3/3 cell samples. Of the 2,139 probe sets with expression over 100 fluorescent units in 3/3 CM samples, 1,432 (67%) have a P or M call assigned to them. This result indicates that only including CM probe sets with an expression value of over 100 did not increase the concordance of overlap between the probe sets in the DLKP cell RNA samples and the probe sets in the DLKP CM samples.

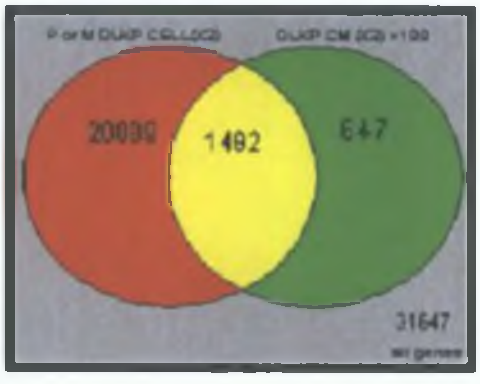


Figure 3.7.6.1 Gene transcripts called P or M in 3 DLKP cell arrays vs. Gene transcripts called P or M in 3 DLKP CM arrays when cut-off of 100 is applied to CM only. The number of probe sets called P or M in all three DLKP cell samples (left circle) are overlapped with the probe sets that are expressed at higher than 100 fluorescent units in the DLKP CM samples (right circle).

3.7.7 Overview of gene ontology of gene transcripts common to DLKP CM and cells using Onto-Express

The gene transcripts called P or M DLKP cell RNA that were also called P or M with an expression cut –off of 100 in the DLKP CM samples were entered into Onto- Express for functional annotation to be assigned (see Section 2.13.9.4). Two pie charts shown in Figs. 3.7.7.1 and 3.7.7.2 show the results obtained from the Onto-Express software. Only categories with a minimum number of 20 gene transcripts were included, if there were less than 20 gene transcripts in a category they were grouped in a segment called "other".

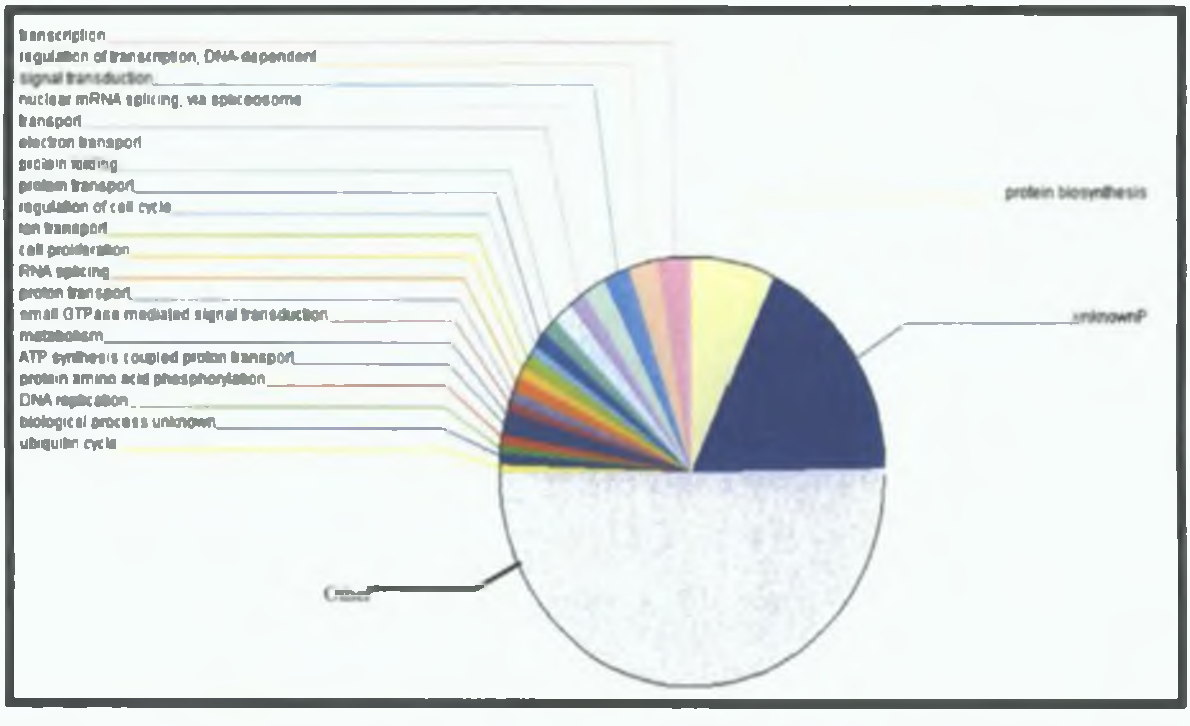


Fig 3.7.7.1 Gene ontology of the biological process related gene transcripts common to DLKP CM and cell samples. UnknownP = unknown process, Other = categories with less than 20 gene transcripts.

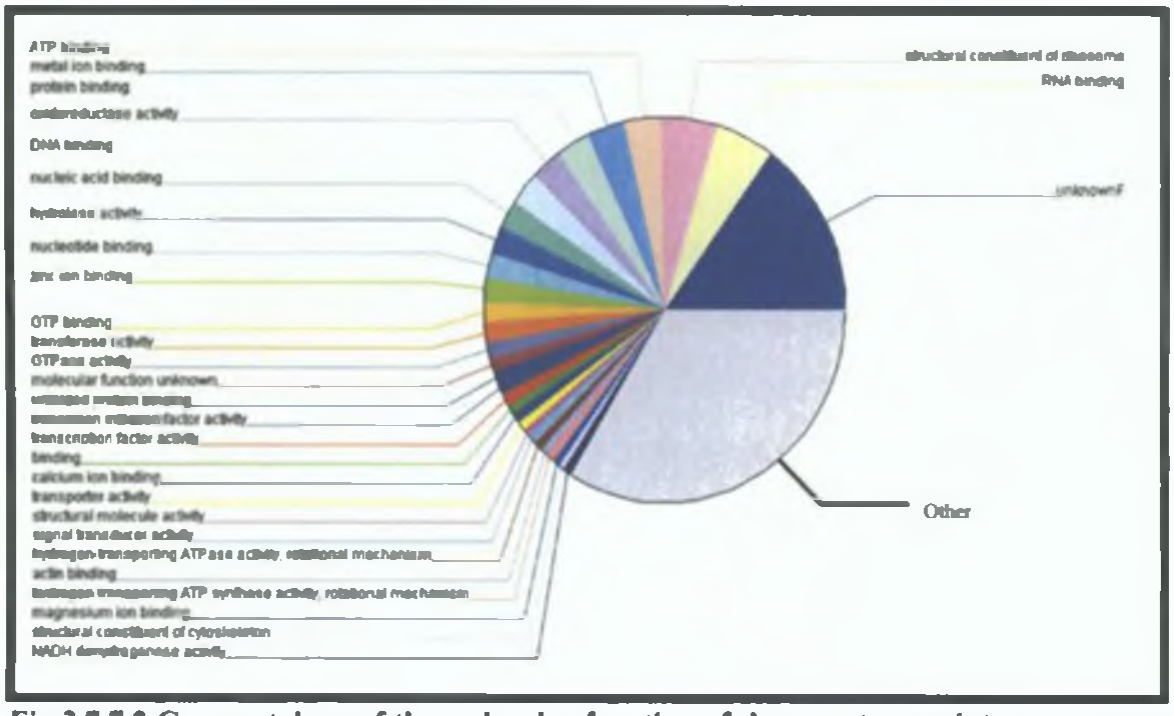


Fig 3.7.7.2 Gene ontology of the molecular function of the gene transcripts common to DLKP CM and cell samples. UnknownF = unknown function, Other = categories with less than 20 gene transcripts.

The aim of Section 3.8 was to assess the changes in CM samples taken from a panel of parental and drug-selected cell lines by microarray to see if any CM specific gene expression profiles were obtainable and also whether any potential markers for drug resistance could be identified. This also involved validation of the microarray Presence using a second technique, RT-PCR, to confirm Present or Absence call assigned and/ or trend of change in expression.

3.8 Analysis of mRNA Expressed By Various CM Using Microarray

The trial run using DLKP CM, to see if it was possible to get valuable array information from CM RNA proved to be successful, so the study was increased to investigate differences in mRNA expression across cell lines. Groups of parental human cancer cell lines and their drug selected variants were cultured (in triplicate and at the same passage number) and their CM was harvested after 48hours and filtered with a 0.45μm filter. The RNA was then extracted and amplified, labelled and hybridised to Affymetrix U133A Plus 2 chips (as described in Section 2.12)

The CM samples were taken from eight different cell lines; NCI H1299 (H1299) and its taxol selected variant (H1299 TAX), RPMI 2650 (RPMI) and its taxol (RPMI TAX) and melphalan (RPMI ML) selected variants and MDA-MB-435-F (MDA-F) and its super invasive adriamycin (MDA-F-ADR-SI) and taxol (MDA-F-TAX-SI) selected variants (see Table 2.5.1 in Section 2.5 for details of cell lines).

One CM sample, RPMI TAX 2, did not yield the minimum requirement of 15μg of cRNA after two attempts of amplification and so was excluded from the study and the remaining 23 CM cRNA samples were hybridised to the Affymetrix Human Genome Plus 2.0 chips.

The following Section contains details of the quality control parameters analysed, scatter plot and Venn diagrams of gene expression, hierarchical clustering of the samples, generation of lists containing significantly changing gene transcripts, gene ontology and RT-PCR studies to assess if the data found using microarray techniques could be reproduced with RT-PCR.

3.8.1 cRNA yield throughout amplification procedure

The average yield of labelled cRNA obtained in this study was 53.5µg± 26.6 (ranging from

15µg-103µg per CM). With the exception of RPMI 2 and H1299 3 CM samples, the cRNA yield obtained for the CM samples were as good as, but mostly much higher than the cRNA yields obtained with the DLKP CM samples (see Table 3.8. 1.1).

Sample		After 1st IVT(ng)	After 2nd IVT(µg)
H1299	1	356.4	39.39
	2	369.9	30.62
	3	420.3	15.04
H1299 TAX	1	363.6	50.45
	2	325.8	55.46
	3	363.6	23.75
RPMI	1	253.8	28.32
	2	268.2	6.97
	3	329.4	28.91
RPMI TAX	1	424.8	49.39.
	2	471.6	69.57
	3	531.9	75.85
RPMI ML	1	476.1	51.12
	2	531	30.03
	3	311.4	34.03
MDA-F	1	431.1	65.74
	2	392.4	78
	3	425.7	92.86
MDA-F-TAX-SI	1	425.7	99.86
	2	530.1	103.91
	3	457.2	48.77
MDA-F-ADR-SI	1	447.3	96.13
	2	298.8	35.95
	3	439.2	74.05

Table 3.8.1.1 Quantity of cRNA generated with 2 rounds of linear amplification (IVT= in vitro transcription)

3.8.2 Microarray QA results for CM samples

The quality control parameters listed in Table 3.8.2.1 show that the scaling factor, noise and background are all at acceptable levels except in the case of samples MDA-F 1 and MDA-F-TAX-SI 2. Both of these samples have high background and noise levels compared to the other samples. They also have lower scaling factors and together these correlate with a decrease in % Present call.

MDA-F-TAX-SI 2 was excluded from further data analysis as it had a scaling factor more than 3 fold different than the other samples in the group in case it biased the results. The MDA-F 1 sample however, although having background levels higher than the other CM samples, was within the accepted range of 20-100, the scaling factors were within a 3 fold difference of the other CM samples so the sample was included. MDA-F-ADR-SI 1 CM sample had a %Present call of 19.4%, which was higher than the other CM samples but as all QA parameters were acceptable it was included.

In general, the present calls for each of the samples were considerably lower than were achieved with the DLKP CM, ranging from 4.8- 19.4% with an average % Present call of 10% compared to an average %Present call of 23.1% with the DLKP CM samples. This may have been due to the time delay in processing the CM once it was taken off the cultured cells. There was a time delay of up to twenty minutes before the filtered CM was added to Tri Reagent because these samples were being collected from cell lines cultured by other researchers. With the DLKP CM samples, there was no time delay and so perhaps more unstable mRNAs were protected from degradation by quick sample processing, i.e. addition of Tri Reagent which inhibits RNase activity.

DChip software was used to normalise the entire set of CM microarray data before further analysis of the data was carried out (see Section 2.13.8.2).

	H1299			H1299 TAX			MDA-F			MDA-F-TAX-SI			MDA-F-ADR-SI		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Scaling Factor	22.5	16.8	20.25	18.34	16.9	17.14	9.86	16.9	14.92	20.44	6.994	24.84	11.18	20.92	15.29
Noise	1.07	1.2	1.09	1.25	1.24	1.15	2.49	1.36	1.4	1.27	3.35	1.06	1.32	1.05	1.59
Background	33.41	36.07	35.01	39.08	38.49	35.7	73.82	40.11	43.51	40.21	113.2	33.33	40.04	34.94	47.82
% Pres	9.7	12.3	11.1	10.4	11.8	11.7	6.3	11.2	11.8	8.5	4.8	8.1	19.4	10.5	8.9
GAPDH 3'/5'	_	3.65	6.74	4.5	11.87	5.73	_	1.77	2.79	_	_	3.13	8.21	5.35	4.81
GAPDH 3'/M	_	4.03	_	_	3.65	_	_	_	_	_	_	_	2.14	_	_

	RPMI		RPMI TAX			RPMI ML		
	1	3	1	2	3	1	2	3
Scaling Factor	24.52	30.5	18.21	18.2	16.37	19.61	20.6	22.9
Noise	1.09	1.02	1.31	1.34	1.4	1.08	1.04	1.15
Background	35.64	32.32	40.62	40.11	42.2	34.01	32.63	35.98
% Pres	7.7	5.9	9.9	10.3	11.7	10.5	10.2	8.3
GAPDH 3'/5'	1.56	2.19	_	_	_	_	_	1.47
GAPDH 3'/M	-	-	-	-	-	-	-	-

Table 3.8.2.1 Quality control results for Affymetrix chips in CM, numbers highlighted indicate background values that are unusually high in comparison to the other CM samples shown.

3.8.3 Expression of Two-Cycle Target Hybridisation and Labelling Controls

All hybridisation controls graphed in Fig.3.8.3.1 were called Present in all of the 23 CM samples. Figure 3.8.3.1 shows that the 5' and 3' probes are also expressed at similar levels across all 23 samples, which indicate that, the hybridisation, washing and staining procedures were without incidence. The targets were detected according to the ratio of their concentration in the initial hybridisation cocktail i.e. *cre* was the most abundant followed by *bioD, bioC* and *bioB.*

The Poly- A controls are *B.* subtilis genes that have been modified by the addition of poly-A tails and cloned into pBluescript vectors, further processing yields sense RNAs which are spiked into the RNA samples before the 2 cycle amplification procedure is carried out to help evaluate the amplification process. The Poly-A controls were called Present in all 23 of the CM samples. Figure 3.8.3.2 shows that the expression of the 3' target probes (yellow lines) is higher in all samples than the expression of the 5' (dark blue lines) or middle probes (pink lines). This indicates that there was a 3' bias in the amplification procedure.

The average expression of the Poly-A controls in all CM samples is shown in Fig. 3.8.3.3 and indicates that the ratios at which the Poly-A controls (detected by microarray) to each other do not correlate fully with the ratios in which they are present in the original stock solution (before amplification).

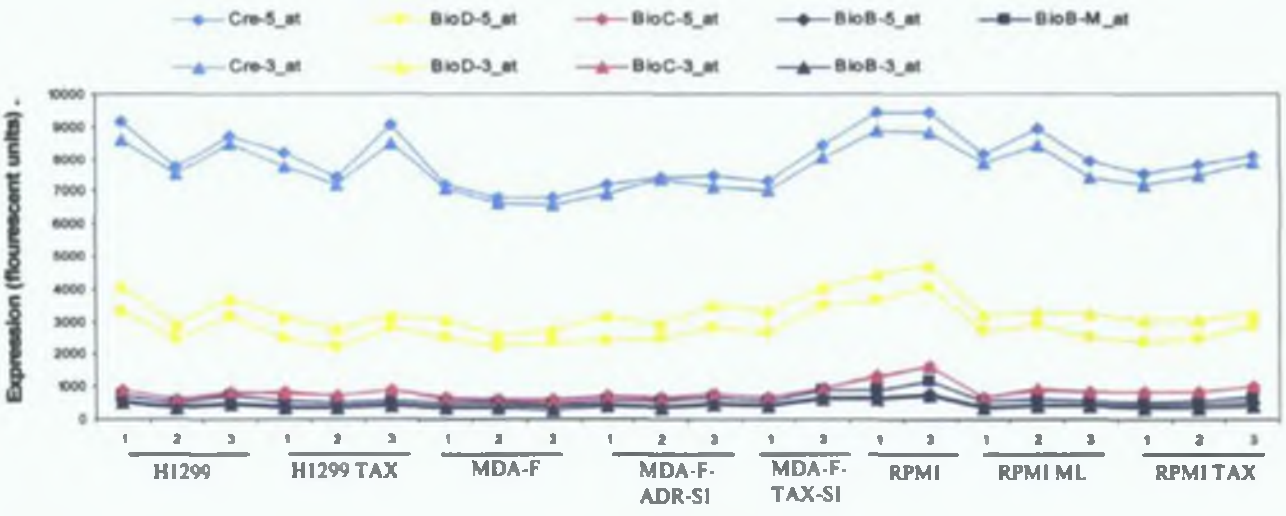


Figure 3.8.3.1 Expression levels of Hybridisation controls. CM samples are plotted on the X-axis with the expression level of each hybridisation control plotted on the Y-axis. Cre, BioD, BioC and BioB are the hybridisation spike-in controls and the suffix denotes the region targeted by the probe set i.e. 3_at targets the 3' end, M_at targets the middle and 5_at targets the 5' end.

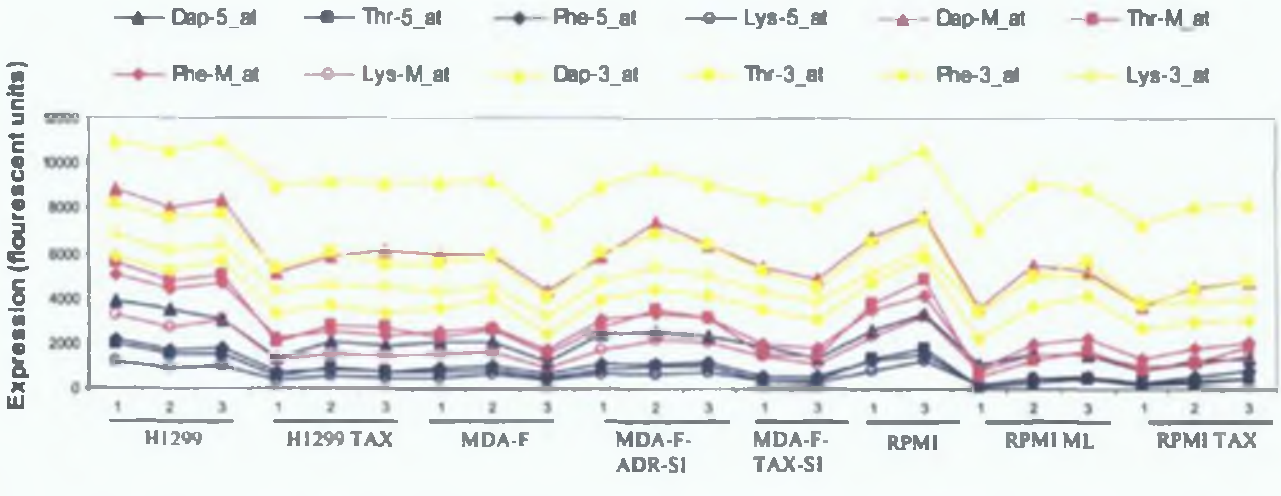


Figure 3.8.3.2 Expression levels of Poly A controls. CM samples are labelled on the X-axis with the expression level of each Poly A- control plotted on the Y-axis. Dap, Thr, Phe, Lys are the Poly-A spike in controls (see Section 2.12.1) and the suffix denotes the region targeted by the probe set i.e. 3_at targets the 3' end, M targets the middle and 5_at targets the 5' end.

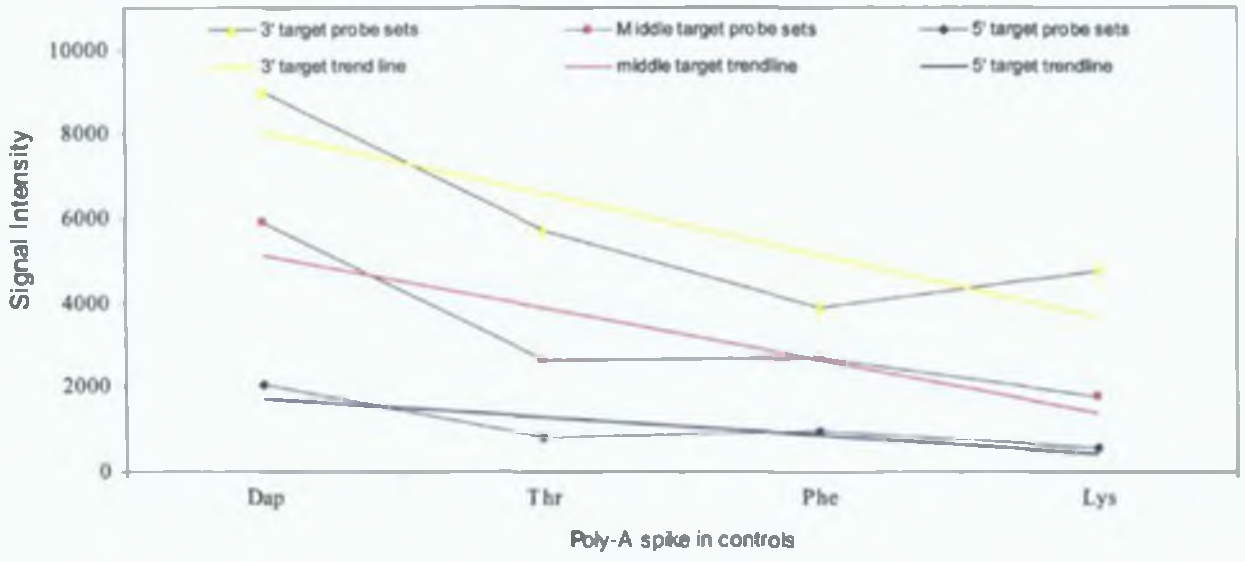


Figure 3.8.3.3 Average expression of Poly-A controls in 23 CM samples. Affymetrix Poly-A controls (see Section 2.12.1) are plotted on the X-axis and the average expression for the controls plotted on the Y axis. The three linear lines (yellow, pink, and dark blue) are the trend lines for each data set and the black lines connect the actual relative expressions for the 5' Δ middle ■, and 3'◆ targeted probe sets

Overall the ratio of the expression levels of the 3' probe sets to either the middle or 5' probe sets were more than 1:1. The average 3' to 5' ratio for all the *dap* and *phe* Poly-A controls is 4:1, and the average 3' to 5' ratios for the *lys* and *thr* Poly-A controls is higher at 7- 8:1. The ratios of the Poly-A controls to each other in each sample should be *dap>thr>phe>lys*. The trend lines in Fig 3.8.3.3 suggest that the ratios of the Poly-A controls to each other match the ratios in the original pre-mixed control, but the actual relative expression of each Poly-A control to each other tells a different story.

The ratios of the 3' targeted Poly-A probes to each other are most similar to the ratios in the original pre-mixed controls (with the exception of *lys*), but they are not as high as they should be (see Section 2.12.1).

A good example of this is with the *dap* Poly A control, in the original spike- in cocktail, it is present at a concentration over three times higher than the *thr* Poly-A control. However from the data generated in the microarray study, out of the 5', middle and 3'

targeted probe sets, only the expression of the *dap* 5' probe set compared to the *thr* 5' probe set comes close to this ratio with an expression level of 2058 compared to 794. The ratios of the expression of the *dap* M and 3' probe sets to the *thr* M and 3' probe sets are 2.2:1 and 1.6:1 respectively instead of 3:1.

Because the Poly-A controls are obtained commercially pre-mixed and thus leaving no chance for the wrong relative ratios to be spiked in, any discrepancies between the relative ratios obtained and the original ratios is attributable to either the amplification procedure itself and/ or different hybridisation efficiencies for the target probe sets.

In summary, there were no problems with the hybridisation of the chips but the poly-A spike-in controls were not amplified to a level that reflected their original concentration. There was also a 3' amplification bias in the 2-cycle amplification procedure, but as the Affymetrix probes target the 3' end of gene transcripts this should not be problematic.

3.8.4 Venn Diagrams of probe set expression for CM samples

The microarray data obtained for these samples was normalised using dChip before being imported into Genespring. Genespring enables visual interpretation of the data using Venn diagrams and scatter plots.

The scatter plot of expression levels of the probe sets in the replicate samples follow the same pattern as for the DLKP CM samples (see Section 3.7.3). The probe sets expressed at levels under 100 are quite variable between the replicate samples resulting in a large spread of dots (which represent the expression level of a particular probe set in each individual sample). At higher levels of expression, the probe sets are more uniformly expressed by the replicate samples, as indicated by the fact that the probe sets centre on a central line, meaning that they are expressed at similar levels in the three replicate samples. The scatterplots for H1299 are shown as an example (see Fig. 3.8.4.1 and 3.8.4.2).

Two different Venn diagrams were also generated using the microarray data for each group of replicate samples. The first overlaps the probe sets called Present (P) or Marginal (M) in each individual sample with the replicate samples and the second Venn diagram overlaps the probe sets expressed at levels over 100 in the individual samples with its replicate samples. The probe sets called P or M in all replicate samples for those groups are in the centre of the Venn diagram. Table 3.8.4 summarises the data shown in the Venn diagrams (Fig. 3.8.4.1-3.8.4.3) for all of the CM samples and shows that, in all cases, there is a higher percentage of overlap between replicate samples when only probe sets with an expression value over 100 are included.

H1299

Normalised data with Present or Marginal call

29,359 gene transcripts are called Absent (A) in all three H1299 CM samples (see Fig 3.8.4.1). 3,269 probe sets are called Present (P) or Marginal (M) in all three samples, which represent just fewer than 13% of the probe sets called P or M in at least one of the three H1299 CM samples. An average of 5,000 probe sets are called P or M in one H1299 CM sample with a further average of 2,400 probe sets P or M in at 2/3 H1299 CM samples. The scatter plot on the left of Fig. 3.8.4.1 shows that the probe sets called P or M in all 3 H1299 CM samples are not all highly expressed, rather they are mixed through the scatter plot. It is interesting that a number of highly expressed probes (signal intensity between 100- 1000) also have an A call in all 3 CM samples, as denoted by the grey dots.

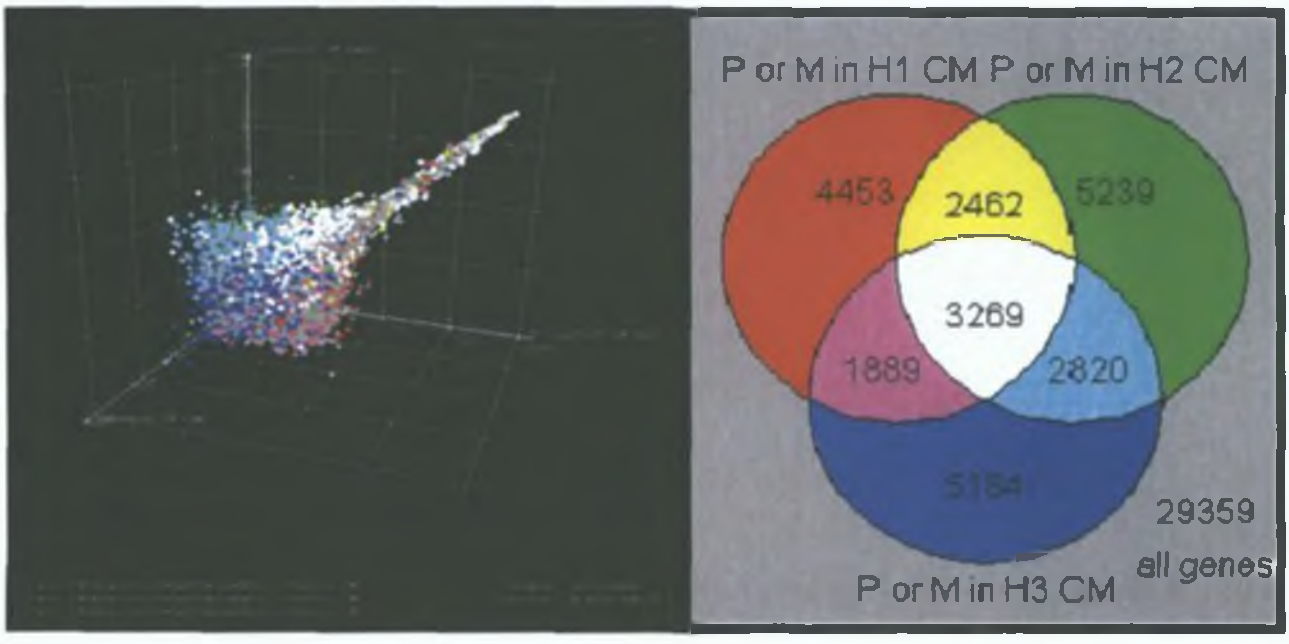


Figure 3.8.4.1 Scatter plot and Venn diagram of probe sets called P or M for H1299 CM. In the scatter plot (left-hand side) H1299 1 CM (H1) is plotted on the X-axis, H1299 2 CM (H2) on the Y-axis and H1299 3 CM (H3) on the Z-axis. Each dot represents a probe set and the colours denote the group in which the probe set belongs in the Venn diagram (right-hand side). P or M = Present or Marginal call

Normalised data with expression >100

When an expression cut- off of 100 expression units is applied to the data the number of probe sets that come through this filter drops by nearly 12 fold compared to the number of probe sets with a P or M call from 25, 316 probe sets to 2,149 probe sets. However the percentage of probe sets with expression over 100 common to the three H1299 CM samples rises to 60% compared to ~13% of probe sets called P or M common to the three H1299 CM samples (Fig. 3.8.4.1). The scatter plot shows that the probe sets with expression levels over 100 common to the three samples (white dots), tend to be similar in the three samples as denoted by the way the white dots cluster more tightly around a central line than for example, the grey dots, which represent the 52, 526 probe sets with expression levels under 100 in all three H1299 CM samples and show wide variation denoted by the bell shaped curve of the data points.

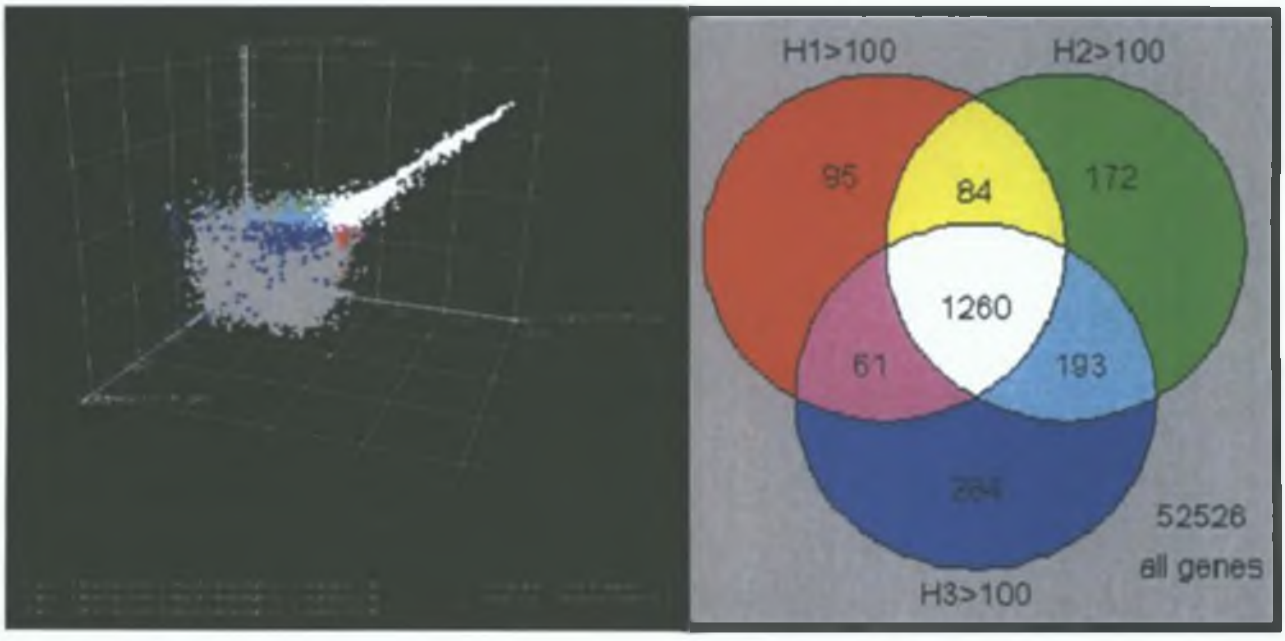


Figure 3.8.4.2 Scatter plot and Venn diagram of probe sets with expression >100 for H1299 CM samples. In the scatter plot (left-hand side) H1299 1 CM (H1) is plotted on the X-axis, H1299 2 CM (H2) on the Y-axis and H1299 3 CM (H3) on the Z-axis. Each dot represents a probe set and the colours denote the group in which the probe set belongs in the Venn diagram (right-hand side).

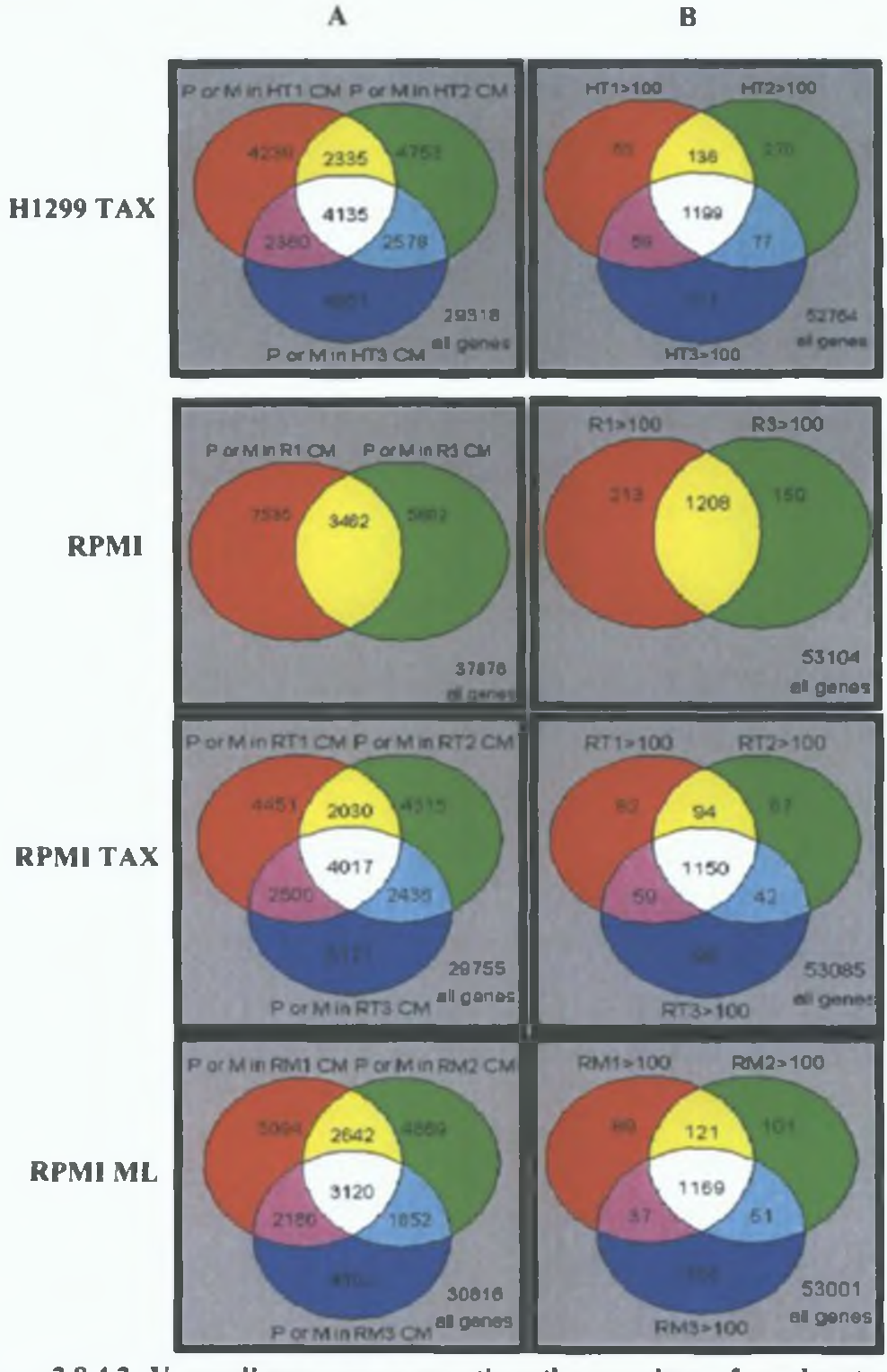


Figure 3.8.4.3 Venn diagrams representing the number of probesets in the individual samples that are similarly (A) called Present or Marginal (P or M) in the replicate samples or (B) expressed at levels >100 in the replicate samples. HT= H1299 TAX, R= RPMI, RT= RPMI TAX, RM= RPMI ML.

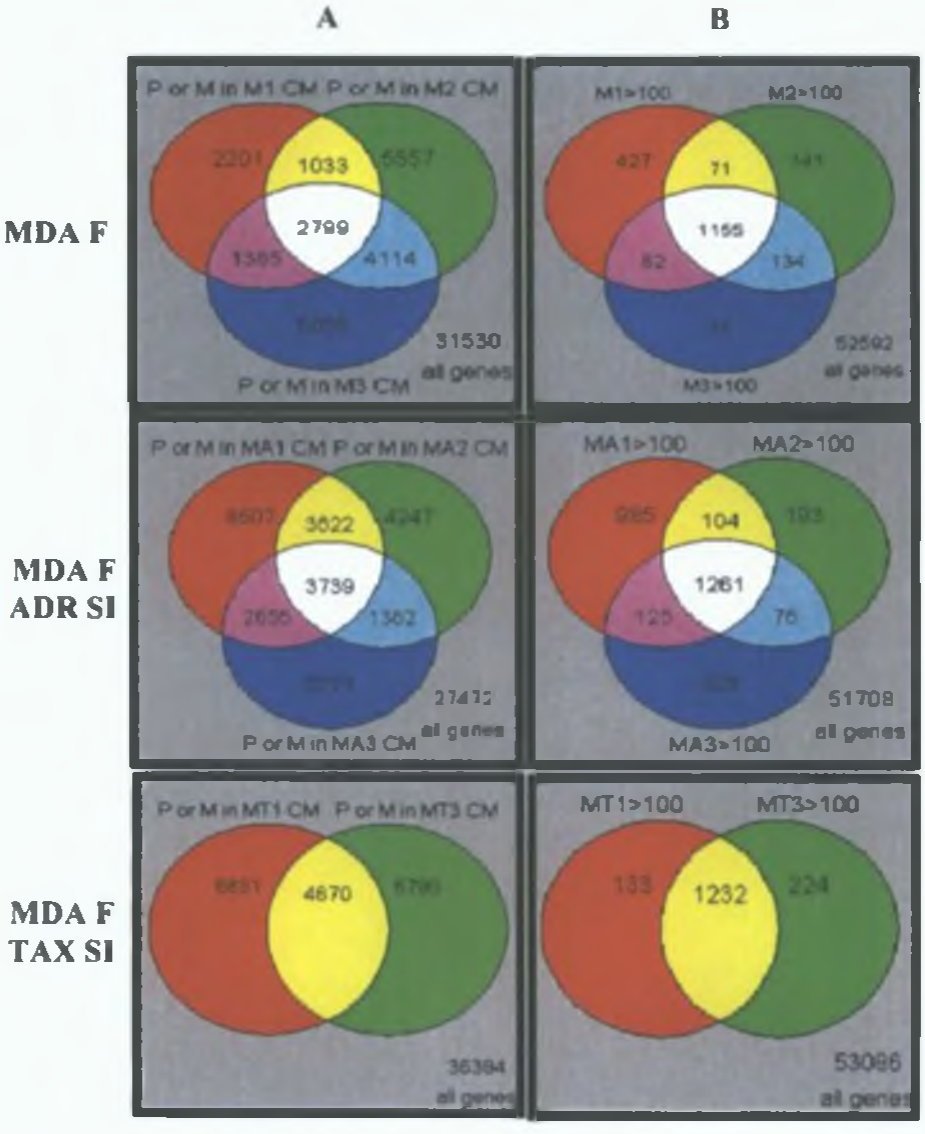


Figure 3.8.4.3 contd. Venn diagrams representing the number of probesets in the individual samples that are similarly (A) called Present or Marginal (P or M) in the replicate samples or (B) expressed at levels >100 in the replicate samples. M= MDA F, MA= MDA F ADR SI, MT= MDA F TAX SI

% overlap between replicate samples		
	P or M	exp>100
H1299	13	60
H1299 TAX	16	63
RPMI	25	77
RPMI ML	13	70
RPMI TAX	16	72
MDA F	12	55
MDA F ADR	14	45
MDA F TAX	20	77
Average	16.125	64.875

Table 3.8.4.1 Summary of the percentage of probe sets either called Present or Marginal (P or M) or expressed at levels >100 (exp>100) in all replicate samples for each group

3.8.5 Gene Ontology of probe sets called Present in all CM samples

To get an idea of the molecular function of the gene transcripts called Present in all 23 CM samples (335 probe sets), they were loaded into Netaffx online software so that Gene Ontology could be assigned (see Section 2.13.9.3). As annotated gene transcripts can have multiple functions, only the primary molecular functions/ parent category (see Fig. 2.13.11 for example) of the gene transcripts assessed were recorded.

As categories of molecular function can be subdivided many ways e.g. binding can be divided into nucleotide binding, ion binding etc, it was decided to chart only the main branch of molecular function, thus avoiding counting the same probe set into many different groups and many different subdivided branches. Although some probe sets were still counted in more than one group, the pie chart in Fig. 3.8.5.1 still serves to give an overall impression of the molecular function of the gene transcripts corresponding to the probe sets called Present in all CM samples.

Just under half (165) of the 335 probe sets were un-annotated. Fig. 3.8.5.1 shows that the molecular function of most of the annotated gene transcripts involves binding. This binding function includes nucleotide-, nucleic acid-, DNA-, damaged DNA-, RNA-, receptor-, integrin-, ion/ cation-, protein-, ATP-, GTP- and transcription factor binding.

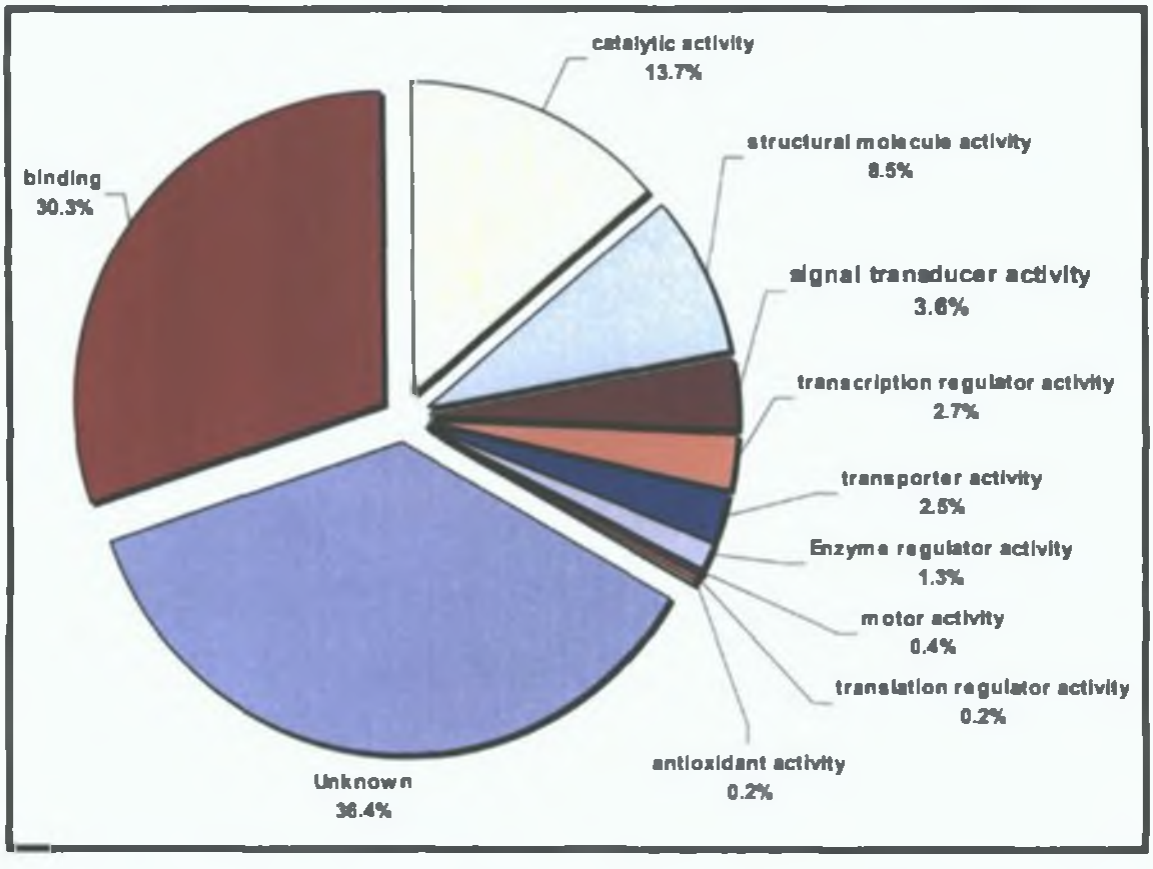


Figure 3.8.5.1 Gene Ontology of probe sets called Present in all CM samples (n= 335)

3.8.6 Hierarchical clustering of CM samples

Hierarchical clustering analysis was carried out on all the CM samples using both Genespring and dChip software to find out how similar the samples are to each other. The Hierarchical clustering of the samples in Genespring was carried out based on the expression of all gene transcripts (54,675) using a Pearson correlation.

Using dChip software, a filter was applied to the CM microarray data to firstly filter out any probe sets whose expression did not change across the samples. The criteria employed required that the ratio of the standard deviation and the mean of a gene's expression values across all samples be greater than a certain threshold. This ratio is also known as Coefficient of Variation (CV). The default formula used in this study was that the variation across the samples (after pooling replicate arrays) was 0.5< (Standard deviation/ Mean) <1000. The more variable a gene is across samples, the larger the ratio is. This is unsupervised sample clustering as the gene transcripts are selected by large variation across samples and the sample group information is not used. A similarity measure was then employed to cluster the samples together.

Genespring was able to cluster at least 2, if not three, of the replicate samples for each sample group beside each other for 5/ 8 groups (Fig. 3.8.6.1).

dChip was able to cluster the different sample groups better than Genespring (Fig. 3.8.6.2). Five groups were classed as significant sample clusters, RPMI TAX (p-value: 0.022727), RPMI ML (p-value: 0.006494), MDA-F-ADR-SI (p-value: 0.002597), H1299 (p-value: 0.012987) and H1299 TAX (p-value: 0.012987). Only the MDA F 1 (M1) sample is out of place, this is possibly due to the fact that the background levels for this chip were much higher than the rest of the samples included in this analysis (see Section 3.8.2).

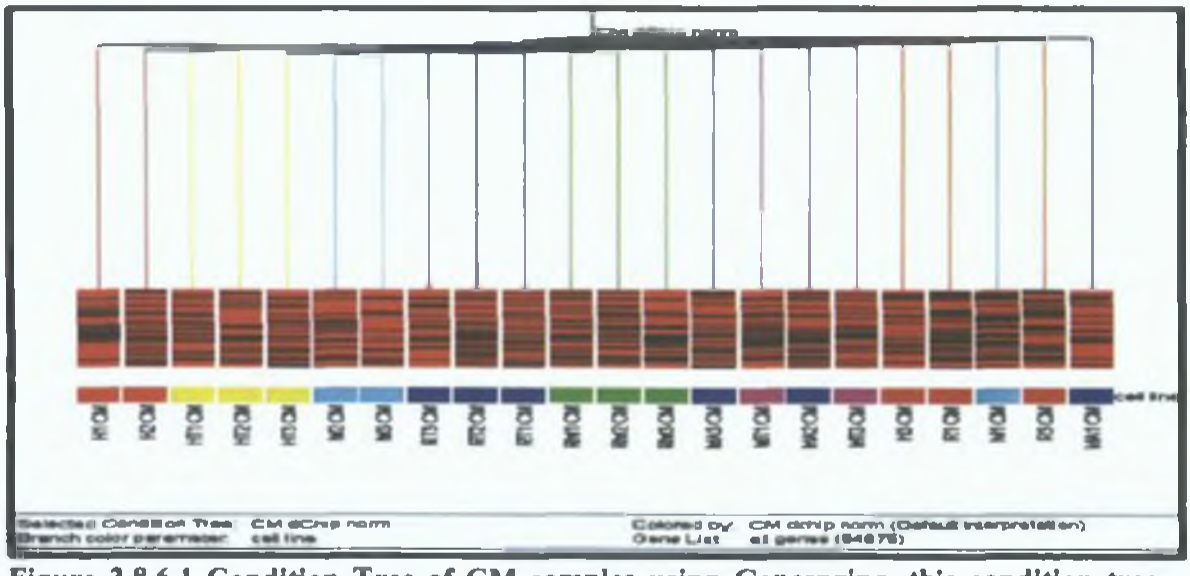


Figure 3.8.6.1 Condition Tree of CM samples using Genespring, this condition tree was performed using a Pearson correlation. H= H1299, HT= H1299 TAX, M= MDA-F, MA= MDA-F-ADR-SI, MT= MDA-F-TAX-SI, R= RPMI, RM= RPMI ML, RT= RPMI TAX.

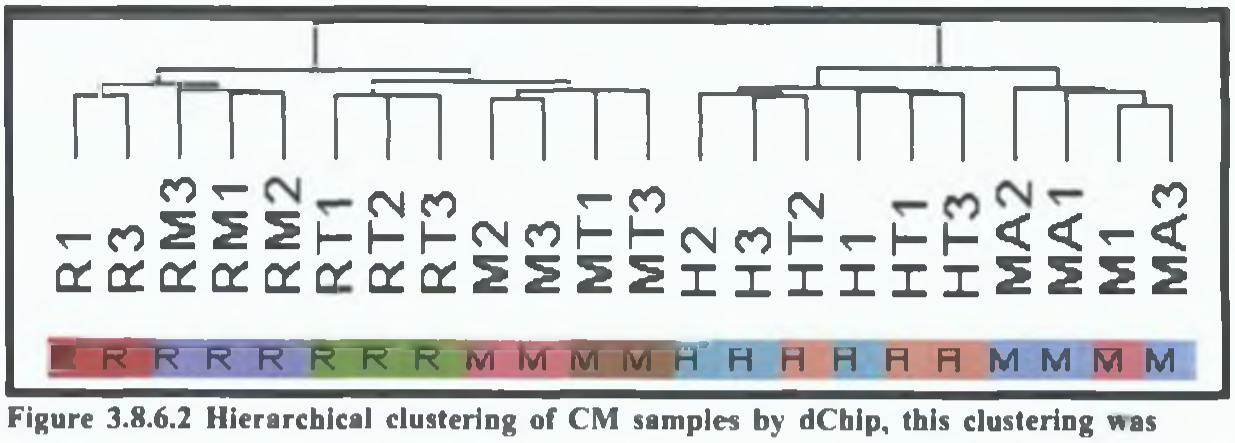


Figure 3.8.6.2 Hierarchical clustering of CM samples by dChip, this clustering was carried out using a similarity measure. Gene transcripts that did not change in expression across all the CM samples were removed and the similarity of the CM samples computed on the 21, 357 remaining gene transcripts. H= H1299, HT= H1299 TAX, M= MDA-F, MA= MDA-F-ADR-SI, MT= MDA-F-TAX-SI, R= RPMI, RM= RPMI ML, RT= RPMI TAX.

3.8.7 Validation of Microarray Data (Part 1)

In Affymetrix arrays there are a set of probe sets designed to target different traditional housekeeping gene transcripts e.g. β-actin and GAPDH. Three different probe sets are designed so that they align to the 5', middle and 3' end of these gene transcripts. In calculating the integrity of the cRNA used for the arrays, the ratio of the expression of the 3' probe set to either the middle or 5' probe set is used. A ratio of 1:1 is ideal but anything up to 3:1 is acceptable. As we had previously designed mRNA specific primers for different regions of the β-actin mRNA transcript, an attempt to validate the Presence/ Absence call assigned to the samples was carried out using these primers.

In addition to this, two transcripts were chosen, one that had a Present call assigned to all 23 of the CM samples and the other which had an Absent call assigned to all 23 of the CM samples. RT- PCR was used to test for amplification of either of these transcripts. Original unamplified RNA was used as a template for reverse transcription in all cases.

3.8.7.1 Expression of "housekeeping gene transcripts" in CM RNA arrays

The expression levels of the GAPDH probe sets in Fig. 3.8.7.1.1 show that according to the microarray data, the expression of GAPDH is slightly higher in the 3' region of the mRNA transcript especially in the H1299/ H1299 TAX and MDA-F-ADR-SI 1 CM samples. In all other CM samples, the expression is negligible.

Although the expression of the 5' region of GAPDH mRNA, in Fig. 3.8.7.1.1 is very close to 0, the accompanying Present/ Absent call data for this probe set indicates that in 16/ 22 of the samples this probe set is assigned "Present" (Table 3.8.7.1). 17/22 of the samples have a Present call assigned to the probe set that targets the 3' end of the GAPDH mRNA, and more interesting is that fact that only 3/22 CM samples have a Present call assigned to

the probe set that targets the middle of the mRNA transcript.

A similar trend in expression is seen in the probe sets that target the different regions of the β-actin gene transcript i.e. there seems to be higher expression of the 3' region of the mRNA than the 5' or middle region (Fig. 3.8.7.1.2). With the exception of the 5' targeted probe set in MDA-F-TAX-SI 1 and the middle region targeted probe set in MDA-F-ADR-SI 1, all probe sets targeted to the 5' and middle region of the β-actin mRNA transcript are called Absent. The probe set that targets the 3' end of the β-actin gene transcript are called Present in 22/22 CM samples.

For the CM samples, calculation of both housekeeping gene 3'/M or 3'/5' ratios were not carried out because the call assigned to the probe sets was Absent more than it was Present (see Tables 3.8.7.1/ 2 for details).

GAPDH

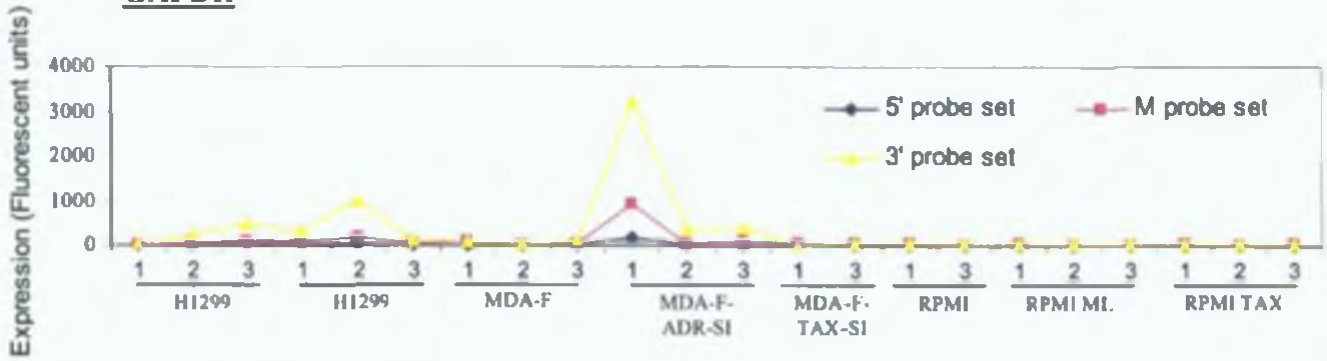


Figure 3.7.1.1 Expression levels (Y-axis) of GAPDH probe sets in CM samples (x-axis). Blue line = 5' targeted probes, pink line = middle region targeted probes, yellow = 3' targeted probes.

	H1299			H1299 TAX			MDA-F			MDA-F-ADR-SI			MDA-F-TAX-SI		RPMI		RPMI ML			RPMI TAX		
	1	2	3	1	2	3	1	2	3	1	2	3	1	3	1	3	1	2	3	1	2	3
5'	A	P	P	P	P	P	A	P	P	P	P	P	P	P	P	P	A	A	P	A	P	A
M	A	P	A	A	P	A	A	A	A	P	A	A	A	A	A	A	A	A	A	A	A	A
3'	A	P	P	P	P	P	P	P	P	P	P	P	A	P	P	P	A	A	P	P	A	P

Table3.7.1.1 Present/ Absent calls assigned to GAPDH probe sets for CM samples.

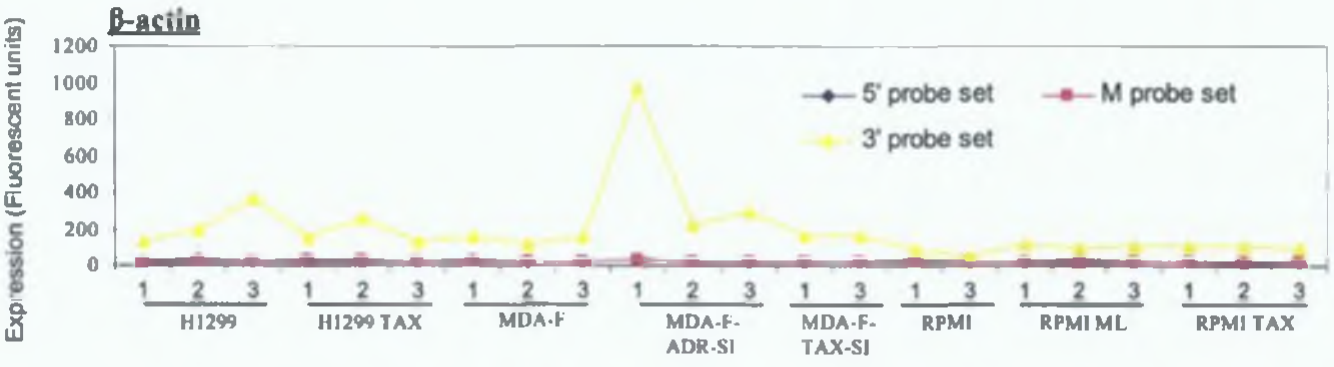


Figure 3.7.1.2 Expression levels (Y-axis) of β- actin probe sets for CM samples (X-axis). Blue line = 5' targeted probes, pink line = middle region targeted probes, yellow = 3' targeted probes.

	H1299			H1299 TAX			MDA-F			MDA-F-ADR-SI			MDA-F-TAX-SI		RPMI		RPMI ML			RPMI TAX		
	1	2	3	1	2	3	1	2	3	1	2	3	1	3	1	3	1	2	3	1	2	3
5'	A	A	A	A	A	A	A	A	A	A	A	A	P	A	A	A	A	A	A	A	A	A
M	A	A	A	A	A	A	A	A	A	P	A	A	A	A	A	A	A	A	A	A	A	A
3'	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P

Table 3.7.1.2 Present/ Absent calls assigned to β-actin probe sets for CM samples.

3.8.7.2 RT-PCR for β-actin using 5' and 3' specific primers

It was decided to carry out RT-PCR on the unamplified RNA CM samples using primers designed for a previous study (see Section 3.3) that amplify different regions of the β-actin gene transcript and 45 cycles of PCR to see if the Presence/ Absent call assigned with the microarray data could be correlated to RT-PCR results.

The results in Figs. 3.8.7.2.1-3, show that whilst detection of β-actin gene transcripts corresponding to the 5' UTR region of the β-actin mRNA was only possible in 6- 14/ 23 CM samples, PCR products were amplifiable in all CM samples targeted by 5' coding and 3' region specific primers. (Expression levels of β-actin as determined by microarray were not possible for the MDA-F-TAX-SI 2 CM sample as it was excluded before normalisation because of bad QA results (see Section 3.8.2) but it was included in this RT-PCR study).

These results however may be a consequence of the 3' bias of the 2 cycle amplification procedure as detailed in Section 3.8.3. As the Affymetrix probes target the 5' and middle regions of the housekeeping gene transcripts as well as the 3' region, it is possible that the Absent call assigned to these regions is a direct result of this bias.

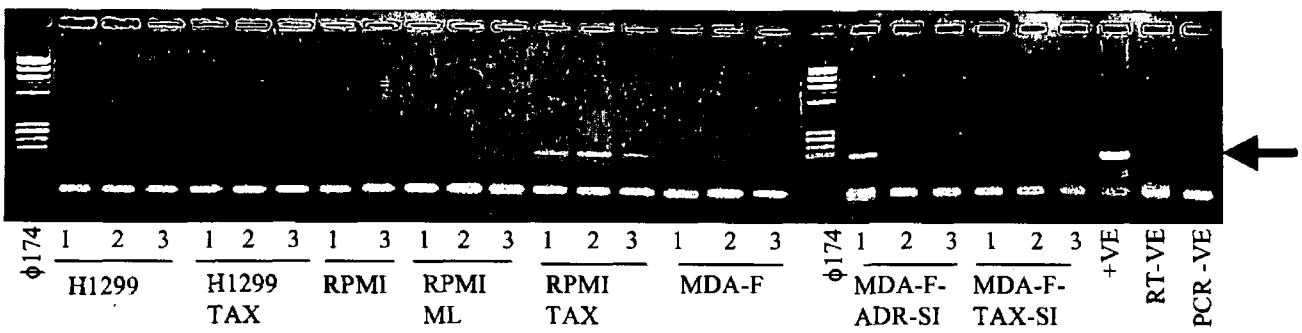


Figure 3.8.7.2.1 Expression of 5'UTR specific β-actin mRNA. β-actin gene transcript (171bp) (arrow) was detected in RPMI 1, RPMI TAX 1, 2 and 3, MDA-F-ADR 1 and MDA-F-TAX 3 CM samples with faint bands visible in at least 8 other CM samples. Amplified product was detected in the positive control and was undetected in the negative control samples (-ve RT = RT reaction with H_20 instead of RNA as control; -ve PCR = PCR reaction with H_20 instead of cDNA as control).

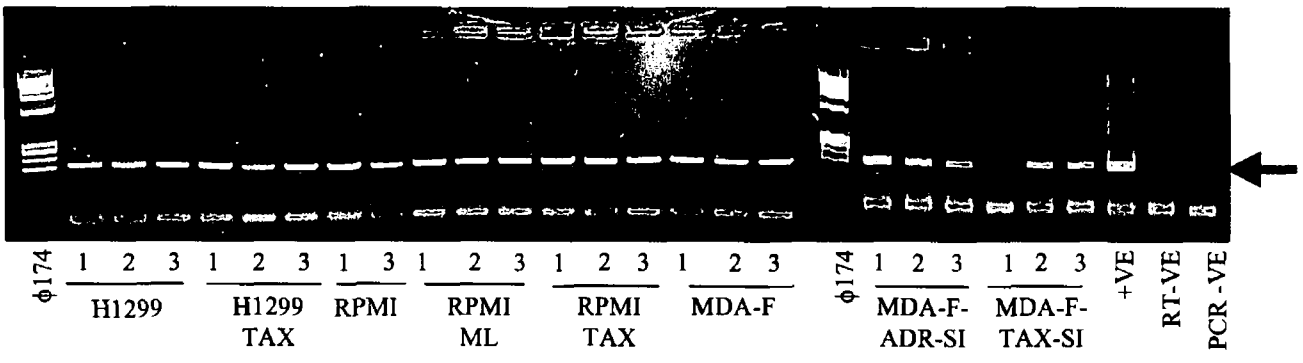


Figure 3.8.7.2.2 Expression of 5'CDS specific β-actin mRNA. β-actin gene transcript (arrow) was detected in all CM samples tested. Amplified product was detected in the positive control and was undetected in the negative control samples (-ve RT = RT reaction with H_20 instead of RNA as control; -ve PCR = PCR reaction with H_20 instead of cDNA as a control).

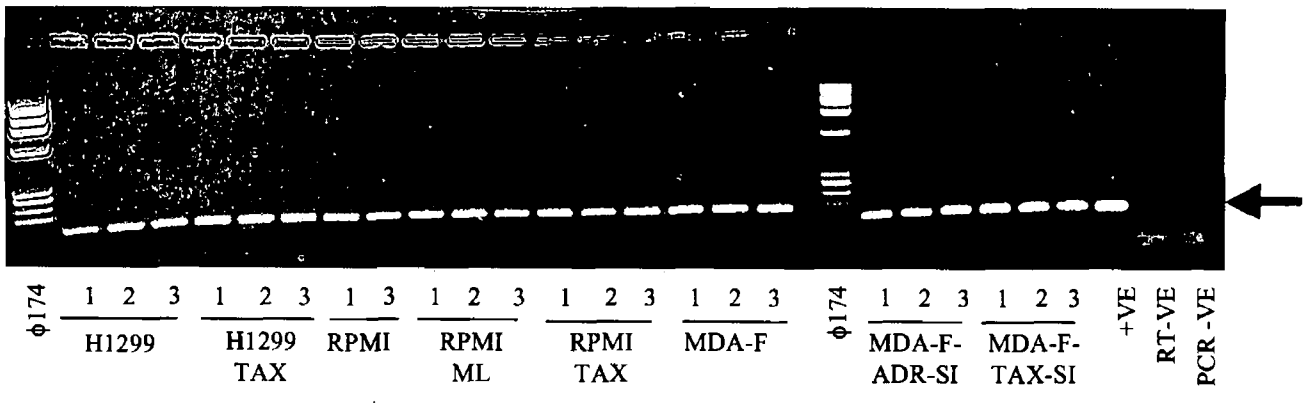


Figure 3.8.7.2.3 Expression of 3' specific β-actin mRNA. β-actin gene transcript (arrow) was detected in all CM samples tested. Amplified product was detected in the positive control and was undetected in the negative control samples (-ve RT = RT reaction with H_20 instead of RNA as control; -ve PCR = PCR reaction with H_20 instead of cDNA as a control).

3.8.7.3 Validation of Presence/ Absence call

It was decided to investigate the correlation of Present/ Absent calls assigned to other mRNA transcripts by microarray with RT-PCR. To do this, two transcripts were chosen, one with a Present call assigned to the probe set in all CM samples, and the other an Absent call assigned to the probe set in all CM samples.

The results shown in Figs. 3.8.7.3.1/ 2 confirm the results previously found (in Section 3.8.7.2) that mRNA gene transcripts assigned as Absent can be detected by PCR. In all 22 CM samples were the detection call assigned to the probe set targeting APAF 1 interacting protein mRNA was Absent, detection of the gene transcript was possible with 45 cycles of PCR. With the probe set that was called Present in all CM samples (220232_at), detection of its target mRNA by RT-PCR was possible.

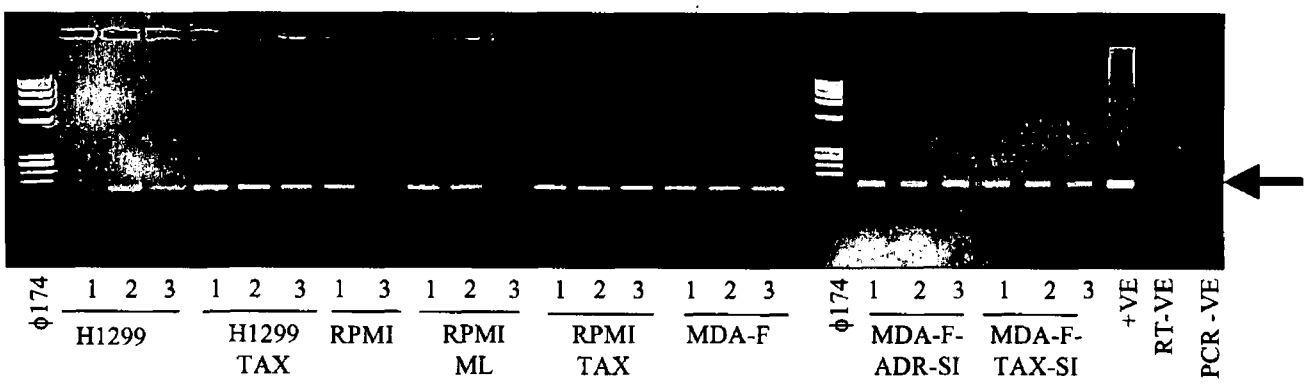


Figure 3.8.7.3.1 RT-PCR of all CM for stearoyl-CoA desaturase 5 (151bp). Stearoyl-CoA desaturase 5 gene transcripts (arrow) were detected in all CM samples tested. The probe set corresponding to Stearoyl-CoA desaturase 5 mRNA (220232_at) is called Present in all CM samples. Amplified product was detected in the positive control and was undetected in the negative control samples (-ve RT = RT reaction with H_20 instead of RNA as control; -ve PCR = PCR reaction with H_20 instead of cDNA as control).

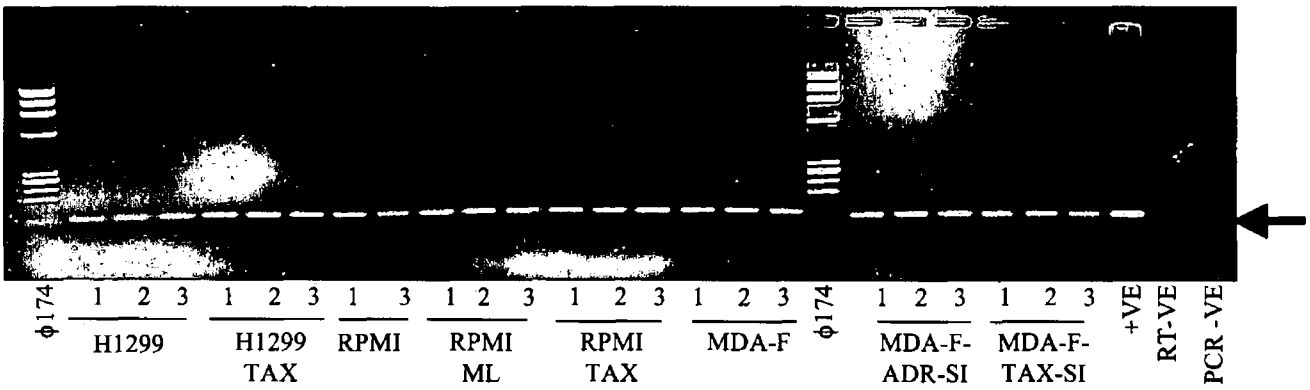


Figure 3.8.7.3.2 RT-PCR of all CM samples for APAF1 interacting protein mRNA. APAF1 interacting protein mRNA gene transcripts (arrow) were detected in all CM samples tested. The probe set corresponding to APAF1 interacting protein mRNA (218698_at), is called Absent in all samples. Amplified product was detected in the positive control and was undetected in the negative control samples (-ve RT = RT reaction with H_20 instead of RNA as control; -ve PCR = PCR reaction with H_20 instead of cDNA as a control).

3.8.8 Gene transcripts significantly up/ down-regulated between parental and drug selected variant CM

The following Section is an overview of the probe sets found to be significantly changed between the parental CM samples and their drug selected variant CM samples. DChip software was used to generate these lists (see Section 2.13.9.2), the mean of the replicate samples in each group were calculated (along with the standard deviation) and compared to other sample groups looking at gene expression of all probe sets available. Presence/ Absence calls assigned to probe sets did not influence the lists generated as results in the previous section (Section 3.8.7) indicated that gene transcripts called Absent can still be detected by RT-PCR. The criteria specified for probe sets to be deemed significantly changed were:

1. A difference in expression >50 fluorescence units

2. A fold change> 1.2
3. A p-value of significance< 0.05

The lists are sorted by fold change, no colour shading indicates gene transcripts that are up-regulated and grey shading indicates gene transcripts that are down-regulated. The dChip classify gene function was also used to determine the molecular function for the significantly changed gene transcripts (see Section 2.13.9.2).

H1299 vs. H1299 TAX

43 probe sets were significantly changed in H1299 TAX CM samples compared to H1299 CM. 34 of these were down-regulated and 9 were up-regulated (see Table 3.8.8.1). Gene ontology annotation was available for 19 of the 43 gene transcripts corresponding to probe sets significantly changed however; no gene category was significantly overrepresented as determined by dChip. 16 of the 19 gene transcripts significantly changed have a known function of binding, including nucleic acid binding (forkhead box A2, eukaryotic translation elongation factor 1 alpha 1, zinc finger protein 573, THAP domain containing 6 and nuclear receptor subfamily 1, group H, member 4), actin binding (myosin IXB) and calcium ion binding (KIAA0494 gene product and calcium channel, voltage-dependent, L type, alpha 1C subunit).

With 12/43 probe sets significantly changed the chromosomal region is known for the gene transcripts they correspond to, six of them to map to the chr1p chromosomal region. Four of these six gene transcripts (DNA-damage inducible protein 2 (224449_at), Leucine zipper and CTNNBIP1 domain containing (226081_at), Similar to bA476I15.3 (novel protein similar to septin- 232953_at) and Hypothetical protein FLJ32784 (1552853_at)) correspond to gene transcripts mapping to the chr1p36 chromosomal region giving an over-representation p-value of 0.000704. All four of the gene transcripts were down-regulated.

Of the 19 gene transcripts significantly changed, pathway annotation was known for 4; eukaryotic translation elongation factor 1 alpha 1 belongs to the GenMAPP Translation Factors pathway, protein kinase, AMP-activated, alpha 2 catalytic subunit belongs to the GenMAPP Fatty_Acid_Synthesis pathway, adenylate kinase 3 belongs to the KEGG Pyrimidine metabolism pathway and calcium channel, voltage-dependent, L type, alpha 1C subunit belongs to the GenMAPP Calcium_Channels pathway.

probe set	gene	Accession	fold change	diff of means	P value
233406_at	gb:AK022100.1 /DB_XREF=gi:10	AK022100	6.02	50.3	0.046382
204347_at	adenylate kinase 3 /// adenylate l	AI653169	4.32	104.06	0.001076
213477_x_at	eukaryotic translation elongation	AL515273	3.48	171.63	0.000362
242323_at	gb:AV714268 /DB_XREF=gi:107	AV714268	3.26	52.19	0.009836
227732_at	ataxin 7-like 1	AB033044	2.23	83.8	0.009627
208397_x_at	potassium inwardly-rectifying cha	U39195	1.9	74.64	0.038166
233658_at	gb:AK022413.1 /DB_XREF=gi:10	AK022413	1.43	144.51	0.025735
207709_at	protein kinase, AMP-activated, al	NM_006252	1.36	51.99	0.008675
238766_at	gb:AI800311 /DB_XREF=gi:53657	AI800311	1.34	163.78	0.044352
213924_at	gb:BF476502 /DB_XREF=gi:1154	BF476502	-25.14	-51.54	0.008211
201777_s_at	KIAA0494 gene product	BC002525	-10.7	-88.22	0.049897
236428_at	gb:D59900 /DB_XREF=gi:961006	D59900	-9.62	-107.06	0.000213
224449_at	DNA-damage inducible protein 2	BC006011	-5.11	-201.62	0.013541
217627_at	zinc finger protein 573	BE515346	-3.75	-90.88	0.015838
242099_at	gb:T83938 /DB_XREF=gi:7.12226	T83938	-3.24	-110.99	0.000656
1552853_at	hypothetical protein FLJ32784	NM_144623	-3.18	-367.12	0.02148
243965_at	Galactose-3-O-sulfotransferase 3	AI860920	-2.87	-64.49	0.032543
240400_at	gb:BF509826 /DB_XREF=gi:1159	BF509826	-2.76	-79.68	0.000883
241904_at	gb:BE068893 /DB_XREF=gi:8413	BE068893	-2.74	-79.37	0.023298
1554128_at	hypothetical protein MGC39724	BC029594	-2.59	-263.25	0.044604
226081_at	Leucine zipper and CTNNBIP1 do	BE219717	-2.57	-92.79	0.047893
231288_at	Similar to 0610010D24Rik protein	BF475370	-2.56	-69.69	0.02801
1567333_at	gb:AJ431618.1 /DB_XREF=gi:18	AJ431618	-2.53	-61.32	0.020938
1554375_a_at	nuclear receptor subfamily 1, gro	AF478446	-2.49	-91.55	0.01629
208452_x_at	myosin IXB	NM_004145	-2.47	-74.53	0.005262
243691_at	gb:AA150455 /DB_XREF=gi:1721	AA150455	-2.4	-69.03	0.043482
206491_s_at	N-ethylmaleimide-sensitive factor	NM_003827	-2.35	-66.73	0.025851
208020_s_at	calcium channel, voltage-depende	NM_000719	-2.22	-136.12	0.046526
211238_at	a disintegrin and metalloproteinas	AF215824	-2.16	-106.79	0.044771
1563754_at	glutamate receptor, ionotropic, ka	AJ252246	-2.1	-138.83	0.012696
232756_at	gb:AK022394.1 /DB_XREF=gi:10	AK022394	-1.81	-62.76	0.022827
242853_at	gb:C18965 /DB_XREF=gi:158056	C18965	-1.78	-54.28	0.03743
232953_at	Similar to bA476I15.3 (novel prote	AL137028	-1.76	-51.89	0.011445
235218_x_at	THAP domain containing 6	BF339201	-1.73	-54.02	0.00364
204158_s_at	T-cell, immune regulator 1, ATPa	NM_006019	-1.69	-367.9	0.02937
241593_x_at	gb:BF476913 /DB_XREF=gi:1154	BF476913	-1.66	-103.62	0.029323
210103_s_at	forkhead box A2	AB028021	-1.65	-289.39	0.014169
242712_x_at	gb:BE856960 /DB_XREF=gi:1037	BE856960	-1.65	-138.6	0.039139
202223_at	integral membrane protein 1	NM_002219	-1.64	-90.4	0.029453
244471_x_at	gb:BF515177 /DB_XREF=gi:1160	BF515177	-1.46	-50.62	0.008994
233412_x_at	Transcribed locus, weakly similar	AW971238	-1.37	-55.07	0.027411
1555792_a_at	hypothetical protein FLJ36046	BC033499	-1.34	-52.99	0.026706
232516_x_at	YY1 associated protein	AU150385	-1.32	-52.41	0.010897

Table 3.8.8.1 Probe sets significantly up or down-regulated between H1299 and H1299 TAX CM samples. The list of gene transcripts is sorted by fold change. All down-regulated gene transcripts are shaded grey with no shading applied to the up-regulated gene transcripts.

RPMI vs. RPMI ML

73 probe sets were significantly changed in RPMI ML CM samples compared to RPMI CM samples, 46 were up-regulated and 26 down-regulated (see Table 3.8.8.2). The gene transcripts corresponding to 40 of the 73 probe sets significantly changed were annotated for gene function. 28 grouped into the organelle gene ontology group giving a significant over representation p value of 0.000202, 9 of these were down-regulated and 19 were up-regulated.

Chromosomal location was known for 23 of the 73 gene transcripts that the significantly changed probe sets correspond to. The gene transcripts mapped to different regions of chromosomes 1-7 and 9 but no chromosome/ chromosomal regions were significantly over represented.

Similarly no pathways were significantly over represented. Of the 73 probe set significantly changed between the RPMI and RPMI ML CM samples only 10 corresponded to gene transcripts belonging to known pathways. The pathways represented included KEGG Fatty acid metabolism (1), gamma-Hexachlorocyclohexane (1) and Tryptophan metabolism (1) pathways and GenMAPP Apoptosis (1), Ribosomal Proteins (2), G_Protein_Signalling (1) and Electron_Tronsport_Chain pathways.

probe set	gene	Accession	fold change	diff of means	P value
220794_at	gremlin 2 homolog, cysteine knot	NM_022469	24.74	53.77	0.006438
222266_at	gb:BF796940 /DB_XREF=gi:1210	BF796940	14.84	66.62	0.005071
204347_at	adenylate kinase 3 /// adenylate k	AI653169	11.02	75.31	0.00913
218464_s_at	hypothetical protein FLJ10700	NM_018182	10.29	54.71	0.018546
219985_at	heparan sulfate (glucosamine) 3-(	NM_006042	10.26	54.97	0.004538
227227_at	Transcribed locus, moderately sir	AI344332	7.18	139.74	0.003352
1566001_at	Ankyrin repeat domain 11	AK096064	6.83	57.27	0.026834
212195_at	Interleukin 6 signal transducer (gp	AL049265	5.67	59.29	0.040941
235983_at	gb:AI360167 /DB_XREF=gi:41117	AI360167	5.28	50.19	0.033165
204914_s_at	SRY (sex determining region Y)-b	AW157202	4.6	59.54	0.049238
241904_at	gb:BE068893 /DB_XREF=gi:8413	BE068893	4.41	58.97	0.045711
229871_at	hypothetical protein FLJ10211	W74622	3.88	153.62	0.00284
1563209_a_at	Homo sapiens, clone IMAGE:451	BC035876	3.64	72.06	0.024161
236811_at	DMRT-like family C2	AI651482	3.4	77.3	0.008786
236484_at	gb:AW850611 /DB_XREF=gi:7941	AW850611	3.23	53.12	0.045036
1569181_x_at	gb:BC017896.1 /DB_XREF=gi:22	BC017896	3.12	61.21	0.037658
213931_at	inhibitor of DNA binding 2, domin	AI819238	3.03	82.59	0.012924
208905_at	cytochrome c, somatic	BC005299	3	52.9	0.043279
241600_at	gb:AI270986 /DB_XREF=gi:38901	AI270986	2.98	74.74	0.040423
215304_at	Clone 23948 mRNA sequence	U79293	2.91	60.62	0.009442
222565_s_at	protein kinase D3	BF978541	2.75	51.29	0.038593
232150_at	gb:AA134418 /DB_XREF=gi:1691	AA134418	2.73	140.68	0.028463
1569180_at	gb:BC017896.1 /DB_XREF=gi:22	BC017896	2.66	53.57	0.019075
214323_s_at	UPF3 regulator of nonsense trans	N36842	2.36	132.87	0.0098
227952_at	Zinc finger protein 595	AI580142	2.26	302.8	0.044858
208668_x_at	high-mobility group nucleosomal I	BC003689	2.25	71.87	0.047983
214938_x_at	high-mobility group box 1	AF283771	2.21	185.24	0.025819
229447_x_at	hypothetical protein LOC200030	N32025	2.15	86.88	0.043675
210718_s_at	ADP-ribosylation factor-like	AF119889	2.14	95.3	0.024139
200630_x_at	SET translocation (myeloid leuke	AV702810	2.03	64.19	0.032103
201103_x_at	hypothetical protein MGC8902 ///	BE299495	1.99	60.94	0.021264
214320_x_at	cytochrome P450, family 2, subfa	T67741	1.91	123.47	0.028632
200817_x_at	ribosomal protein S10	NM_001014	1.78	80.63	0.049138
212854_x_at	AG1	AB051480	1.74	57.58	0.043794
213356_x_at	heterogeneous nuclear ribonuclec	AL568186	1.68	63.12	0.048771
242896_at	gb:BF223302 /DB_XREF=gi:1113	BF223302	1.68	52.39	0.027131
207631_at	neighbor of BRCA1 gene 2	NM_005821	1.65	99.13	0.026481
229567_at	similar to CG10671-like	AI742370	1.52	62.99	0.043195
201220_x_at	C-terminal binding protein 2	NM_001329	1.45	61.89	0.045658
215852_x_at	KIAA0889 protein	AK022023	1.44	479.93	0.013725
239127_at	CDNA FLJ42086 fis, clone TESO	T61954	1.44	66.22	0.040592
210334_x_at	baculoviral IAP repeat-containing	AB028869	1.34	76.1	0.010615
1561573_at	Homo sapiens, clone IMAGE:552	BC039476	1.33	57.99	0.019732
52005_at	Widely-interspaced zinc finger m	AA422049	1.23	52.47	0.02343
223815_at	Hypothetical protein BC009518	AF130069	1.23	102.91	0.01258
243895_x_at	gb:AW452435 /DB_XREF=gi:699	AW452435	1.23	779.45	0.047436
233816_at	Solute carrier family 8 (sodium/ca	AW668617	1.21	52.85	0.034609
215724_at	phospholipase D1, phophatidylch	AJ276230	-6.63	-68.74	0.023609
219058_x_at	lipocalin 7	NM_022164	-4.87	-197.85	0.01626
220808_at	Theg homolog (mouse)	NM_016585	-4.75	-119.28	0.004057
1554128_at	hypothetical protein MGC39724	BC029594	-4.15	-381.87	0.022574
218922_s_at	LAG1 longevity assurance homol	NM_024552	-3.72	-85.32	0.009085
226538_at	mannosidase, alpha, class 2A, m	AW700323	-3.66	-80.42	0.015963
218698_at	likely ortholog of mouse monocyt	NM_015957	-3.47	-141.78	0.043591
240304_s_at	Transmembrane channel-like 5	BG484769	-3.32	-138.84	0.014492
208020_s_at	calcium channel, voltage-depende	NM_000719	-3.09	-176.81	0.017252
206491_s_at	N-ethylmaleimide-sensitive factor	NM_003827	-2.58	-66.53	0.014097
1554375_a_at	nuclear receptor subfamily 1, grou	AF478446	-2.17	-69.6	0.040681
208452_x_at	myosin IXB	NM_004145	-2.16	-66.32	0.017701
231590_x_at	gb:AI379823 /DB_XREF=gi:41898	AI379823	-1.87	-60.38	0.043531
225276_at	G1 to S phase transition 1	AA143579	-1.85	-50.93	0.048889
220725_x_at	Dynein, axonemal, heavy polypep	NM_025095	-1.73	-258.1	0.015389
239805_at	gb:AW136060 /DB_XREF=gi:614	AW136060	-1.68	-52.37	0.018374
231286_at	chromosome 9 open reading fram	AW452080	-1.63	-57.66	0.009371
236389_x_at	gb:T90064 /DB_XREF=gi:718577	T90064	-1.58	-92.93	0.011942
208817_at	catechol-O-methyltransferase	BC000419	-1.49	-338.75	0.010668
216153_x_at	reversion-inducing-cysteine-rich p	AK022897	-1.48	-215.22	0.028096
238701_x_at	FLJ45803 protein	BE176566	-1.4	-156.34	0.049751
217696_at	fucosyltransferase 7 (alpha (1,3) f	AA767713	-1.32	-285.59	0.028329
216362_at	gb:AJ251844.1 /DB_XREF=gi:938	AJ251844	-1.28	-67.92	0.01762
210201_x_at	bridging integrator 1	AF001383	-1.27	-173.45	0.014539
214614_at	homeo box HB9	AI738662	-1.26	-176.95	0.033232
219113_x_at	dehydrogenase/reductase (SDR f	NM_016246	-1.23	-479.37	0.012137

Table 3.8.8.2 Probe sets significantly up or down-regulated between RPMI and RPMI ML CM samples. The list of gene transcripts is sorted by fold change. All down-regulated gene transcripts are shaded grey with no shading applied to the up-regulated gene transcripts.

RPMI vs. RPMI TAX

The RPMI CM samples compared to the RPMI TAX CM samples showed 150 probe sets that were significantly different (Tables 3.8.8.3(a-b)). More probe sets were up-regulated (92) than down-regulated (58) as was the case with RPMI CM samples compared to RPMI ML CM samples (see Table 3.8.8.2). Gene ontology annotation was known for just under half (70) of the gene transcripts corresponding to the probe sets on the list but no gene ontology group was determined to be significantly overrepresented by dChip. 11 of the 70 annotated gene transcripts have a known molecular function in regulation of metabolism and another 12 have a function in protein metabolism. 11 of the 70 have a known molecular function in the regulation of DNA-dependent transcription and 38 of the 70 are annotated to have a known function of binding, 16 of which are nucleic acid/ DNA binding and a different 15 nucleotide binding. One of these gene transcripts with a function of adenyl nucleotide binding and transferase activity, adenylate kinase 3, was also up-regulated in the taxol- selected CM samples by 4.3 fold compared to the parental CM sample in the H1299 CM vs. H1299 TAX CM sample group. Other molecular function categories represented by the significantly changed gene transcripts in the RPMI CM samples against the RPMI TAX CM samples include catalytic activity (29), cell communication (14) and receptor/ signal transducer activity (10).

16 of the 70 gene transcripts significantly changed had known pathway annotations for 16 different pathways, none of which were deemed to be over represented. The pathways represented included GenMAPP TGF_Beta_Signaling_Pathway (1), Inflammatory_Response_Pathway (1) and Ribosomal_Proteins (2) and KEGG Phosphatidylinositol signalling system (3) and Pyrimidine metabolism (2).

Chromosomal annotations were known for 49/70 gene transcripts corresponding to the significantly changed probe sets. 13/ 49 of these gene transcripts mapped to chromosome 7

giving an over represented p-value of 0.000399. 5 of the 13 further mapped to the chr7q21 region of the chromosome giving a significantly over represented p-value of 0.000087. Another 4 of the 49 chromosome annotated probe sets mapped to the chr5p15 region and the region was also deemed over represented with a p-value of 0.000370)

probe set	gene	Accession	fold change	difference of means	P value
214854_at	Transcribed locus, moderately simil	AC004490	53.31	52.31	0.037248
216480_x_at	gb:AF060927.1 /DB_XREF=gi:3139(	AF060927	22.79	52.65	0.002647
236520_at	gb:AW972380 /DB_XREF=gi:81622:	AW972380	13.2	50.74	0.012573
204347_at	adenylate kinase 3 /// adenylate kin	AI653169	10.65	72.6	0.010142
244350_at	gb:N33403 /DB_XREF=gi:1153802 /	N33403	10.58	70.4	0.008354
215657_at	gb:AK025044.1 /DB_XREF=gi:1043	AK025044	8.09	55.25	0.022038
232330_at	Hypothetical protein FLJ10803	AK024861	7.74	58.84	0.032577
215321_at	Rap2-binding protein 9	AI826798	7.57	99.60	0.0188
236472_at	gb:AI806586 /DB_XREF=gi:5393152	AI806586	7.57	72.84	0.010493
224851_at	cyclin-dependent kinase 6	AW274756	7.42	58.46	0.01727
227227_at	Transcribed locus, moderately simil	AI344332	7.03	136.32	0.014621
243902_at	gb:AI434868 /DB_XREF=gi:429919(	AI434868	6.97	67.63	0.011572
241617_x_at	gb:BE961949 /DB_XREF=gi:11764:	BE961949	6.57	61.11	0.007142
1570505_at	ATP-binding cassette, sub-family B	BC020618	6.61	118.83	0.011712
236404_at	gb:AW197320 /DB_XREF=gi:64765:	AW197320	5.58	64.43	0.014634
225431_x_at	Aminoacylase 1-like 2	BE779764	5.32	121.88	0.011081
202850_at	ATP-binding cassette, sub-family D	NM_002856	5.3	64.55	0.004703
242108_at	Two pore segment channel 1	AA001615	5.19	81.11	0.026949
230466_s_at	gb:AI092770 /DB_XREF=gi:3431748	AI092770	5.07	66.63	0.013971
204712_at	WNT inhibitory factor 1	NM_007191	4.82	76.27	0.037368
215386_at	Autism susceptibility candidate 2	AU148161	4.76	62.85	0.023977
235361_at	Hepatitis B virus x associated prote	BF033855	4.73	51.6	0.032679
241600_at	gb:AI270986 /DB_XREF=gi:3890153	AI270986	4.63	136.58	0.008626
219108_x_at	DEAD (Asp-Glu-Ala-Asp) box polyp	NM_017695	4.51	68.77	0.029036
237330_at	gb:AA603494 /DB_XREF=gi:243736	AA603494	4.43	50.6	0.030548
244866_at	gb:H72108 /DB_XREF=gi:1043924 /	H72108	4.33	53.9	0.043566
236484_at	gb:AW850611 /DB_XREF=gi:79461:	AW850611	4.15	75.24	0.044761
205814_at	glutamate receptor, metabotropic 3	NM_000840	3.88	83.55	0.017967
218412_s_at	GTF2I repeat domain containing 1	NM_016328	3.86	52.09	0.048293
213578_at	bone morphogenetic protein recepto	AI678679	3.81	50.71	0.033799
238549_at	core-binding factor, runt domain, alp	AI420611	3.77	115.41	0.028758
226462_at	iroquois homeobox protein 2	AI928035	3.74	204.56	0.007047
1557599_a_at	CDNA clone IMAGE:5274693, partia	BC041378	3.56	71.65	0.013607
237157_at	gb:AW117547 /DB_XREF=gi:60861:	AW117547	3.53	71.5	0.031215
234258_at	MRNA; cDNA DKFZp434F0472 (fror	AL137331	3.52	54.26	0.03713
230713_at	gb:BF115786 /DB_XREF=gi:109852	BF115786	3.47	50.75	0.02118
239082_at	Homo sapiens, clone IMAGE:53113	BF437161	3.41	59.27	0.028984
226332_at	hypothetical protein MGC40405	AI569932	3.4	58.42	0.039815
203298_s_at	Jumonji, AT rich interactive domain	NM_004973	3.18	57.2	0.04161
225845_at	HSPC063 protein	BG253884	3.16	90.72	0.019224
232113_at	Hypothetical gene supported by BX	N90870	3.14	92.51	0.010047
202391_at	brain abundant, membrane attachec	NM_006317	2.94	75.77	0.032298
200833_s_at	RAP1B, member of RAS oncogene	NM_015646	2.78	52.3	0.034405
221740_x_at	Transcribed locus, moderately simil	AI140364	2.76	52.05	0.019407
203917_at	coxsackie virus and adenovirus rece	NM_001338	2.74	57.22	0.039028
242759_at	gb:AI821726 /DB_XREF=gi:5440805	AI821726	2.69	83.31	0.005837
236802_at	gb:AI732132 /DB_XREF=gi:5053267	AI732132	2.65	93.41	0.020767
218858_at	DEP domain containing 6	NM_022783	2.58	51.64	0.035523
232611_at	RAN binding protein 2-like 1	AK022838	2.47	77.12	0.022823
244612_at	gb:AW117181 /DB_XREF=gi:60857(	AW117181	2.43	120.94	0.032037
1557996_at	gb:AK091784.1 /DB_XREF=gi:2175(	AK091784	2.39	129.39	0.031114
216364_s_at	fragile X mental retardation 2	AJ001650	2.34	58.58	0.040893
243271_at	gb:AI064690 /DB_XREF=gi:6358962	AI064690	2.26	57.59	0.015778
228854_at	gb:AI492388 /DB_XREF=gi:4393391	AI492388	2.25	70.96	0.017299
230497_at	bruno-like 5, RNA binding protein (D	BE503640	2.24	56.07	0.026641
232256_s_at	Hypothetical gene supported by AK	AF143329	2.23	96.04	0.031816
212854_x_at	AG1	AB051480	2.09	85.14	0.040579
200817_x_at	ribosomal protein S10	NM_001014	2.08	111.68	0.040589
213881_x_at	SMT3 suppressor of mif two 3 homc	AI971724	2.08	125.83	0.046759
233037_at	CDNA clone IMAGE:4801089, partia	AF138869	2.05	57.7	0.009922
242920_at	gb:AW590838 /DB_XREF=gi:72779(	AW590838	2.05	71.43	0.039764
209134_s_at	ribosomal protein S6	BC000524	1.97	109.68	0.011572
214320_x_at	cytochrome P450, family 2, subfam	T67741	1.95	120.77	0.008449
222380_s_at	Similar to Microneme antigen	AI907083	1.9	157.62	0.008893
200080_s_at	H3 histone, family 3A /// H3 histone	AI955655	1.86	57.87	0.034131
81811_at	Similar to Microneme antigen	AI744451	1.85	96.21	0.019962
1561471_at	Homo sapiens, clone IMAGE:48311	BC039114	1.85	88.58	0.03691
207631_at	neighbor of BRCA1 gene 2	NM_005821	1.76	115.18	0.037144

Table 3.8.8.3 Probe sets significantly up-regulated in the RPMI TAX compared to the RPMI CM samples. The list of gene transcripts is sorted by fold change. All down-regulated gene transcripts are shaded grey with no shading applied to the up-regulated gene transcripts.

probe set	gene	Accession	fold change	difference of means	P value
200801_x_at	actin, beta	NM_001101	1.73	59.12	0.027541
229567_at	similar to CG10671-like	AI742370	1.72	86.94	0.033326
231831_at	hypothetical protein LOC90639	AU153267	1.67	79.96	0.027296
210718_s_at	ADP-ribosylation factor-like	AF119889	1.65	53.91	0.035008
1561573_at	Homo sapiens, clone IMAGE:55287	BC039476	1.64	110.74	0.014761
1560562_a_at	Hypothetical protein MGC48625	AK026366	1.56	83.28	0.026495
212525_s_at	H2A histone family, member X	AA760862	1.53	115.33	0.034113
236267_at	zinc finger protein 346	BG178775	1.52	68.65	0.030424
1570257_x_at	gb:BC034527.1 /DB_XREF=gi:2290	BC034527	1.51	62.98	0.028286
207134_x_at	tryptase beta 2	NM_024164	1.47	95.99	0.040107
231686_at	Hypothetical LOC388572	AL137655	1.47	51.75	0.040335
210722_at	pancreatic lipase-related protein 1	BC005233	1.46	88.58	0.045881
215852_x_at	KIAA0889 protein	AK022023	1.46	510.43	0.01824
221270_s_at	queuine tRNA-ribosyltransferase 1 (	NM_031209	1.44	81.11	0.036975
239127_at	CDNA FLJ42086 fis, clone TESOP1	T61954	1.44	66.24	0.038714
239311_at	gb:AI367034 /DB_XREF=gi:413677	AI367034	1.42	62.38	0.016186
243724_at	gb:AW979182 /DB_XREF=gi:81704	AW979182	1.42	53.17	0.034216
231715_s_at	pyrroline-5-carboxylate reductase fa	NM_013328	1.34	58.54	0.047846
240397_x_at	gb:AI801626 /DB_XREF=gi:536709	AI801626	1.28	179.26	0.046197
210965_x_at	cell division cycle 2-like 5 (cholinest	BC001274	1.27	412.48	0.025637
203045_at	ninjurin 1	NM_004148	1.25	142.63	0.030042
214488_at	RAP2B, member of RAS oncogene	NM_002886	1.25	77.47	0.021701
243895_x_at	gb:AW452435 /DB_XREF=gi:69932	AW452435	1.24	819.39	0.006006
228700_at	Chromosome X open reading frame	AA063608	1.22	57.13	0.047997
201671_x_at	ubiquitin specific protease 14 (tRNA	BC003556	1.2	365.64	0.000614
238572_at	hypothetical protein MGC16211	AI082836	-35.42	-63.84	0.017643
215724_at	phospholipase D1, phophatidylcholi	AJ276230	-14.13	-75.22	0.032333
206491_s_at	N-ethylmaleimide-sensitive factor at	NM_003827	-5.89	-90.17	0.016714
1554128_at	hypothetical protein MGC39724	BC029594	-5.33	-408.63	0.000983
240304_s_at	Transmembrane channel-like 5	BG484769	-4.79	-157.25	0.03476
222365_at	gb:AW974666 /DB_XREF=gi:81658	AW974666	-4.69	-58.84	0.03185
218698_at	likely ortholog of mouse monocyte r	NM_015957	-3.89	-147.93	0.007462
1554375_a_at	nuclear receptor subfamily 1, group	AF478446	-3.8	-95.13	0.036448
208020_s_at	calcium channel, voltage-dependent	NM_000719	-3.59	-188.58	0.006098
218922_s_at	LAG1 longevity assurance homolog	NM_024552	-3.46	-82.91	0.005124
225276_at	G1 to S phase transition 1	AA143579	-3.38	-77.99	0.027384
201422_at	interferon, gamma-inducible protein	NM_006332	-3.07	-67.49	0.046016
226538_at	mannosidase, alpha, class 2A, mer	AV700323	-2.97	-73.43	0.033573
208452_x_at	myosin IXB	NM_004145	-2.91	-68.73	0.01053
231598_x_at	gb:AI379823 /DB_XREF=gi:418967	AI379823	-2.75	-82.4	0.014557
1557330_at	CDNA clone IMAGE:4836946, parti	BC027619	-2.48	-72.14	0.007888
220808_at	Theg homolog (mouse)	NM_016585	-2.35	-86.89	0.009031
1564155_x_at	gb:BC041466.1 /DB_XREF=gi:2737	BC041466	-2.31	-72.71	0.016593
236186_x_at	interleukin 17 receptor E	AW003256	-1.98	-71.31	0.030851
206936_x_at	gb:NM_022335.1 /DB_XREF=gi:116	NM_022335	-1.9	-119.28	0.031369
241675_s_at	gb:AI349737 /DB_XREF=gi:408694	AI349737	-1.9	-60.15	0.023155
244490_at	gb:AA748520 /DB_XREF=gi:278847	AA748520	-1.88	-50.27	0.021363
1553326_at	leucine-rich repeat-containing G pro	AF453828	-1.88	-85.37	0.032413
236389_x_at	gb:T90064 /DB_XREF=gi:718577 /D	T90064	-1.86	-116.29	0.025084
240882_at	MRNA; cDNA DKFZp313N0236 (fro	R85522	-1.71	-256.08	0.042304
239805_at	gb:AW136060 /DB_XREF=gi:61401	AW136060	-1.64	-50.2	0.011323
217684_at	thymidylate synthetase	BG281679	-1.62	-87.56	0.047231
219358_s_at	centaurin, alpha 2	NM_018404	-1.62	-166.05	0.012815
216153_x_at	reversion-inducing-cysteine-rich prot	AK022897	-1.6	-246.45	0.021247
216362_at	gb:AJ251844.1 /DB_XREF=gi:9368	AJ251844	-1.51	-103.03	0.022366
216560_x_at	Ig lambda-chain V-J-C region (HCV-	D87021	-1.45	-56.24	0.01853
201109_s_at	thrombospondin 1	AV726673	-1.39	-503.84	0.032746
203565_s_at	menage a trois 1 (CAK assembly fa	NM_002431	-1.39	-98.06	0.009458
214614_at	homeo box HB9	AI738662	-1.39	-243.22	0.036664
1561305_at	Homo sapiens, clone IMAGE:48375	BC040325	-1.39	-178.01	0.046302
201378_s_at	ubiquitin associated protein 2-like	NM_014847	-1.34	-500.38	0.011166
207742_s_at	nuclear receptor subfamily 6, group	NM_001489	-1.34	-788.3	0.007115
209881_s_at	linker for activation of T cells	AF036905	-1.31	-186.9	0.029199
1561614_at	Solute carrier family 8 (sodium/calci	AW452398	-1.3	-55.28	0.019365
207138_at	PHD finger protein 2	NM_005392	-1.29	-925.98	0.0104
233605_x_at	heterogeneous nuclear ribonucleopr	AK022050	-1.29	-513.96	0.001636
231597_x_at	gb:AI371550 /DB_XREF=gi:415030	AI371550	-1.28	-628.64	0.04252
218836_at	ribonuclease P 21kDa subunit	NM_024839	-1.26	-255.48	0.025326
219113_x_at	dehydrogenase/reductase (SDR fan	NM_016246	-1.22	-459.86	0.033826
208148_at	myosin, heavy polypeptide 4, skelet	NM_017533	-1.21	-212.35	0.04595

Table 3.8.3.3 (contd) Probe sets significantly up or down-regulated between RPMI and RPMI TAX CM samples. The list of gene transcripts is sorted by fold change. All down- regulated gene transcripts are shaded grey with no shading applied to the up-regulated gene transcripts.

MDA-F vs. MDA-F-ADR-SI

Tables 3.8.8.4 (a) and (b) list the probe sets called significantly changed between MDA-F and MDA-F ADR CM samples. When this list was generated a function in dChip was selected that allowed for the expression value of probe sets in individual CM samples deemed to be "array outliers" to be excluded. This was used because the large difference in %Present call in the MDA-F-ADR 1 CM sample meant that a gene that was called significantly changed when MDA-F-ADR 2 and 3 CM samples were compared to the three MDA-F CM samples, would possibly not have been determined significantly changed when the expression value for the MDA-F-ADR 1 sample was also included. Of the 94 probe sets that were significantly changed between the MDA-F CM and the MDA-F-ADR CM samples, 72 of these would not have been deemed significantly changed if the "treat array outlier as missing expression" option (see Section) had not been checked.

A large majority of the 94 gene transcripts significantly changed were up-regulated (85). 71 of these probe sets were annotated for molecular function and some of the significantly overrepresented gene ontology groups were protein complex (33), macromolecule biosynthesis (29), structural constituent of ribosome (25), RNA binding (19) each with a significant over representation p-value of 0.00000.

The GenMAPP Ribosomal_Proteins pathway was significantly over represented with 19 of the 29 non- redundant probe sets that correlated to gene transcripts with known pathway annotations, belonging to it giving a significantly over represented p-value of 0.00000.

probe set	gene	Accession	fold change	diff of means	P value
217733_s_at	thymosin, beta 10	NM_021103	15.05	123.23	0.020834
1553570_x_at	cytochrome c oxidase II	NM_173705	12.86	223.4	0.002722
211969_at	heat shock 90kDa protein 1, alph	BG420237	11.07	209.55	0.000107
1553588_at	gb:NM_173710.1 /DB_XREF=gi:2	NM_173710	10.2	329.16	0.000069
1553567_s_at	ATP synthase 6	NM_173702	8.75	86.17	0.009519
211072_x_at	tubulin, alpha, ubiquitous /// tubul	BC006481	8.55	163.06	0.018613
224373_s_at	NADH dehydrogenase 4 /// NADH	AF253979	8.51	76.98	0.040145
224372_at	NADH dehydrogenase 4 /// NADH	AF253979	8.29	249.73	0.005462
212082_s_at	myosin, light polypeptide 6, alkali	BE734356	8.2	50.6	0.006232
200610_s_at	nucleolin	NM_005381	7.19	56.06	0.008283
208628_s_at	nuclease sensitive element bindir	BC002411	6.87	285.46	0.000215
214687_x_at	aldolase A, fructose-bisphosphate	AK026577	6.07	110.41	0.003151
200006_at	Parkinson disease (autosomal re	NM_007262	5.98	57.39	0.03086
214114_x_at	FAST kinase	AK023141	5.92	50.15	0.043193
213414_s_at	ribosomal protein S19	BE259729	5.89	186.98	0.008803
200966_x_at	aldolase A, fructose-bisphosphate	NM_000034	5.68	101.6	0.008403
1553538_s_at	gb:NM_173704.1 /DB_XREF=gi:2	NM_173704	5.63	112.43	0.003017
204892_x_at	eukaryotic translation elongation	NM_001402	5.28	254.84	0.001117
201615_x_at	caldesmon 1	AI685060	5.26	56.57	0.00463
229353_s_at	nuclear ubiquitous casein kinase	AW515443	5.21	173.78	0.000024
200651_at	guanine nucleotide binding protei	NM_006098	4.95	52.88	0.046231
211073_x_at	ribosomal protein L3 /// ribosomal	BC006483	4.81	86.58	0.004552
201550_x_at	actin, gamma 1	NM_001614	4.74	62.43	0.005118
211970_x_at	actin, gamma 1	BG026805	4.66	77.56	0.046674
215963_x_at	gb:Z98200 /DB_XREF=gi:600229	Z98200	4.64	82.36	0.003245
211620_x_at	runt-related transcription factor 1	L21756	4.55	142.11	0.048554
211983_x_at	actin, gamma 1	BE741683	4.54	80.96	0.019703
219762_s_at	ribosomal protein L36	NM_015414	4.5	50.99	0.042117
201217_x_at	ribosomal protein L3	NM_000967	4.45	79.46	0.00763
211968_s_at	heat shock 90kDa protein 1, alph	AI962933	4.29	99.94	0.028257
203034_s_at	ribosomal protein L27a	NM_000990	4.28	66.62	0.006574
212988_x_at	actin, gamma 1	AL515810	4.28	51.14	0.046588
200031_s_at	ribosomal protein S11 /// ribosom	NM_001015	4.24	94.47	0.041127
206559_x_at	eukaryotic translation elongation	NM_001403	4.24	229.36	0.004421
200909_s_at	ribosomal protein, large P2	NM_001004	4.19	52.49	0.027566
200858_s_at	ribosomal protein S8	NM_001012	4.13	84.31	0.003764
212039_x_at	ribosomal protein L3	BG339228	4.1	77.41	0.047231
213453_x_at	glyceraldehyde-3-phosphate dehy	BF689355	3.91	66.6	0.03134
213614_x_at	eukaryotic translation elongation	BE786672	3.9	404.14	0.004142
201530_x_at	eukaryotic translation initiation fa	NM_001416	3.87	87.27	0.004562
201429_s_at	ribosomal protein L37a	NM_000998	3.8	319.04	0.004944
226131_s_at	ribosomal protein S16	AA583817	3.77	239.74	0.025219
200926_at	ribosomal protein S23	NM_001025	3.63	71.3	0.041146
223734_at	ovary-specific acidic protein	AF329088	3.61	71.81	0.031442
201105_at	lectin, galactoside-binding, solubl	NM_002305	3.6	56.62	0.027375
201305_x_at	acidic (leucine-rich) nuclear phos	AV712577	3.57	1017.32	0.027835
200834_s_at	ribosomal protein S21	NM_001024	3.4	96.13	0.00346
213583_x_at	gb:BE964125 /DB_XREF=gi:117E	BE964125	3.31	528.73	0.003062

Table 3.8.8.4 (a) Probe sets significantly up/ down-regulated between MDA-F and MDA-F-ADR CM samples .The list of gene transcripts is sorted by fold change. All down- regulated gene transcripts are shaded grey with no shading applied to the up-regulated gene transcripts.

probe set	gene	Accession	fold change	diff of means	P value
224616_at	dynein, cytoplasmic, light interme	BG110975	3.26	62.74	0.039549
200763_s_at	ribosomal protein, large, P1	NM_001003	3.25	89.93	0.007455
212284_x_at	tumor protein, translationally-cont	BG498776	3.09	127.31	0.018554
200089_s_at	ribosomal protein L4 /// ribosomal	AI953886	3.06	99.72	0.002703
216342_x_at	gb:AL121916 /DB_XREF=gi:7406	AL121916	3.02	52.96	0.019242
1559489_a_at	hypothetical protein LOC257358	AK093725	2.97	118.9	0.044835
217398_x_at	glyceraldehyde-3-phosphate dehy	AK026525	2.86	55.15	0.020471
211787_s_at	eukaryotic translation initiation fa	BC006210	2.82	79.76	0.010322
221775_x_at	ribosomal protein L22	BG152979	2.82	112.19	0.022805
201306_s_at	acidic (leucine-rich) nuclear phos	NM_006401	2.76	131.08	0.027315
200594_x_at	heterogeneous nuclear ribonuclec	NM_004501	2.75	60.44	0.030674
1558173_a_at	leucine zipper protein 1	AK093016	2.71	56.99	0.012152
209134_s_at	ribosomal protein S6	BC000524	2.62	350.63	0.003657
208768_x_at	ribosomal protein L22	D17652	2.59	102.86	0.025973
213757_at	gb:AA393940 /DB_XREF=gi:204E	AA393940	2.52	246.29	0.016118
210646_x_at	ribosomal protein L13a	BC001675	2.47	97.63	0.013184
212790_x_at	ribosomal protein L13a	BF942308	2.47	96.78	0.013129
208695_s_at	ribosomal protein L39	BC001019	2.45	294.85	0.000971
200716_x_at	ribosomal protein L13a	NM_012423	2.4	96.06	0.012137
200817_x_at	ribosomal protein S10	NM_001014	2.29	295.82	0.016535
220209_at	peptide YY, 2 (seminalplasmin)	NM_021093	2.19	59.8	0.030716
200715_x_at	ribosomal protein L13a	BC000514	2.11	68.81	0.017213
211542_x_at	ribosomal protein S10	BC004334	2.1	252.01	0.024807
200095_x_at	ribosomal protein S10 /// ribosom	AA320764	2.09	307.34	0.023518
201254_x_at	ribosomal protein S6	NM_001010	2.01	78.98	0.015147
213941_x_at	ribosomal protein S7	AI970731	2.01	213.11	0.015483
221488_s_at	chromosome 6 open reading fram	AF230924	1.74	72.47	0.017006
1552319_a_at	kallikrein 8 (neuropsin/ovasin)	NM_144506	1.71	160.25	0.045895
213867_x_at	actin, beta	AA809056	1.63	119.95	0.009981
200801_x_at	actin, beta	NM_001101	1.61	97.16	0.03492
62987_r_at	calcium channel, voltage-depende	AI675178	1.59	87.07	0.012971
214938_x_at	high-mobility group box 1	AF283771	1.56	164.84	0.031716
205827_at	cholecystokinin	NM_000729	1.55	50.57	0.02541
221544_s_at	thyroid hormone receptor associa	BG339606	1.52	158.66	0.001977
216153_x_at	reversion-inducing-cysteine-rich p	AK022897	1.45	157.94	0.025483
59999_at	hypoxia-inducible factor 1, alpha :	W37897	1.41	67.16	0.012406
208817_at	catechol-O-methyltransferase	BC000419	1.37	190.71	0.000717
229961_x_at	FLJ44968 protein	AI871270	1.3	65.29	0.033242
210601_at	cadherin 6, type 2, K-cadherin (fe	BC000019	-2.48	-85.23	0.033697
1563052_at	Homo sapiens, clone IMAGE:529	BC040832	-2.46	-55.4	0.048218
215197_at	Mannosyl (alpha-1,6-)-glycoprotei	AK023838	-2.07	-59.66	0.021508
219610_at	Rho-guanine nucleotide exchange	NM_022448	-2.05	-80.22	0.003526
230077_at	succinate dehydrogenase comple	W90764	-1.59	-54.69	0.022441
1554663_a_at	nuclear mitotic apparatus protein	BC043499	-1.4	-118.73	0.048052
233658_at	gb:AK022413.1 /DB_XREF=gi:10	AK022413	-1.36	-164.4	0.037888
1553258_at	hypothetical protein FLJ30679	NM_153017	-1.32	-57.2	0.035894

Table 3.8.8.4 (b) Probe sets significantly up or down-regulated between MDA-F and MDA-F-ADR CM samples. The list of gene transcripts is sorted by fold change. All down- regulated gene transcripts are shaded grey with no shading applied to the up-regulated gene transcripts.

MDA-F vs. MDA-F-TAX-SI

39 probe sets were significantly different between MDA-F CM samples and MDA-F-TAX-SI CM samples (see Table 3.8.8.5). The majority of these (31) were down-regulated, with 8 up-regulated. Only 15 of the 39 probe sets are annotated. One gene ontology group "oxidoreductase activity\, acting on the CH-NH group of donors" was determined to be overrepresented (p-val 0.000299) as 2 of the 15 annotated gene transcripts alcohol dehydrogenase 4 (class II), pi polypeptide (223781_x_at), and Quinoid dihydropteridine reductase (238616_at) belong to the group which itself only contains 45 gene transcripts.

Of the 14/ 39 gene transcripts with a known chromosomal location, no region was determined to be over represented but the chromosomes represented include 1, 2, 4- 7 and the X chromosome..

probe set	gene	Accession	fold change	diff of means	P value
228452_at	chromosome 17 open reading frar	AA827865	3.38	56.63	0.019761
220406_at	gb:NM_003238.1 /DB_XREF=gi:4	NM_003238	2.64	53.13	0.037813
212854_x_at	AG1	AB051480	2.25	186.95	0.002983
222380_s_at	Similar to Microneme antigen	AI907083	1.96	287.96	0.013562
1552771_a_at	TUDOR gene similar	NM_173083	1.53	56.15	0.008524
213811_x_at	transcription factor 3 (E2A immur	AW062341	1.28	60.12	0.046806
230162_at	gb:AI951640 /DB_XREF=gi:5743	AI951640	1.28	61.08	0.031092
1552504_a_at	KIAA1811 protein	NM_032430	1.27	69.7	0.039655
1560325_at	gb:AL832767.1 /DB_XREF=gi:21	AL832767	-208.95	-207.95	0.023244
217878_s_at	cell division cycle 27	AI203880	-2.92	-78.22	0.022409
211921_x_at	prothymosin, alpha (gene sequen	AF348514	-2.54	-302.12	0.045763
242136_x_at	Hypothetical LOC403340	T66145	-2.48	-55.05	0.039278
235693_at	gb:D81004 /DB_XREF=gi:117888	D81004	-2.45	-76.15	0.020459
231598_x_at	gb:AI379823 /DB_XREF=gi:41898	AI379823	-2.39	-68.1	0.028666
243489_at	gb:BF514098 /DB_XREF=gi:1159	BF514098	-2.2	-882.52	0.012861
219754_at	hypothetical protein FLJ11016	NM_018301	-2.1	-125.78	0.041206
217793_at	RAB11B, member RAS oncogene	AL575337	-2.08	-58.23	0.847095
202015_x_at	gb:NM_006838.1 /DB_XREF=gi:5	NM_006838	-1.99	-476.45	0.034393
1552860_at	transcription elongation factor B p	NM_145653	-1.97	-60.54	0.030515
201293_x_at	peptidylprolyl isomerase A (cyclo	NM_021130	-1.95	-51.79	0.045805
217137_x_at	gb:K00627.1 /DB_XREF=gi:33765	K00627	-1.72	-150.03	0.020739
1561813_at	MRNA full length insert cDNA clo	AL109711	-1.69	-111.24	0.025285
215545_at	Excision repair cross-complemen	AK024185	-1.66	-52.92	0.023454
219610_at	Rho-guanine nucleotide exchange	NM_022448	-1.64	-60.83	0.013611
233912_x_at	ELMO domain containing 2	AK021525	-1.6	-127.94	0.044832
234753_x_at	gb:AK026434.1 /DB_XREF=gi:10	AK026434	-1.54	-220.92	0.013638
240222_at	gb:BF347758 /DB_XREF=gi:1129	BF347758	-1.49	-206.12	0.007601
224261_at	gb:AF119917.1 /DB_XREF=gi:77	AF119917	-1.47	-60.28	0.008088
201564_s_at	fascin homolog 1, actin-bundling	NM_003088	-1.46	-53.9	0.047014
233658_at	gb:AK022413.1 /DB_XREF=gi:10	AK022413	-1.43	-188.48	0.031617
216051_x_at	gb:AK022045.1 /DB_XREF=gi:10	AK022045	-1.41	-98.26	0.048445
208404_x_at	gb:NM_000890.2 /DB_XREF=gi:4	NM_000890	-1.35	-128.16	0.036164
217126_at	Transcribed locus, moderately sir	K00627	-1.32	-314.05	0.025867
238616_at	Quinoid dihydropteridine reductas	AA844132	-1.32	-72.42	8.023784
226544_x_at	muted homolog (mouse)	AV734582	-1.28	-170.81	0.035781
214605_x_at	G protein-coupled receptor 1	AL046992	-1.25	-70.94	0.038136
223781_x_at	alcohol dehydrogenase 4 (class II	M15943	-1.24	-275.8	0.018678
234723_x_at	CDNA: FLJ21228 fis, clone COLO	AK024881	-1.23	-530	0.047218
1562905_at	Attractin-like 1	BC042015	-1.2	-110.15	0.037857

Table 3.8.8.5 Probe sets significantly up or down-regulated between MDA-F and MDA-F-TAX CM samples. The list is sorted by fold change. All down- regulated gene transcripts are shaded grey with no shading applied to the up- regulated gene transcripts.

H1299 +RPMI +MDA-F vs. H1299 TAX +RPMI TAX +MDA-F-TAX-SI

34 probe sets were significantly different when the parental CM samples H1299, RPMI and MDA-F were pooled and compared to the pooled taxol selected variant CM samples. The average expression of the parental CM group was compared to the average expression of the taxol- selected variant CM group. 8 gene transcripts were up-regulated in the taxol selected variants compared to the parental CM samples, with the majority of significantly changed gene transcripts down-regulated in the taxol selected CM samples.

18 of these probe sets corresponded to gene transcripts that were annotated with one gene ontology group designated as overrepresented. Both myosin IXB (208452_x_at) and Rho GTPase activating protein 26 (215955_x_at) belong to the Rho GTPase activator activity group, which contains only 43 gene transcripts giving an over representation p-value of 0.000385.

Of the 34 probe sets changed, eight have a known function of binding, including nucleotide binding; DNA binding and transcription factor activity and calcium ion binding function; calcium channel (208020_s_at) and similar to NOTCH2 protein (214722_at). The rest of the probe sets have known functions from transporter activities to regulation of cell cycle.

The KEGG Pyrimidine metabolism and GenMAPP Calcium_Channels pathways are represented with one gene from the gene list in each.

The chromosomal location was known for 11/ 34 gene transcripts found to be significantly different between the groups. The chromosomes that had at least one gene from the list on it were chromosomes 1 and 2 and chromosomes 5-8. No regions were determined as significantly over represented.

Whilst the parental CM samples as one group compared to the taxol- selected variants as one group yielded 34 gene transcripts deemed to be significantly changed, it was decided to look more closely at the significance of these gene transcripts in the individual parental vs.

taxol selected variant to see if any gene transcripts were consistently determined as significant in the individual selections as well as in the larger grouping listed in Table 3.8.8.6.

Table 3.8.8.7 shows that when the individual parental CM samples compared to their respective taxol selected CM samples, the gene transcripts found to be significantly changed (listed in Table 3.8.8.6) were not always found to be significantly changed or, in a few cases, even show the same patterns in change of expression in the individual groupings. For example the gene similar to NOTCH2 protein (214722_at) is significantly up-regulated when the parental CM samples are grouped against the taxol selected variants, however when the individual parental CM samples are tested against their taxol selected variant, although the three groups show up regulation of the gene, it is only found to be significant in the RPMI vs. RPMI TAX CM sample grouping.

Overall the gene transcripts do follow the same trend in change of expression in the individual parental vs. taxol selected variant groupings with 6 exceptions, 5 of which are in the MDA-F vs. MDA-F-TAX-SI group.

None of the 34 gene transcripts found to be significant in the larger parental grouping vs. taxol selected variant grouping listed in Table 3.8.8.6 were found to be significantly expressed in all three of the individual groupings listed in Table 3.8.8.7, especially in the MDA-F vs. MDA-F-ADR-SI. This meant that this comparison did not yield any candidate gene transcripts that may be possible candidate markers for taxol resistance. This may be a reflection of the fact that whilst the RPMI TAX and H1299 TAX cells were solely selected based on their resistance to taxol, the MDA-F-TAX-SI cells were a subgroup of the taxol selected MDA-F-TAX cells that were also found to be super invasive and so may have different mRNA gene expression profiles related more to super invasion than taxol-resistance.

probe set	gene	Accession	fold change	diff of means	P value
204347_at	adenylate kinase 3 /// adenylate I	AI653169	2.52	56.98	0.004665
241600_at	gb:AI270986 /DB_XREF=gi:38901	AI270986	1.8	55.01	0.034356
1559054_a_at	Protein phosphatase 1, regulatory	BC027905	1.68	66.59	0.009337
227732_at	ataxin 7-like 1	AB033044	1.65	76.59	0.005944
214722_at	similar to NOTCH2 protein	AW516297	1.29	98.28	0.019443
1559956_at	synaptotagmin VII	BF109699	1.26	68.8	0.04762
211792_s_at	cyclin-dependent kinase inhibitor	U17074	1.24	58.04	0.01278
242659_at	gb:T70285 /DB_XREF=gi:681433	T70285	-6.86	-163.47	0.029237
215955_x_at	Rho GTPase activating protein 26	Y10388	-5.01	-291.95	0.020765
224449_at	DNA-damage inducible protein 2 /	BC006011	-4.67	-129.86	0.013029
222359_x_at	Follicular lymphoma variant transl	BF573849	-4.19	-146.24	0.030241
242099_at	gb:T83938 /DB_XREF=gi:712226	T83938	-3.06	-71.19	0.019875
1552853_at	hypothetical protein FLJ32784	NM_144623	-3.04	-249.53	0.020049
1554128_at	hypothetical protein MGC39724	BC029594	-2.92	-248.41	0.00131
1563754_at	glutamate receptor, ionotropic, ka	AJ252246	-2.78	-165.54	0.042744
1567333_at	gb:AJ431618.1 /DB_XREF=gi:18	AJ431618	-2.75	-62.27	0.005717
206491_s_at	N-ethylmaleimide-sensitive factor	NM_003827	-2.74	-61.3	0.000438
237031_at	Full length insert cDNA clone YPI	AI743452	-2.69	-190.9	0.004832
240304_s_at	Transmembrane channel-like 5	BG484769	-2.39	-83.12	0.001933
208452_x_at	myosin IXB	NM_004145	-2.38	-56.43	0.004152
225489_at	transmembrane protein 18	AI720705	-2.37	-51.63	0.008922
1554375_a_at	nuclear receptor subfamily 1, grou	AF478446	-2.26	-58.17	0.022124
218698_at	likely ortholog of mouse monocyt	NM_015957	-2.25	-87.55	0.006169
208020_s_at	calcium channel, voltage-depende	NM_000719	-2.18	-115.69	0.002549
217512_at	gb:BG398937 /DB_XREF=gi:1325	BG398937	-2.15	-56.8	0.042082
229421_s_at	gb:BF435329 /DB_XREF=gi:1144	BF435329	-2.09	-62.55	0.045025
211238_at	a disintegrin and metalloproteinas	AF215824	-2.04	-88.86	0.007309
229453_at	gb:N64025 /DB_XREF=gi:121185	N64025	-2.04	-55.8	0.003252
210103_s_at	forkhead box A2	AB028021	-1.86	-316.41	0.045586
241593_x_at	gb:BF476913 /DB_XREF=gi:1154	BF476913	-1.53	-76.9	0.025298
204158_s_at	T-cell, immune regulator 1, ATPa	NM_006019	-1.52	-238.44	0.036178
243489_at	gb:BF514098 /DB_XREF=gi:1159	BF514098	-1.48	-499.9	0.020936
1558875_at	sterol regulatory element binding	S66168	-1.38	-55.19	0.032389
219358_s_at	centaurin, alpha 2	NM_018404	-1.27	-75.57	0.019505

Table 3.8.8.6 Probe sets significantly up or down-regulated between H1299, RPMI and MDA-F CM samples vs. H1299 TAX, RPMI TAX and MDA-F-TAX-SI CM samples. The list of gene transcripts is sorted by fold change. All down- regulated gene transcripts are shaded grey with no shading applied to the up- regulated gene transcripts.

			H1299 v H1299 TAX			RPMI v RPMI TAX			MDA-F v MDA-F-TAX-SI		
probe set	gene	Accession	fold change	difference of means	P value	fold change	difference of means	P value	fold change	difference of means	P value
204347_at	adenylate kinase 3 /// ad	AI653169	4.32	104.06	0.001076	10.65	72.6	0.010142	-1.08	-4.7	0.687948
241600_at	gb:AI270986 /DB_XREF=	AI270986	-1.39	-26.18	0.412058	4.63	136.58	0.008626	2.08	69.92	0.111566
1559054_a_at	Protein phosphatase 1, r	BC027905	2.19	93.27	0.070647	1.55	50.11	0.052653	1.42	56.65	0.544534
227732_at	ataxin 7-like 1	AB033044	2.23	83.8	0.009627	1.99	107.7	0.109151	1.27	46.41	0.033728
214722_at	similar to NOTCH2 protei	AW516297	1.09	33.83	0.408408	1.94	230.32	0.081387	1.17	65.39	0.502534
1559956_at	synaptotagmin VII	BF109699	1.26	63.26	0.114592	1.6	152.04	0.064288	-1.07	-19.16	0.757612
211792_s_at	cyclin-dependent kinase	U17074	1.12	28.39	0.282474	1.58	117.03	0.091668	1.13	34.07	0.350073
242659_at	gb:T70285 /DB_XREF=gi	T70285	-8.25	-241.92	0.114172	-21.17	-242.43	0.418001	-1.54	-22.97	0.690136
215955_x_at	Rho GTPase activating p	Y10388	-4.09	-333.2	0.088783	-15.17	-596.22	0.229522	-1.6	-39.86	0.315857
224449_at	DNA-damage inducible p	BC006011	-5.11	-201.62	0.013541	-8.89	-180.77	0.290179	-1.46	-15.83	0.542548
222359_x_at	Follicular lymphoma vari	BF573849	-3.63	-178.58	0.097553	-11.32	-299.73	0.245289	-1.16	-6.46	0.770529
242099_at	gb:T83938 /DB_XREF=gi	T83938	-3.24	-110.99	0.000656	-6.1	-108.21	0.229696	-1.1	-3.17	0.869932
1552853_at	hypothetical protein FLJ	NM_14462	-3.18	-367.12	0.02148	-5.77	-410.58	0.264662	-1.22	-22.77	0.537239
1554128_at	hypothetical protein MGC	BC029594	-2.59	-263.25	0.044604	-5.33	-408.63	0.000983	-1.95	-117.65	0.148508
1563754_at	glutamate receptor, ionot	AJ252246	-2.1	-138.83	0.012696	-6.59	-349.97	0.335966	-1.03	-2.48	0.935712
1567333_at	gb:AJ431618.1 /DB_XRE	AJ431618	-2.53	-61.32	0.020938	-4.63	-107.01	0.291644	-1.85	-31.63	0.340363
206491_s_at	N-ethylmaleimide-sensiti	NM_00382	-2.35	-66.73	0.025851	-5.89	-90.17	0.016714	-1.79	-31.37	0.191407
237031_at	Full length insert cDNA c	AI743452	-2.49	-195.34	0.124621	-4.96	-336.06	0.19093	-1.65	-80.16	0.220376
240304_s_at	Transmembrane channel	BG484769	-1.92	-70.37	0.123513	-4.79	-157.25	0.03476	-1.62	-36.77	0.289076
208452_x_at	myosin IXB	NM_00414	-2.47	-74.53	0.005262	-2.91	-68.73	0.01053	-1.48	-16.63	0.341669
225489_at	transmembrane protein 1	AI720705	-2.53	-67.24	0.095783	-2.82	-69.55	0.222022	-1.94	-26.95	0.287469
1554375_a_at	nuclear receptor subfami	AF478446	-2.49	-91.55	0.01629	-3.8	-95.13	0.036448	-1.35	-14.53	0.577499
218698_at	likely ortholog of mouse	NM_01595	-2.31	-98.76	0.187531	-3.89	-147.93	0.007462	-1.26	-22.94	0.446629
208020_s_at	calcium channel, voltage	NM_00071	-2.22	-136.12	0.046526	-3.59	-188.58	0.006098	-1.29	-33.4	0.463812
217512_at	gb:BG398937 /DB_XREF	BG398937	-2.57	-83.74	0.212319	-2.94	-89.82	0.180323	-1.13	-6.48	0.510801
229421_s_at	gb:BF435329 /DB_XREF	BF435329	-2.19	-91.1	0.086264	-3.98	-120.27	0.070546	1.1	4.71	0.721226
211238_at	a disintegrin and metallo	AF215824	-2.16	-106.79	0.044771	-3.35	-160.88	0.156656	-1.12	-11.94	0.699475
229453_at	gb:N64025 /DB_XREF=g	N64025	-2.3	-65.64	0.094404	-2.37	-74.47	0.18023	-1.58	-33.31	0.171475
210103_s_at	forkhead box A2	AB028021	-1.65	-289.39	0.014169	-3.75	-792.37	0.27893	-1	-0.09	0.999062
241593_x_at	gb:BF476913 /DB_XREF	BF476913	-1.66	-103.62	0.029323	-2.16	-151.44	0.132736	1.03	4.39	0.910418
204158_s_at	T-cell, immune regulator	NM_00601	-1.69	-367.9	0.02937	-2.16	-412.58	0.24005	1.12	52.79	0.600987
243489_at	gb:BF514098 /DB_XREF	BF514098	-1.04	-65.98	0.825766	-1.77	-572.14	0.128885	-2.2	-882.52	0.012861
1558875_at	sterol regulatory element	S66168	-1.26	-39.26	0.052425	-1.92	-122.77	0.274113	-1.02	-3.94	0.888329
219358_s_at	centaurin, alpha 2	NM_01840	-1.15	-38.44	0.511605	-1.62	-166.05	0.012815	-1.14	-43.34	0.06136

Table 3.8.8.7 The 34 gene transcripts determined to be significantly differentially expressed between parental CM group (H1299 + RPMI+ MDA-F) against the taxol selected variants (H1299 TAX+ RPMI TAX+ MDA-F-ADR-TAX-SI) are examined in the individual parental vs. taxol selected variant groups. The list is sorted based on gene placement in Table 3.8.8.6. The probe set, gene and Accession columns are shaded according to the fold change of the grouped parental vs. drug selected variant in Table 3.8.8.6, no shading= up-regulated, grey shading = down-regulated. The fold change, difference of means and p-value information listed in the three columns H1299 vs. H1299 TAX, RPMI vs. RPMI TAX and MDA-F vs. MDA-F-ADR-SI are coloured based on the fold change of the gene transcripts in the individual groupings e.g. H1299 vs. H1299 TAX.

3.8.9 Gene Ontology of probe sets significantly changed between parental CM and drug selected variant CM

The probe sets that showed a change in expression greater than 100 fluorescent units as well as having a fold change greater than 1.2 and a p-value less than 0.05 in the drug selected variants compared to the parental CM samples were submitted to netaffx (see section 2.10.82). The molecular function of these probe sets were graphed in a pie chart in (see Fig 3.8.9.1).

The pie-chart in Fig. 3.8.9.1 is quite similar to the one generated for the probe sets assigned as Present in all CM samples (see Fig. 3.8.5.1). The amount of represented gene transcripts in the categories shift though, after the categories binding and catalytic activity, in the gene transcripts called Present in all CM samples the next largest represented groups were structural molecule activity, signal transducer activity, transcription regulator activity, transporter activity, enzyme regulator activity, motor activity and translation activity. In Fig. 3.8.9.1, the most represented categories after binding and (a smaller) catalytic activity group, are transcription regulator activity, signal transducer activity, transporter activity and a noticeably smaller structural molecule activity group. Both pie charts show that a high percentage of the probe sets in the list are un-annotated (Figs. 3.8.5.1 and 3.8.9.1).

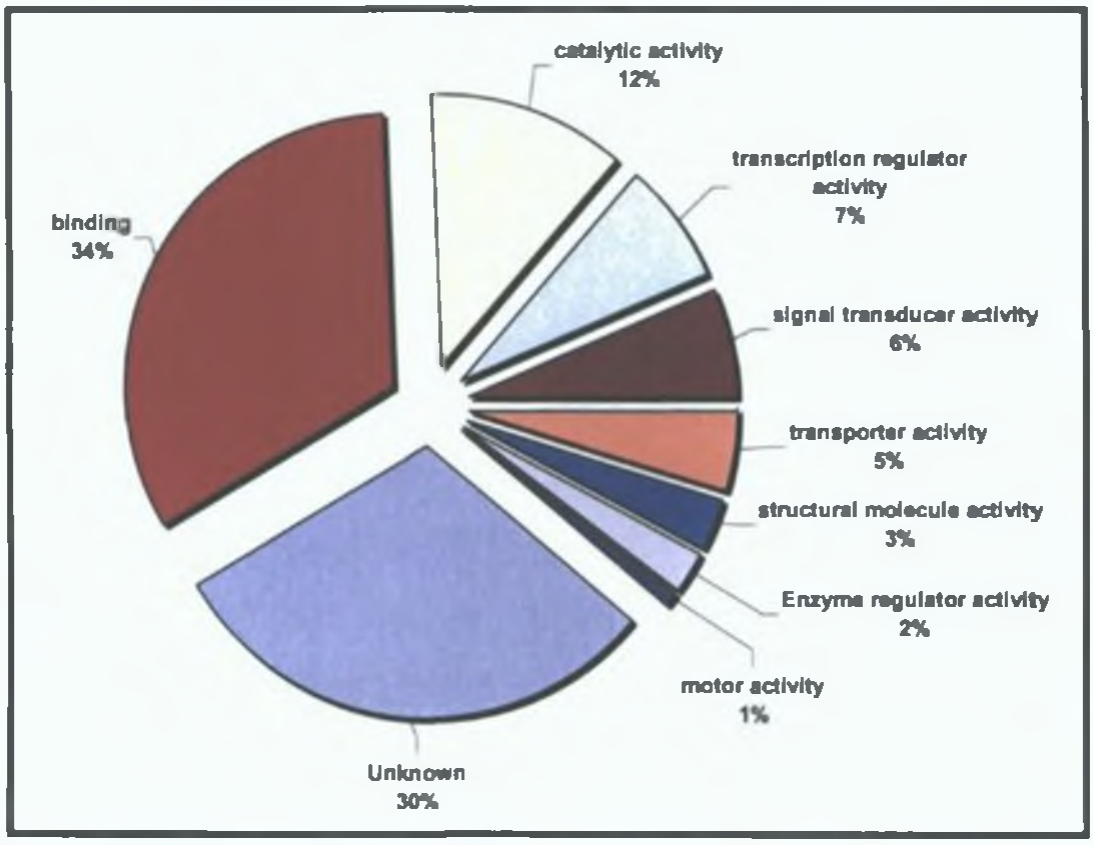


Figure 3.8.9.1 Gene ontology of significantly changing gene transcripts between parental and drug selected variant CM samples (n=88).

3.8.10 Number of gene transcripts significantly different between individual CM sample groups

The numbers of gene transcripts found to be different between each CM sample group are listed in Table 3.8.10.1. The CM sample groups with the highest amount of differentially expressed gene transcripts are the parental RPMI CM samples compared to the MDA-F-ADR-SI CM samples. The next 2 groups of CM samples with the highest number of differentially expressed gene transcripts also involves the parental RPMI CM group and the MDA-F-ADR-SI group, the former in comparison to its drug selected variant group RPMI TAX and the latter with the RPMI ML CM group.

There seems to be no pattern to how different the sample groups are from each other as determined by the number of gene transcripts significantly differentially expressed between the sample groups.

SAMPLE	# significant gene transcripts
RPMI v MDA-F-ADR-SI	167
RPMI v RPMI TAX	150
RPMI ML v MDA-F-ADR-SI	133
H1299 v RPMI	128
RPMI TAX v MDA-F-ADR-SI	125
H1299 TAX v RPMI	123
H1299 v MDA-F	110
H1299 v MDA-F-ADR-SI	106
MDA-F-ADR-SI v MDA-F-TAX-SI	99
MDA-F v MDA-F-ADR-SI	94
H1299 v RPMI TAX	93
RPMI v MDA-F	92
H1299 v MDA-F-TAX-SI	81
RPMI v RPMI ML	73
H1299 TAX v MDA-F-TAX	71
RPMI TAX v MDA-F	66
RPMI TAX v MDA-F-TAX-SI	66
H1299 TAX v MDA-F	65
RPMI v MDA-F-TAX-SI	63
RPMI ML v RPMI TAX	59
RPMI ML v MDA-F	56
H1299 v RPMI ML	55
H1299 TAX v MDA-F-ADR-SI	48
H1299 TAX v RPMI TAX	44
H1299 v H1299 TAX	43
H1299 TAX v RPMI ML	43
MDA-F v MDA-F-TAX-SI	39
RPMI ML v MDA-F-TAX-SI	38

Table 3.8.10.1 Number of gene transcripts *significantly* different between all CM sample groups. Each of the sample groups were compared with each other using dChip and the criteria for selection as significantly different were; expression difference>100, fold change>1.2 and p-value<0.05. The list is sorted in order of the sample groups with the most differences.

3.8.11 Validation of Microarray Data (Part 2)

Gene transcripts significantly up/ down- regulated were selected for confirmation of change in trend of expression by RT-PCR. Probes sets that showed a change in expression of >100 fluorescent units coupled with a fold change >1.2 and a p-value<0.05 between parental CM samples and their drug selected variant samples were assessed for suitability.

From this list, only probes sets with the suffix "_at" were considered as they indicated a unique probe set that only hybridised to one gene. From this shortened list all mRNA sequences connected to the probe sets were entered in to a BLAT search (see Section 2.11.1) to check that exon- intron boundaries were known enabling selection of mRNA specific primers to target this region. Of the probe sets that successfully passed all of these criterion, ten were randomly chosen but they included probe sets that had differing ranges of fold change between sample groups e.g. from 1.3 to 5.6 fold.

Due to the results obtained in the validation of the Presence/ Absence call i.e. that gene transcripts with Absent calls assigned to their corresponding probe sets were amplifiable by RT-PCR, Presence/ Absent call of a probe set in any/ all CM samples had no influence on the probe sets chosen for validation. RT-PCRs were carried out to validate trend of change onely, i.e. up/ down regulation The microarray expression data, RT-PCR results and densitometry analysis for all probe sets chosen for validation are shown in Figs. 3.8.11.2-10. Table 3.8.11.1 summarises the results showing that 5/10 of the RT-PCRs validated the trend of change also found in the microarray results. However, results shown in Section 3.8.11.1 indicate that some of this may be attributed to amplification of genomic DNA. Details of the primer sets chosen and the region of the mRNA transcript amplified along with the region of the transcript targeted by the Affymetrix array chip, are summarised in Table 2.11.1.3 (Section 2.11.1).

Hyp Protein FLJ32784 mRNA (1552853_at)

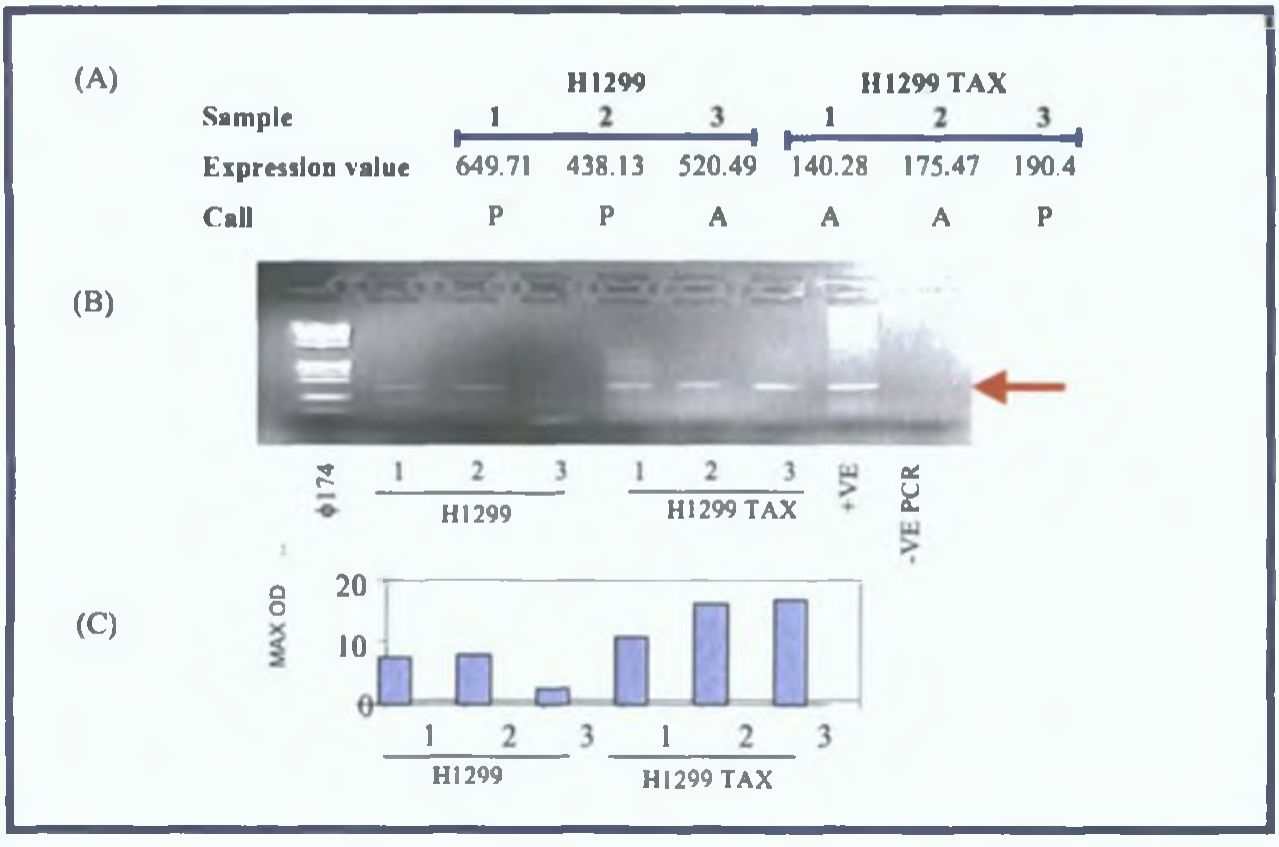


Figure 3.8.11.2 (A) Expression levels and Presence/ Absent call for probe set 1552853_at (Hyp Protein FLJ32784). Hyp Protein FLJ32784 mRNA is 2.5 fold down regulated in the H1299 TAX CM samples compared to the H1299 CM samples with a significant p- value of 0.02. (B) RT-PCR of CM samples for HYP Protein FLJ32784 (158bp). Hyp Protein FLJ32784 mRNA (arrow) was expressed, to some extent, by all H1299 and H1299 TAX samples analysed. Amplified product was detected in the positive control (RPMI TAX cell RNA) and was undetected in the negative control samples (-VE = PCR reaction with H_20 instead of cDNA as control). (C) Densitometry analysis of RT-PCR for HYP Protein FLJ32784 mRNA indicating that the transcript is upregulated in the H1299 TAX samples compared to the H1299 samples. CM samples are plotted on the X-axis with maximum OD for each PCR band plotted on the Y-axis.

Small Adipocyte factor 1 (SMAF1) (1554128_at)

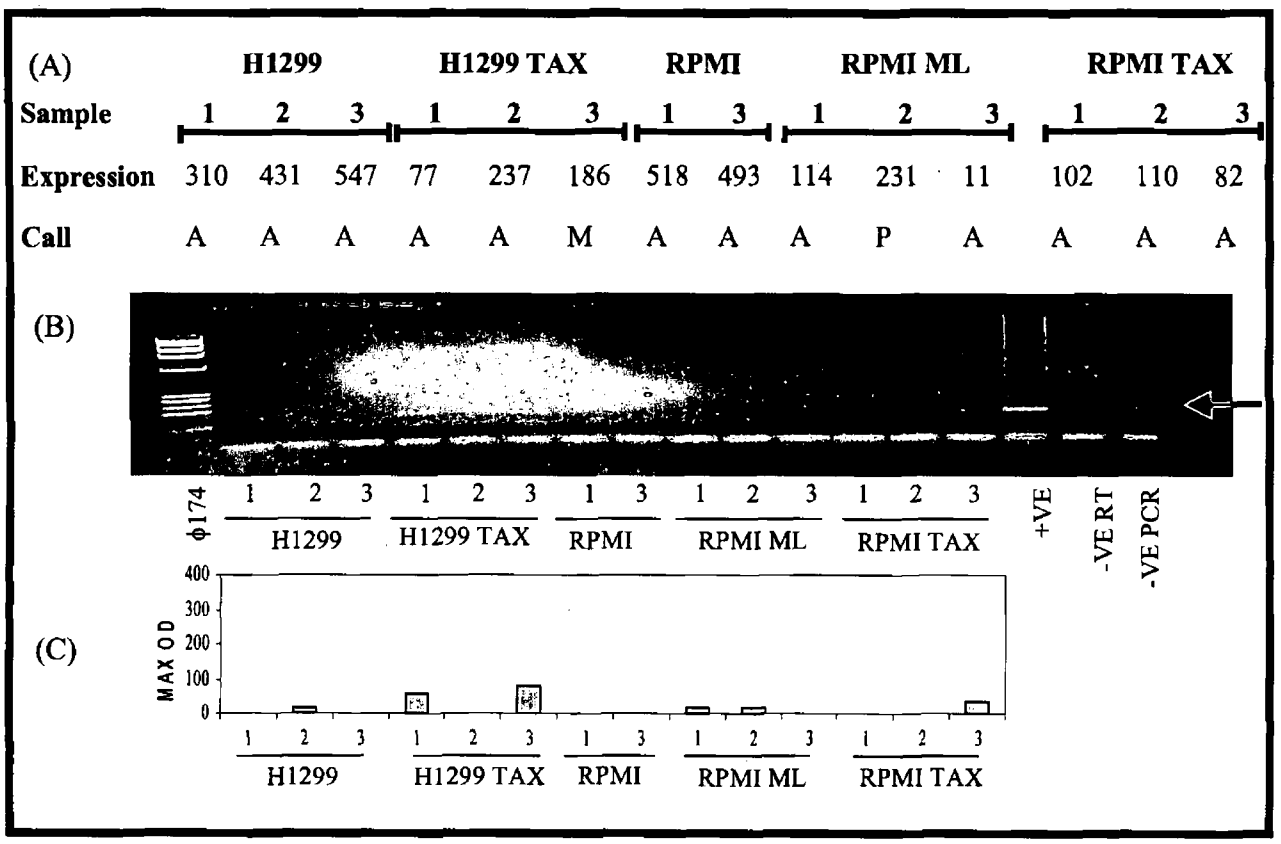


Figure 3.8.11.2(A) Expression levels and Presence/ Absent call for probe set 1554128_at (SMAF1 mRNA). SMAF1 mRNA is 2.6 fold down-regulated in H1299 TAX CM compared to H1299 CM with a p- value of 0.044, and 4 fold down-regulated in RPMI ML CM samples compare to RPMI CM samples with a p-value of 0.02 and finally 5 fold down-regulated in RPMI TAX CM samples compared to RPMI CM samples with a p-value of 0.0009. (B) RT-PCR of SMAF 1 mRNA (152bp). SMAF 1 mRNA (black arrow) transcripts were expressed in H1299 TAX 1, H1299 TAX 3, RPMI TAX 3 CM samples, with faint bands in H1299 2, RPMI ML 1 and RPMI ML 2 CM samples. Amplified product was detected in the positive control (RPMI TAX cell RNA), and was undetected in the negative control samples (-VE RT = RT reaction with = H_20 instead of RNA, -VE PCR = PCR reaction with H_20 instead of cDNA as control). (C) Densitometry analysis of RT-PCR for SMAF1 in (B). CM samples are plotted on the X-axis with maximum OD for each PCR band plotted on the Y-axis.

Clone IMAGE: 5528716 (1561573_at)

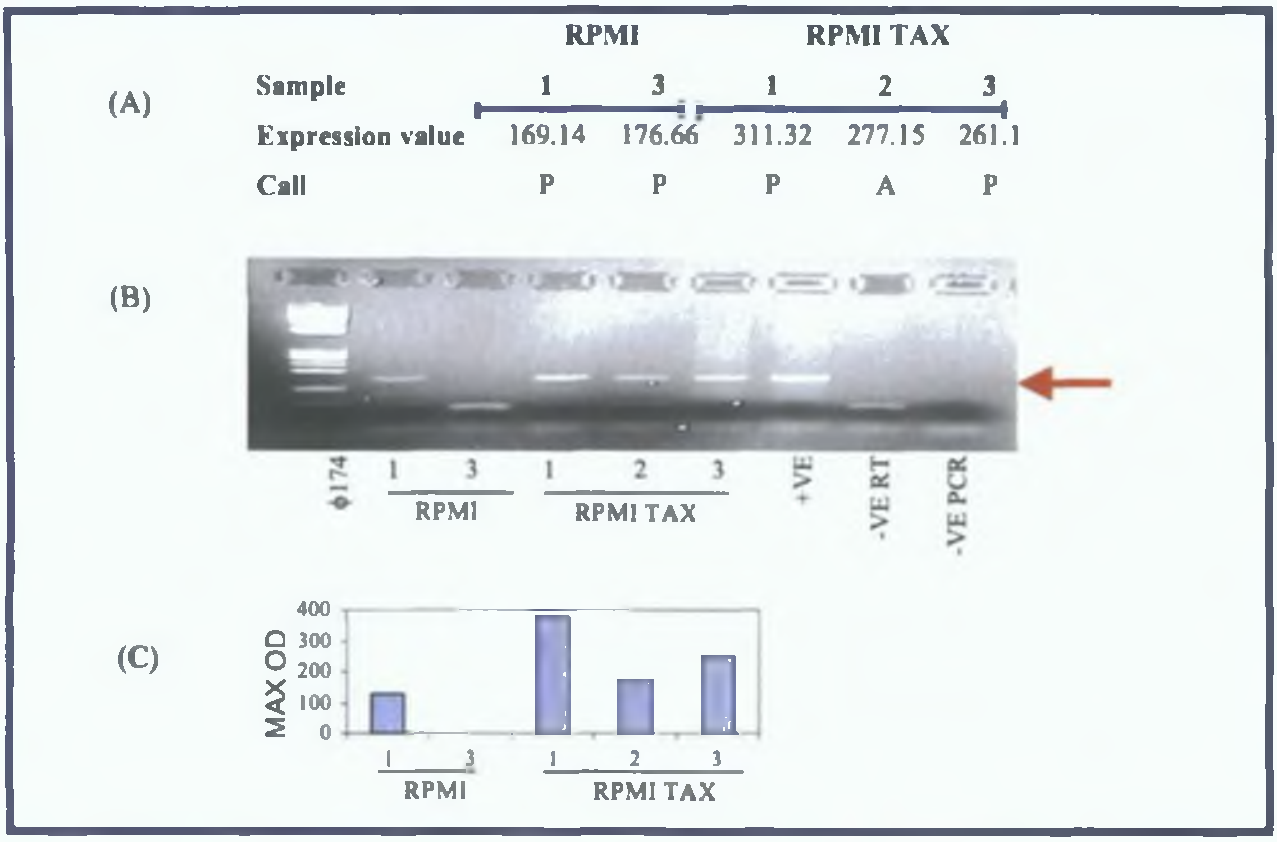


Figure 3.8.11.3 (A) Expression levels and Presence/ Absent call for probe set 1561573_at. The probe set is called Present in all RPMI and RPMI TAX CM samples except RPMI TAX 2. Clone IMAGE: 5528716 mRNA is 1.65 fold up-regulated in the RPMI TAX CM samples compared to the RPMI CM samples with a p- value of 0.04. (B) RT-PCR of CM samples for Clone IMAGE: 5528716 (152bp). The Clone IMAGE: 552876 mRNA (arrow) is expressed in all samples except RPMI 3 CM. Amplified product was detected in the positive control (RPMI TAX cell RNA), and was undetected in the negative control samples (-VE RT = RT reaction with = H_20 instead of RNA, -VE PCR = PCR reaction with H_20 instead of cDNA as control). (C) Densitometry analysis of RT-PCR for Clone IMAGE: 5528716. CM samples are plotted on the X-axis with maximum OD for each PCR band plotted on the Y-axis.

ATP-binding cassette, subfamily (MDR/TAP), member 4 (ABCB4) (1570505_at)

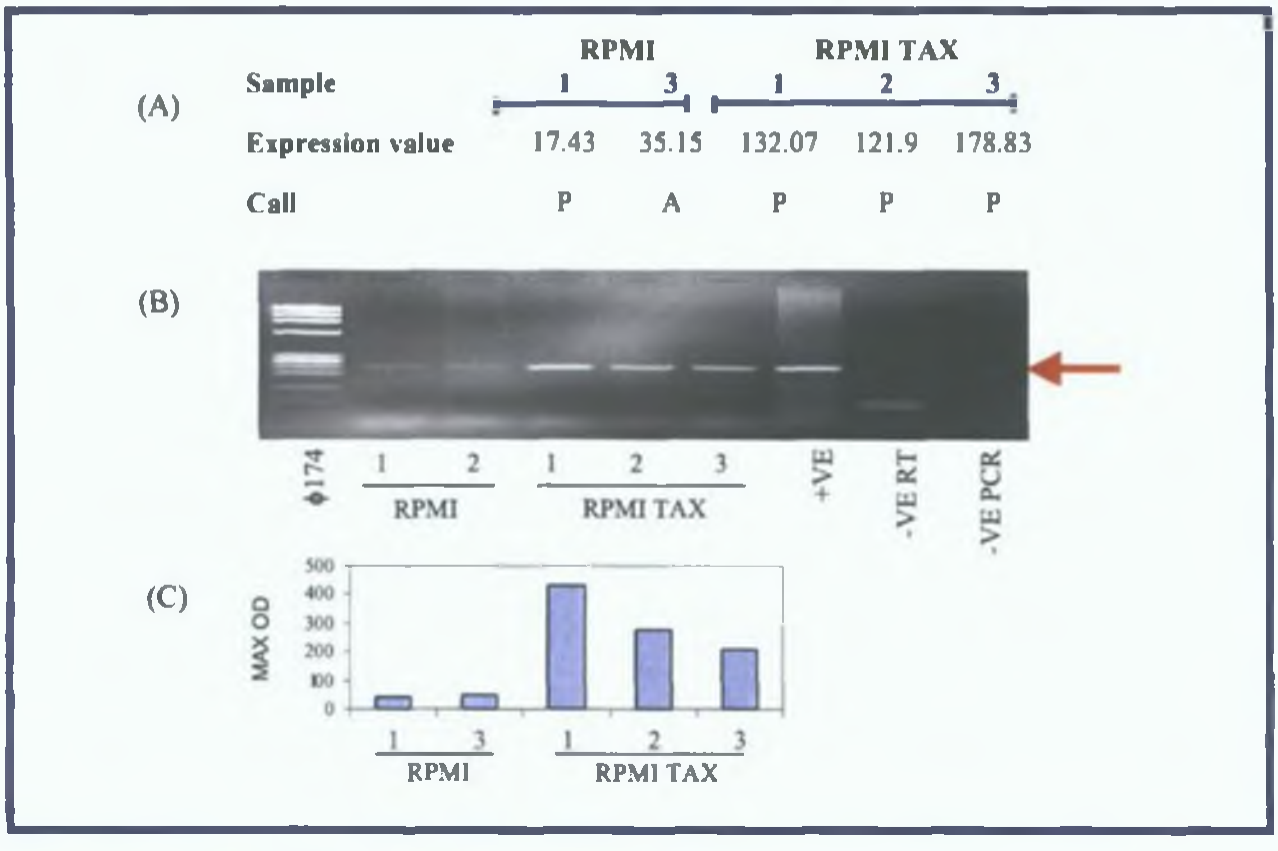


Figure 3.8.11.4 (A) Expression levels and Presence/ Absent call for probe set 1570505_at. The probe set is called Present in all samples with the exception of RPMI 3. ABCB 4 mRNA is 5.6 fold up-regulated in the RPMI TAX CM samples compared to the RPMI CM samples with a p-value of 0.011. (B) RT-PCR of CM samples for ABCB4 mRNA (239 bp). ABCB4 mRNA gene transcripts (arrow) were amplifiable in all RPMI and RPMI TAX CM samples. Amplified product was detected in the positive control (RPMI TAX cell RNA), and was undetected in the negative control samples (-VE RT = RT reaction with = H_20 instead of RNA, -VE PCR = PCR reaction with H_20 instead of cDNA as control). (C) Densitometry analysis of RT-PCR for ABCB4. CM samples are plotted on the X-axis with maximum OD for each PCR band plotted on the Y-axis.

cDNA FLJ34465 fis, clone HLUNG2003061 (1557996_at)

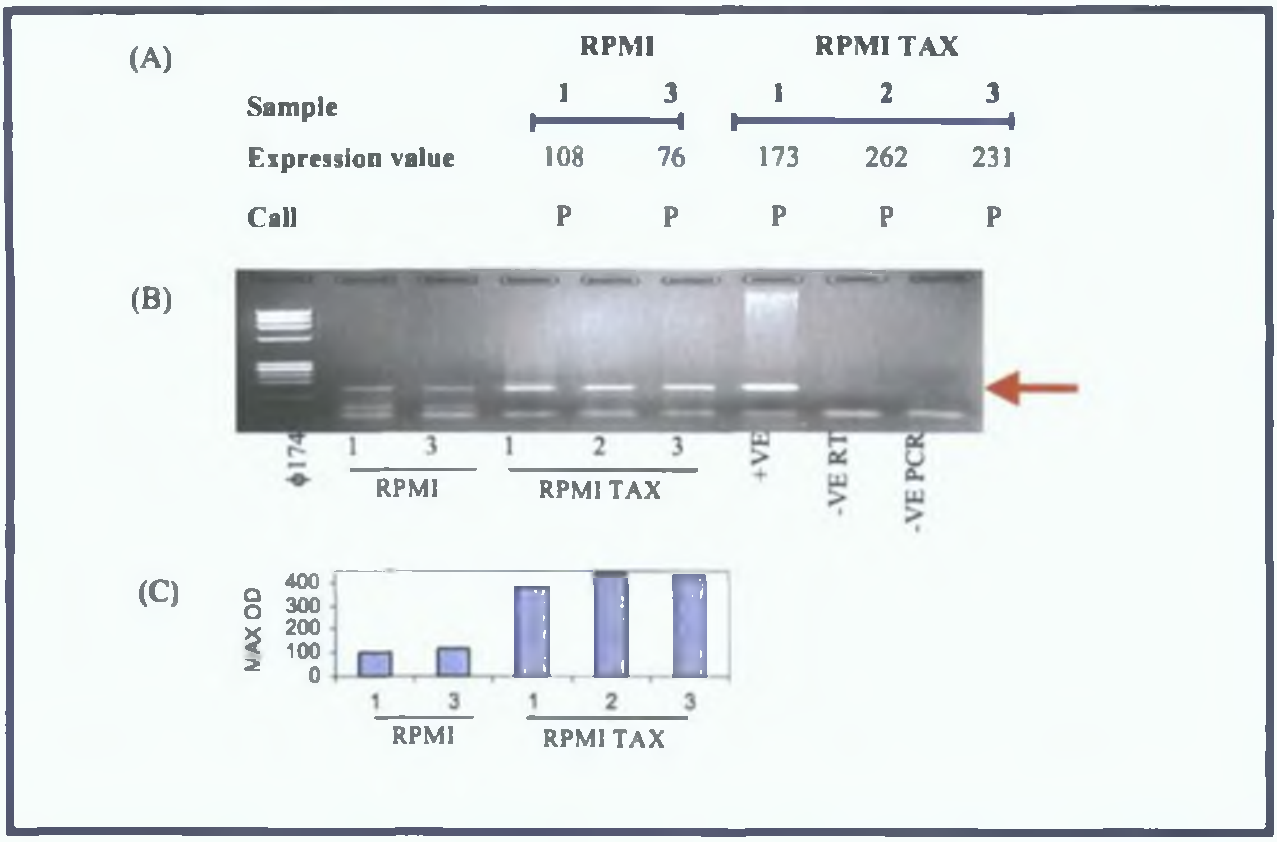


Figure 3.8.11.5 (A) Expression levels and Presence/ Absent call for probe set 1557996_at. The probe set is called Present in all RPMI and RPMI TAX CM samples and is calculated to be 2.4 fold up-regulated in the RPMI TAX CM samples compared to the R CM samples with a p- value of 0.03. (B) HLUNG2003061 mRNA gene transcripts (arrow) are expressed in all R and RT CM samples. Amplified product was detected in the positive control (RPMI TAX cell RNA), and was undetected in the negative control samples (-VE RT = RT reaction with = H_20 instead of RNA, -VE PCR = PCR reaction with H_20 instead of cDNA as control). (C) Densitometry analysis of RT-PCR for clone HLUNG2003061. CM samples are plotted on the X-axis with maximum OD for each PCR band plotted on the Y-axis.

Neighbour of BRCA1 gene (NBR2) (207631_at)

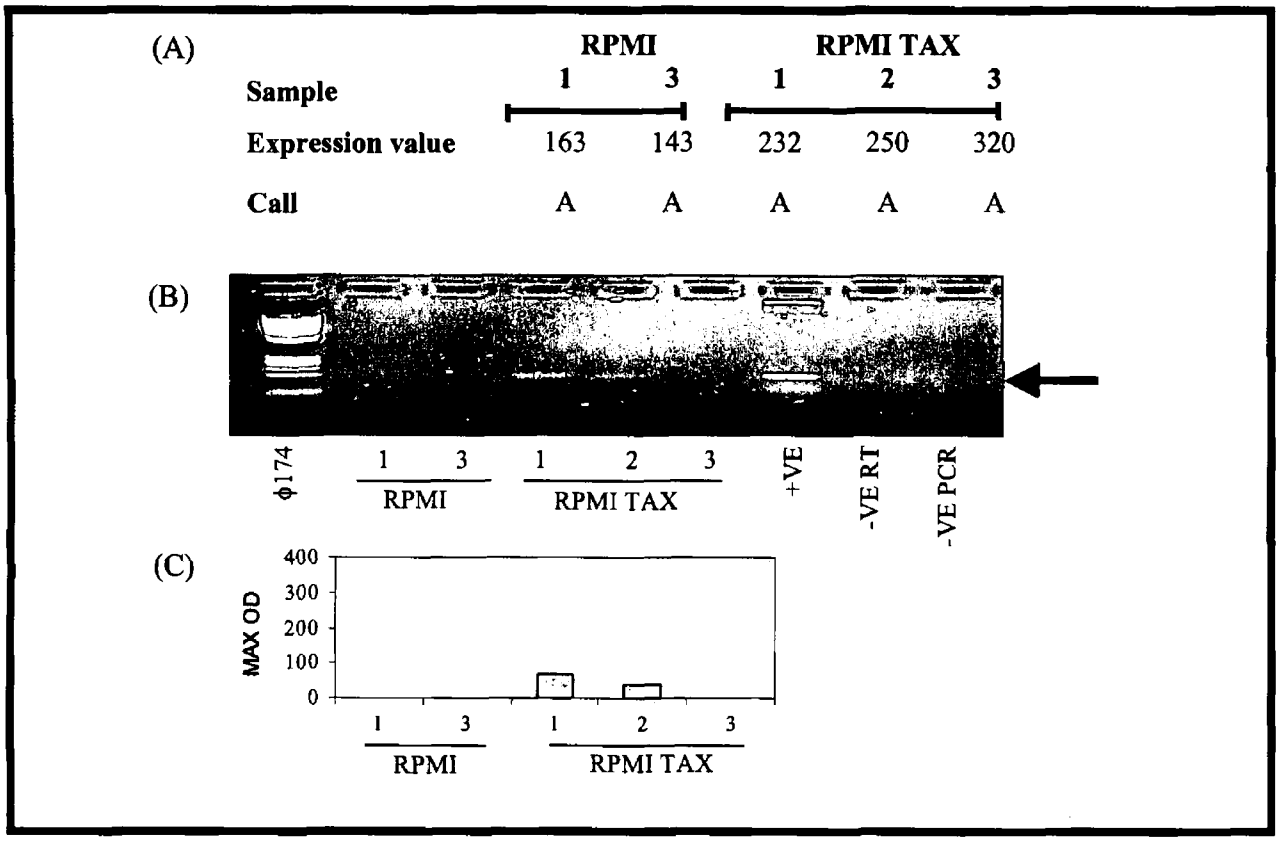


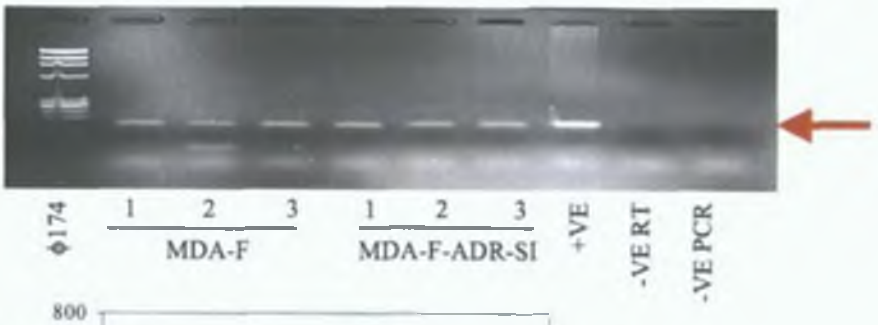
Figure 3.8.11.6 (A) Expression levels and Presence/ Absent call for probe set 207631_at. The probe set 207631_at was assigned as Absent in all RPMI and RPMI TAX CM samples. Despite this there was still a 1.76 fold up-regulation of the probe set in RPMI TAX CM samples compared to RPMI CM samples with a p-value of 0.037. (B) RT-PCR of CM samples for NBR2 mRNA (179bp). NBR2 mRNA gene transcripts (arrow) were only faintly detected in RPMI TAX 1 and RPMI TAX 2 CM samples. Amplified product was detected in the positive control (RPMI TAX cell RNA), and was undetected in the negative control samples (-VE RT = RT reaction with = H_20 instead of RNA, -VE PCR = PCR reaction with H_20 instead of cDNA as control). (C) Densitometry analysis of RT-PCR for NBR2. CM samples are plotted on the X-axis with maximum OD for each band plotted on the Y-axis.

Catechol-O-methyltransferase (COMT) (208817_at)

(A)

Sample	MDA-F 1	MDA-F 2	MDA-F 3	MDA-F-ADR-SI 1	MDA-F-ADR-SI 2	MDA-F-ADR-SI 3
Expression value	536	499	519	697	732	714
Call	A	A	A	A	A	A

(B)



(C)

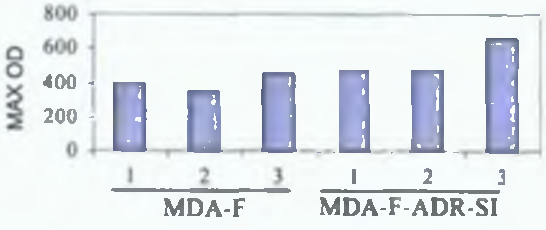


Figure 3.8.11.7 (A) Expression levels and Presence/ Absent call for probe set 208817_at. The probe set 208817_at is called Absent in all MDA-F and MDA-F-ADR-SI CM samples. The expression values however are fairly high ranging from 519 to 732 and there is an increase in expression of 1.37 fold with a p-value of 0.0007. (B) RT-PCR of CM samples for COMT mRNA (151bp). COMT mRNA gene transcripts (arrow) were amplifiable in all samples tested. Amplified product was detected in the positive control (RPMI TAX cell RNA), and was undetected in the negative control samples (-VE RT = RT reaction with = H_20 instead of RNA, -VE PCR = PCR reaction with H_20 instead of cDNA as control). (C) Densitometry analysis of RT-PCR for COMT. CM samples are plotted on the X-axis with maximum OD for each band plotted on the Y-axis.

Homeo box HB9 (HLXB9) (214614_at)

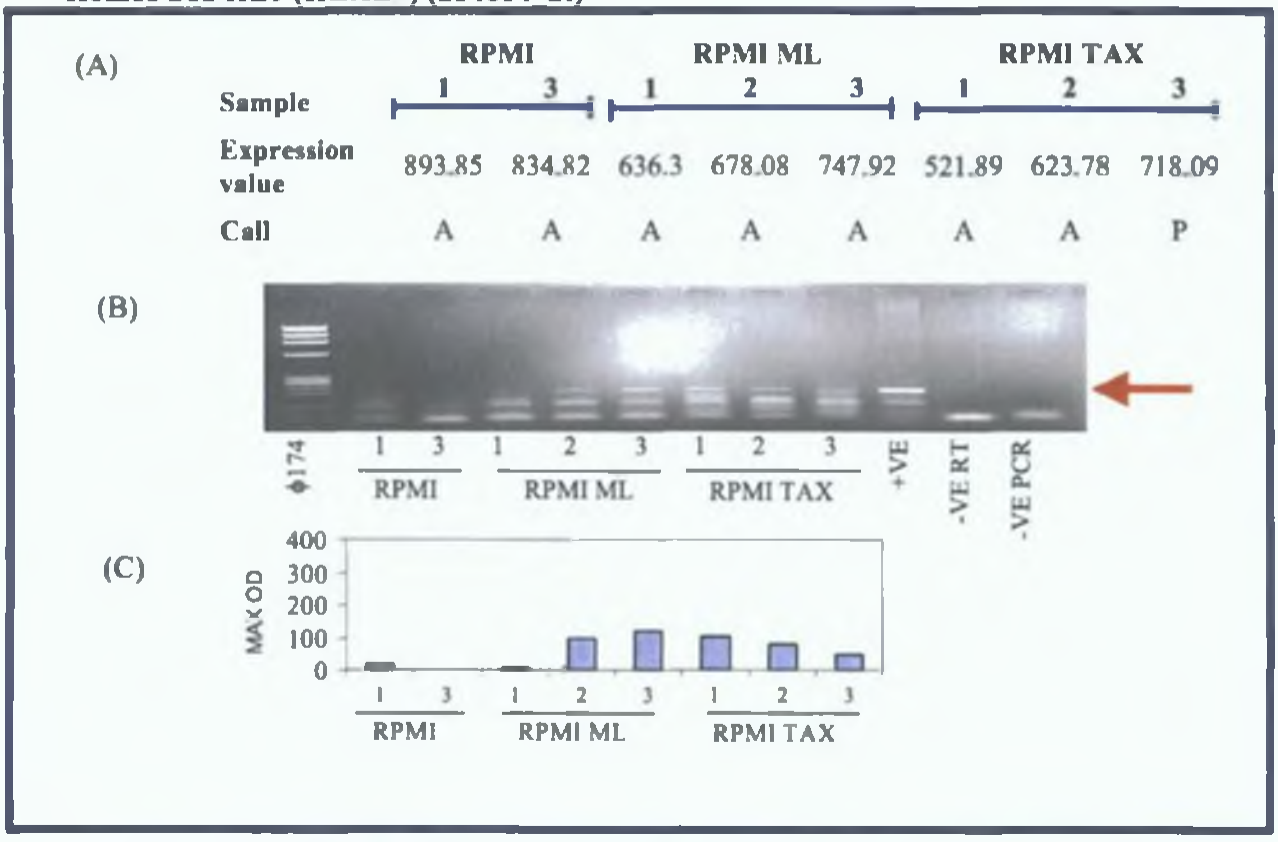


Figure 3.8.11.8 (A) Expression levels and Presence/ Absent call for probe set 214614_at. HLXB9 mRNA is assigned as Absent in all samples with the exception of RPMI TAX 3 CM. The expression values for this probe set are high in all samples ranging from 521 to 893. HLXB9 mRNA is significantly down-regulated in both RPMI ML and RPMI TAX CM samples compared to RPMI CM samples, with a fold change (and p- value) of 1.26 (0.033) and 1.39 (0.036) respectively. (B) RT-PCR of CM samples for HLXB9 mRNA (arrow) (180bp). Amplified product was detected in the positive control (RPMI TAX cell RNA), and was undetected in the negative control samples (-VE RT = RT reaction with = H_20 instead of RNA, -VE PCR = PCR reaction with H_20 instead of cDNA as control). (C) Densitometry analysis of RT-PCR for HLXB9. CM samples are plotted on the X-axis with maximum OD for each PCR band plotted on the Y-axis.

DNA- damage inducible protein 2 (DDI2) (224449_at)

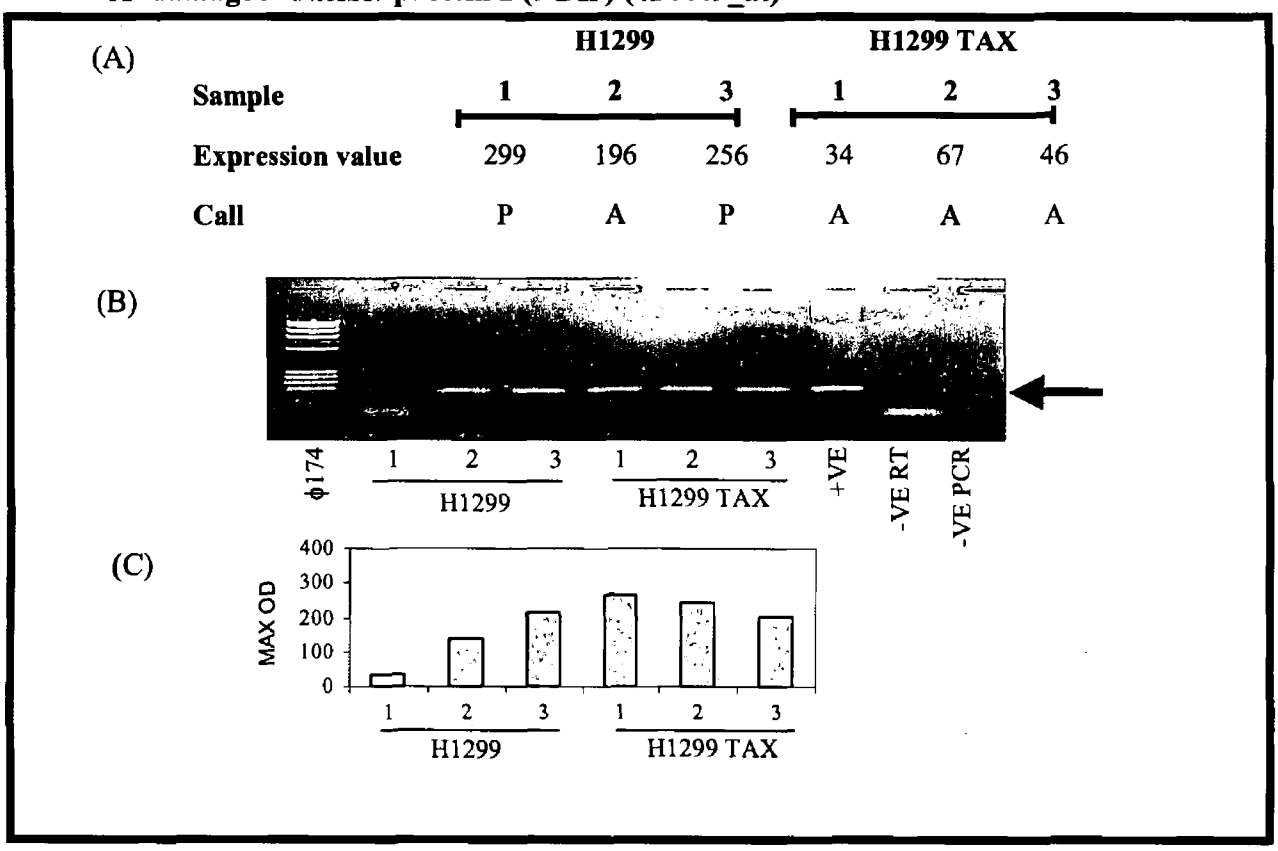


Figure 3.8.11.9 (A) Expression levels and Presence/ Absent call for probe set 224449_at. The probe set 224449_at is assigned as Present in 2/3 H1299 CM samples and Absent in 3/ 3 H1299 TAX CM samples. The probe set is also deemed to be 5 fold down-regulated in H1299 TAX CM samples compared to H1299 CM samples with a p-value of 0.013. (B) RT-PCR of CM samples for DDI2 mRNA (163bp). DDI2 mRNA (arrow) was detected in all samples tested. Amplified product was detected in the positive control (RPMI TAX cell RNA), and was undetected in the negative control samples (-VE RT = RT reaction with = H_20 instead of RNA, -VE PCR = PCR reaction with H_20 instead of cDNA as control). (C) Densitometry analysis of RT-PCR for DDI2. CM samples are plotted on the X-axis with maximum OD for each PCR band plotted on the Y-axis.

APAF1 interacting protein (APIP) (218698_at)

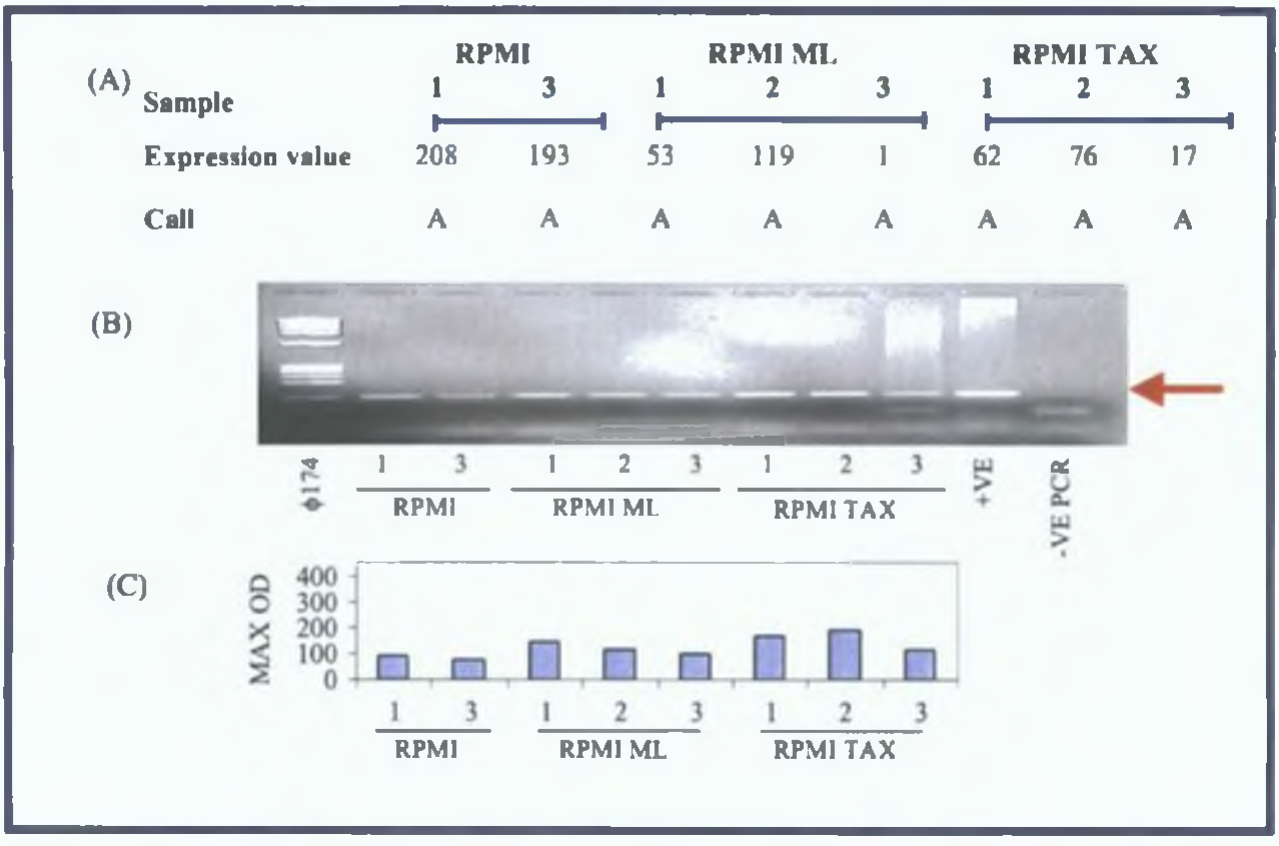


Figure 3.8.11.10 (A) Expression levels and Presence/ Absent call for probe set 218698_at. The probe set 218698_at is assigned an Absent call in all CM samples. The expression level of the probe set however is deemed to be down-regulated by just under 4 fold in both RPMI ML and RPMI TAX CM samples compared to the RPMI CM samples with p-values of 0.04 and 0.007 respectively. (B) RT-PCR of CM samples for APIP mRNA (119bp). APIP mRNA gene transcripts (arrow) were amplifiable in all CM samples tested. Amplified product was detected in the positive control (RPMI TAX cell RNA), and was undetected in the negative control samples (-VE PCR = PCR reaction with H_20 instead of cDNA as control). (C) Densitometry analysis of RT-PCR for APIP. CM samples are plotted on the X-axis with maximum OD for each PCR band plotted on the Y-axis.

Table 3.8.11.11 shows that of the ten probe sets chosen for the validation study, five validated the trend of change found by the microarray analysis i.e. up/ down regulation in the drug selected variant CM compared to the parental CM samples. Of the five probe sets that validated the microarray data trends, four had primers that amplified regions close to or in the Affymetrix target region, three had a majority Present call among the samples tested, and three were from the set of 4 gene transcripts that were functionally annotated. The fourth well annotated mRNA, HLXB9, was called Absent in 7/8 CM samples tested and the region amplified by RT-PCR was not close to the region targeted by the Affymetrix probe sets.

Probe ID	Accession No. (NCBI)	Gene Name	PCR target close to/ overlap Affy target region?	Gene function annotated?	How many samples with Present call?	Validate Array Result?
1552853_at	NM_144623	Hyp Protein FLJ32784	Y	N	3/6	N
1554128_at	NM_001018082	small adipocyte factor 1 (SMAF1)	N	N	1/14	N
1561573_at	BC039476	Clone IMAGE:5528716	Y	N	4/5	Y
1570505_at	BC020618	ATP-binding cassette, subfamily (MDR/TAP), member 4 (ABCB4)	Y	Y	4/5	Y
1557996_at	AK091784	cDNA FLJ34465 fis, clone HLUNG2003061	N	N	5/5	Y
207631_at	NM_005821	Neighbour of BRCA1 gene (NBR2)	Y	Y	0/5	Y
208817_at	NM_000754/ NM_007310	Catechol-O-methyltransferase (COMT)	Y	Y	2/6	Y
214614_at	NM_005515	Homeo box HB9 (HLXB9)	N	Y	1/8	N
224449_at	BC006011	DNA- damage inducible protein 2 (DDI2)	N	N	2/6	N
218698_at	NM_015957	APAF1 interacting protein (APIP)	Y	N	0/8	N

Table 3.8.11.1 Summary of gene transcripts selected for validation by RT-PCR and the results obtained. The microarray trend changes validated by RT-PCR are shaded in grey. N= No, Y= Yes, Affy= Affymetrix

3.8.11.1 No reverse transcriptase controls

All of the primers designed for the RT-PCR work to validate the microarray results were designed to be specific for mRNA transcripts. To do this, one of the primers in the primer pair was designed to cross an exon-intron boundary (see Section 2.11.1 for example). For this reason the controls included for the RT-PCR study were water controls, i.e. water instead of RNA in the RT part and water instead of cDNA in the PCR part. These controls helped to ensure that no components (reagents/ tips/ tubes) were contaminated with either cDNA or PCR product carried over from previous experiments.

After the validation study was carried out PCR was carried out using RNA from some of the microarray samples to confirm that the results obtained were not due to amplification of genomic DNA sequences. PCR products were found in the (-) reverse transcription samples for 10/11 primer sets designed for the study (see Figs. 3.8.11.1.2/3 and 5). This was a surprising result because the primers pairs used in the RT-PCR study in Section 3.4, which were designed exactly the same way as those used in the microarray validation study did not show amplified product in the (-) reverse transcriptase samples. This indicates that the validation RT-PCR results shown in Section 3.8.11 may have been influenced by the amplification of genomic DNA sequences.

An added control was tested for some of these samples also, where all components were added except Oligo dT. The results shown in Figs. 3.8.11.1.9 and 3.8.11.1.10 (B) however, show that these samples were positive for PCR product whilst the (-) reverse transcriptase samples were negative. The primers for GSTπ mRNA align to two different exons, and therefore amplify a larger band if genomic DNA is present, however the PCR product was the same size as the mRNA amplicon. This indicates that either the mRNA is somehow

capable of self- priming but most likely that one of my RT components is contaminated with Oligo dT.

The primer design for two of the microarray validation primer sets are shown in Figs 3.8.11.1.1 and 3.8.11.1.3. The primer design for two of the PCR primer pairs used in the RT-PCR study from Section 3.4 are shown in Figs. 3.8.11.1.6 and 3.8.11.1.8 for comparison. The latter primer pairs do not amplify genomic DNA in comparison to the microarray primers that do. There are no distinctly obvious differences between the designs of the primers. Another possible explanation is that the gene transcripts chosen for validation of the microarray results have unknown intron-less pseudogenes, which would be amplified if genomic DNA was present and yield a PCR product of exactly the same size as expected with RNA.

Clone IMAGE: 5528716 mRNA, 152bp

Figure 3.8.11.1.1 Multalin results for mRNA sequence (black letters) with Affy target region (red letters). Forward and reverse RT-PCR primers are designated by the continuous blue line, dashed blue line is the amplified sequence and the region encircled (red) is the exon-exon junction site.

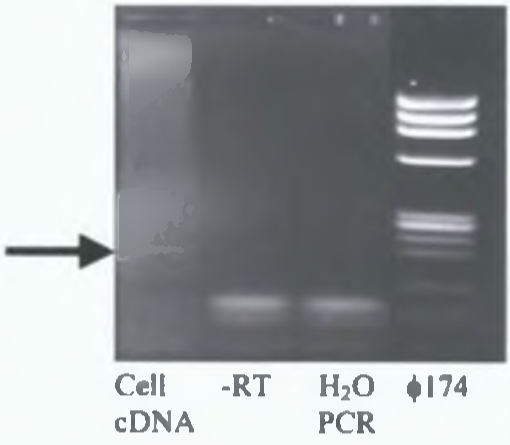


Figure 3.8.11.1.2 RT-PCR results for Clone IMAGE: 5528716 mRNA, 152bp. RPMI TAX cell cDNA was used for the cDNA sample, -RT is the RPMI TAX CM sample that was processed using PCR with no reverse transcription step. UHP was used in the H_2O PCR sample instead of cDNA as a control.

ABCB4 mRNA, 239bp

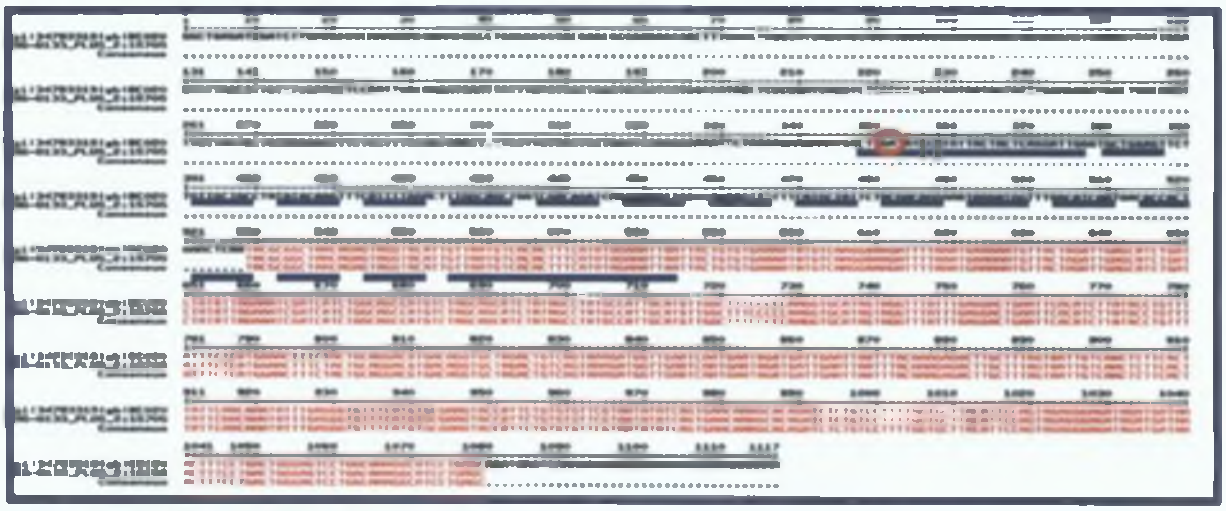


Figure 3.8.11.1.3 Multalin results for mRNA sequence (black letters) with Affy target region (red letters). Forward and reverse RT-PCR primers are designated by the continuous blue line, dashed blue line is the amplified sequence and the region encircled (red) is the exon-exon junction site.

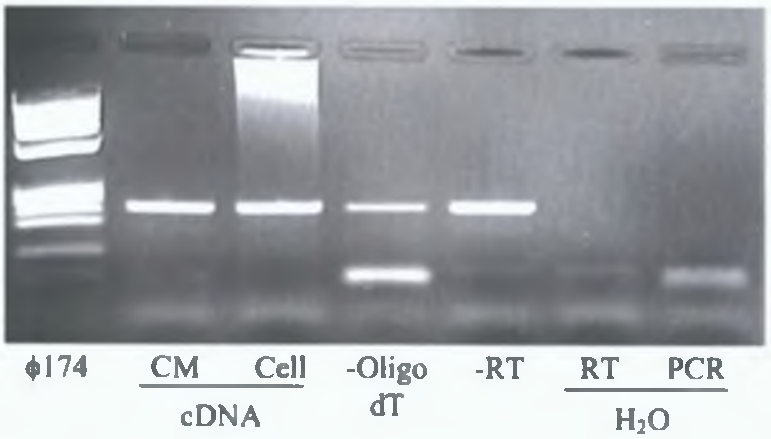


Figure 3.8.11.1.4 RT-PCR results for ABCB4 mRNA, 239bp. RPMI TAX CM and cell cDNA were used for the cDNA samples, -Oligo dT is were everything except oligo dT was added to the samples, -RT is the RPMI TAX CM sample that was processed using PCR with no reverse transcription step. UHP was used in the H_2O RT and PCR samples instead of RNA and cDNA as controls respectively.

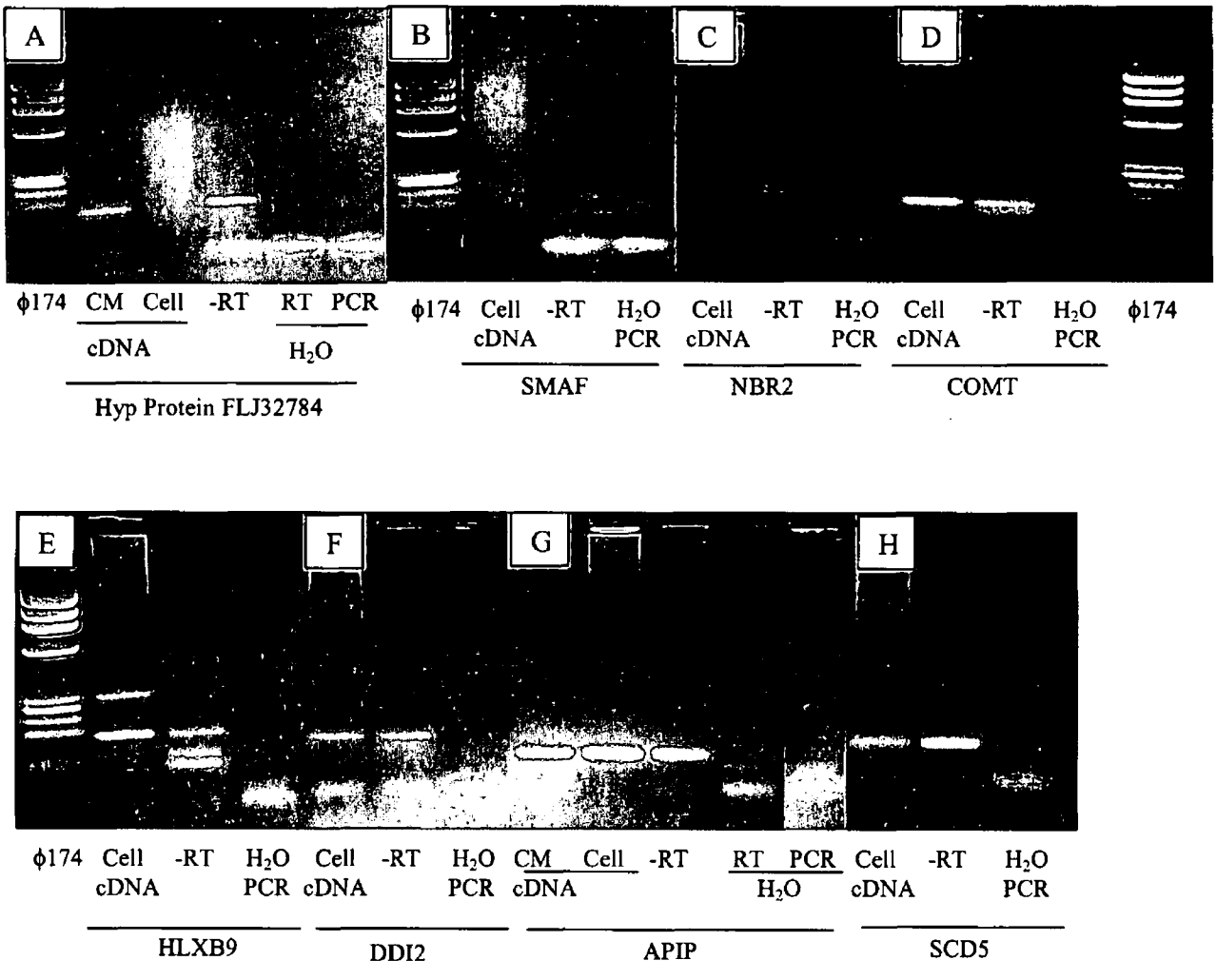


Figure 3.8.11.1.5 RT-PCR results for primers targetted to (A) Hyp Protein FLJ32784 mRNA, (B) SMAF mRNA, (C) NBR2 mRNA, (D) COMT mRNA, (E) HLXB9 mRNA, (F) DDI2 mRNA, (G) APIP mRNA and (H) SCD5 mRNA. RPMI TAX CM and cell cDNA were used for the cDNA samples, everything except oligo dT was added to the -Oligo dT samples, -RT is the RPMI TAX CM sample that was processed using PCR with no reverse transcription step. UHP was used in the H_2O RT and PCR samples instead of RNA and cDNA as controls respectively.

β-actin 171bp

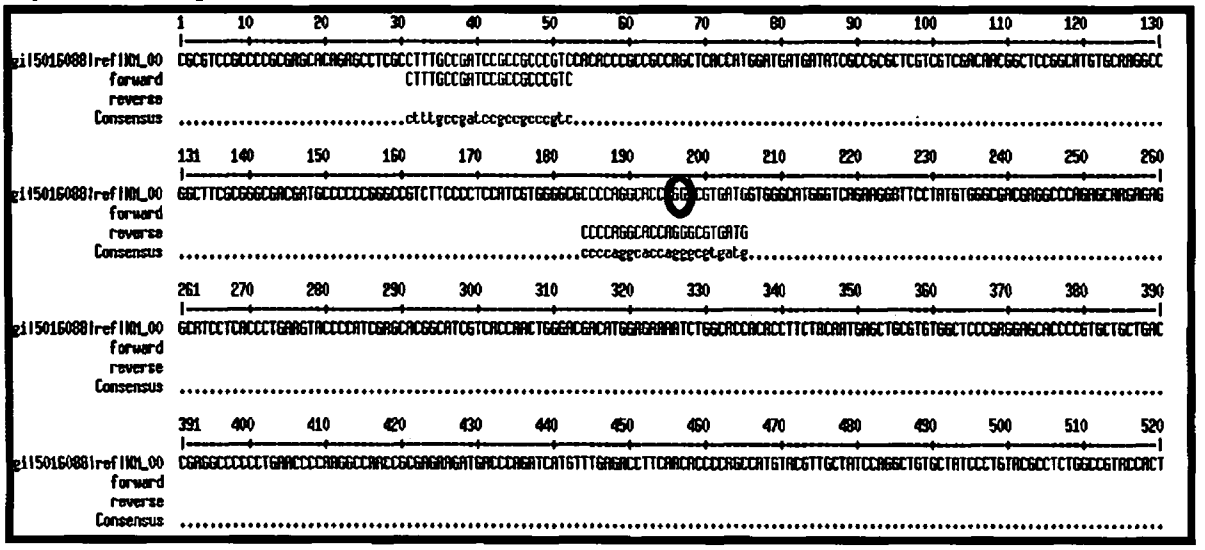


Figure 3.8.11.1.6 Multalin results for mRNA sequence (black letters) with primers (blue letters) and the region encircled (red) is the exon-exon junction site.

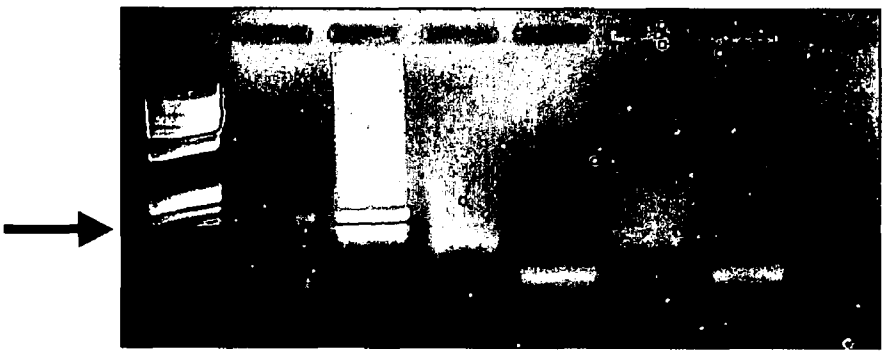


Figure 3.8.11.1.7 RT-PCR results for β-actin mRNA, 171bp. RPMI TAX CM and cell cDNA were used for the cDNA samples, everything except oligo dT was added to the -Oligo dT samples, -RT is the RPMI TAX CM sample that was processed using PCR with no reverse transcription step. UHP was used in the H_2O RT and PCR samples instead of RNA and cDNA as controls respectively.

HnRNPB1 155bp

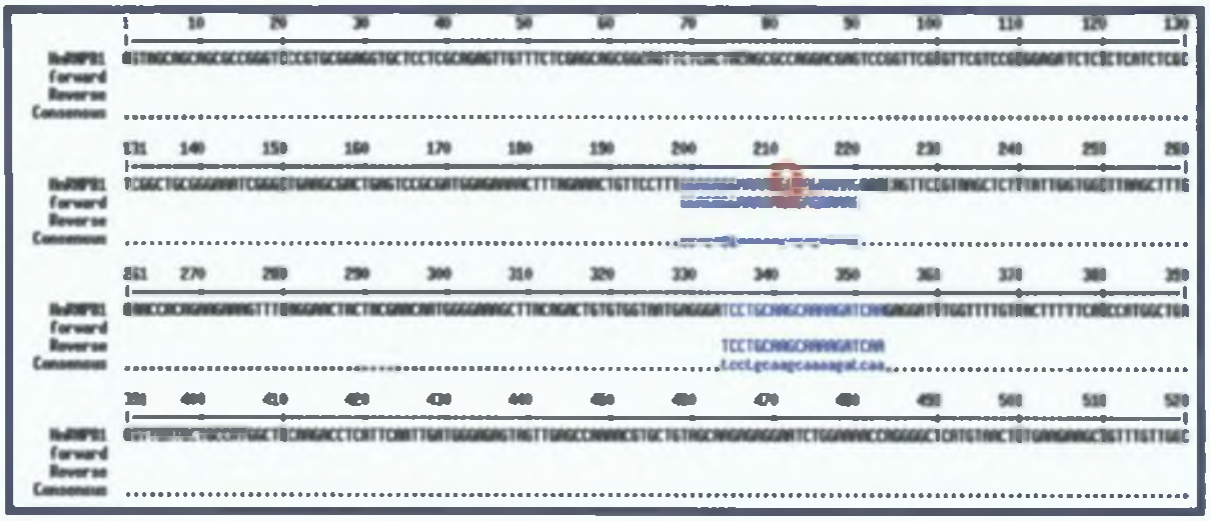


Figure 3.8.11.1.8 Multalin results for mRNA sequence (black letters) with primers (blue letters) and the region encircled (red) is the exon-exon junction site.

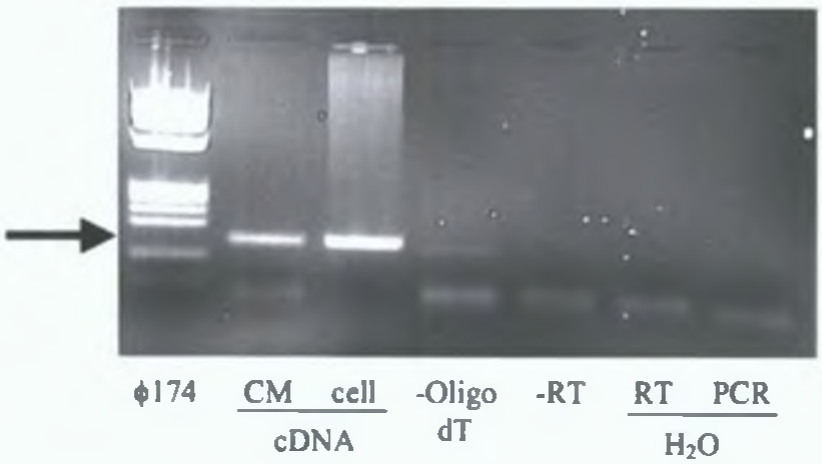


Figure 3.8.11.1.7 RT-PCR results for β-actin mRNA, 171bp. RPMI TAX CM and cell cDNA were used for the cDNA samples, everything except oligo dT was added to the -Oligo dT samples, -RT is the RPMI TAX CM sample that was processed using PCR with no reverse transcription step. UHP was used in the H_2O RT and PCR samples instead of RNA and cDNA as controls respectively.

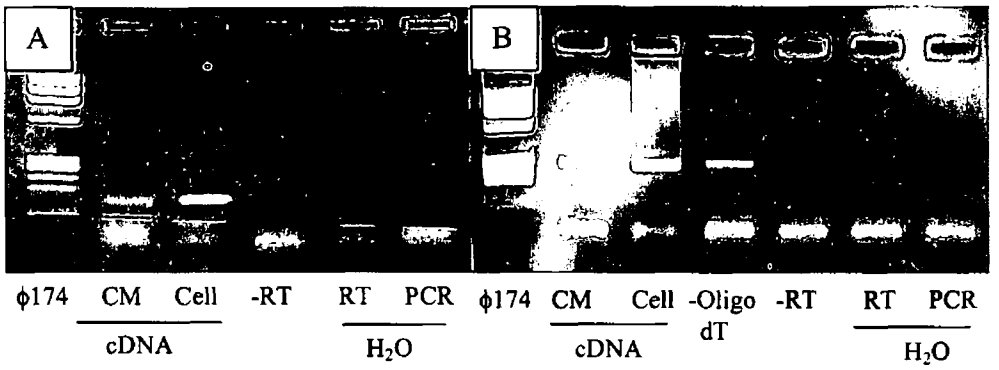


Figure 3.8.11.1.8 RT-PCR results for (A) MDR1, (B) GST π and (C) Bcl2α. RPMI TAX CM and cell cDNA were used for the cDNA samples, everything except oligo dT was added to the -Oligo dT samples, -RT is the RPMI TAX CM sample that was processed using PCR with no reverse transcription step. UHP was used in the H_2O RT and PCR samples instead of RNA and cDNA as controls respectively.

3.8.12 CM mRNA vs. CELL mRNA

The cells, from which CM was taken for the microarray study, were washed, trypsinised and RNA was extracted by other researchers (H1299 and H1299 TAX by Laura Breen, MDA-F, MDA-F-ADR and MDA-F-TAX by Helena Joyce and RPMI, RPMI ML and RPMI TAX by Isabella Bray). The RNA was extracted at the same time that RNA was extracted from the CM taken from the samples. The cell RNA was also amplified and labelled using a 1 cycle amplification kit and hybridised to Affymetrix U133A Plus 2.0 chips by the researchers mentioned. As the microarray data generated directly related to the microarray data generated with the CM samples the researchers kindly gave access to the data so that cell /CM profiles could be compared.

All of the raw data was normalised exactly as the CM data and imported into Genespring. Venn diagrams were then plotted to ascertain the number of probe sets called Present or Marginal in the CM groups and also in the Cell groups.

The cell RNA microarray data was also assessed to see if probe sets significantly up/ down-regulated in the parental against their drug selected variant CM samples, showed similar expression patterns. In order to allow larger numbers of probe sets to be included in the study, the only criteria specified was that probe sets differentially expressed between the sample groups with a p-value less than 0.05 were included, fold change difference in expression and Presence/ Absence call were not included.

H1299 CM vs. Cell

Fig. 3.8.12.1 shows that 57% of the gene transcripts called P or M in all 3 H1299 CM samples are also called P or M in all three H1299 cell RNA samples. If the gene transcripts with an expression value of over 100 in the 3 H1299 CM samples are overlapped with the gene transcripts called Present in the 3 H1299 cell RNA samples, the percentage of gene

transcripts common to both samples drops to 50%. However if both of these criterion i.e. Present call in 3/3 CM samples with an expression value over 100 are applied and the number of gene transcripts passing this criteria (371) are overlapped with the gene transcripts called Present in 3/3 cell RNA samples the percentage of gene transcripts common to both samples rises to 80%. On closer inspection of the 76 gene transcripts (20%) still not common to both the cell and CM RNA samples, only 23 of these were called absent in all three cell RNA samples.

To summarise, 80% of the gene transcripts called Present in the 3 H1299 CM samples with an expression value over 100 were called also called Present in each of the 3 H1299 cell RNA samples and 6% were called Absent in the 3 H1299 cell RNA samples. The other 14% related to probe sets that were not uniformly called Present/ Absent or Marginal in 3/3 cell RNA samples e.g. P in 1 or 2 of the cell RNA samples and A in the remaining sample. This means that the poor correlation of the probe sets called Present in both the CM and cell samples is attributable to the noise in the CM samples possibly due to the 2 cycle amplification procedure or maybe a reflection of the RNA extractable from the CM samples.

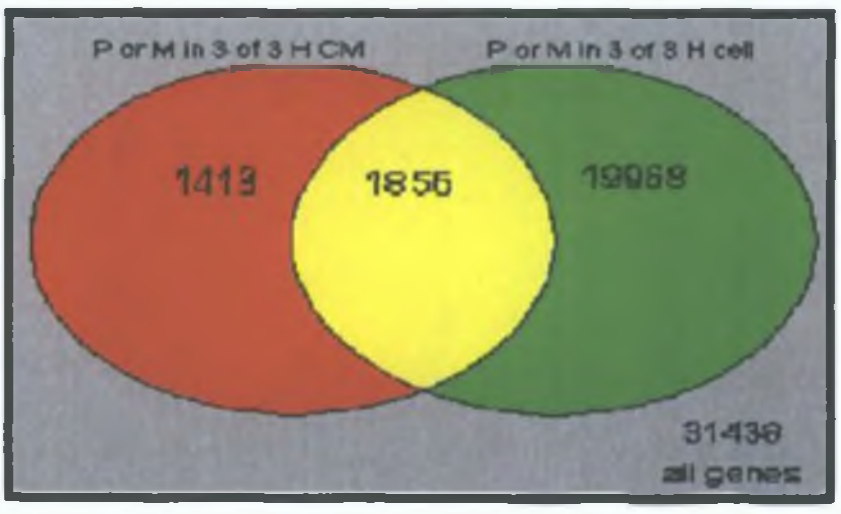


Figure 3.8.12.1 Probe sets called P or M in all H1299 CM and cell samples. The number of probe sets assigned P or M in the CM and cell RNA samples are overlapped. H= H1299.

H1299 TAX CM vs. Cell

Fig. 3.8.12.2 shows that 65% of the gene transcripts called P or M in all 3 H1299 TAX CM samples are also called P or M in all three H1299 TAX cell RNA samples. If the gene transcripts with an expression value over 100 in all 3 CM samples (1,199) were overlapped with the cell RNA samples, the number of gene transcripts common to both would drop to 50% (610). However if both of these criteria are applied to the CM samples and they are overlapped with the gene transcripts called P in all 3 cell RNA samples, the number of overlapping gene transcripts rises to 79% (394 probe sets called P with an expression value over 100 in all 3 H1299 TAX CM samples, of which 312 are called P in the 3 H1299 TAX cell RNA samples. Only 5% (20) of the gene transcripts called P with an expression value over 100 in the 3 H1299 TAX CM samples are called A in all 3 H1299 TAX cell RNA samples. The remaining 16% is attributable to probe sets with a mixture of P/A/M calls between the replicate H1299 TAX cell RNA samples e.g. P in 1 or 2 of the cell RNA samples and A in the remaining sample.

Figure 3.8.12.2 Venn diagram of probe sets called P or M in all H1299 TAX CM and Cell samples. HT = H1299 TAX.

RPMI CM vs. Cell

Fig. 3.8.12.3 shows that 54% of the gene transcripts called P or M in both RPMI CM samples are also called P or M in all three RPMI cell RNA samples. The percentage of gene transcripts with an expression value over 100 in all 3 RPMI CM samples also called P in all 3 RPMI cell RNA samples is 43%. The gene transcripts common to the cell and CM RNA samples rises to 74% however when the combination criteria of a P call with an expression value over 100 is applied to the CM samples and they are overlapped with the gene transcripts called P in all 3 RPMI cell RNA samples. Of the remaining 26%, 12% (39) are gene transcripts that are called P in all 3 CM samples, with an expression value over 100 and called A in all 3 cell RNA samples. The remaining 14% is attributable to probe sets with a mixture of P/A/M calls within the 3 RPMI cell RNA samples e.g. P in 1 or 2 of the cell RNA samples and A in the remaining sample.

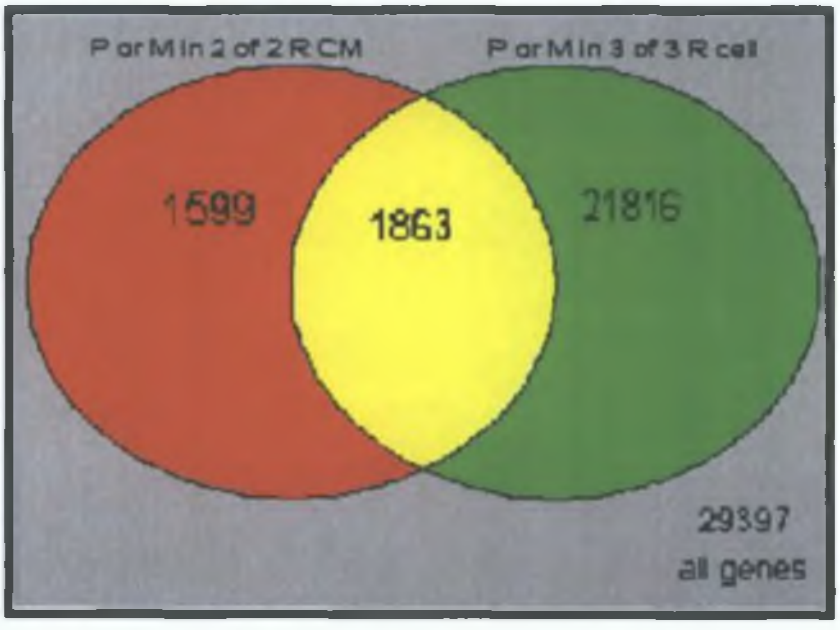


Figure 3.8.12.3 Venn diagram of probe sets called P or M in all RPMI CM and Cell samples. R= RPMI.

RPMI ML CM vs. Cell

Fig. 3.8.12.4 shows that 52% of the gene transcripts called P or M in all 3 RPMI ML CM samples are also called P or M in all three RPMI cell RNA samples. Only 41% of gene transcripts with an expression value over 100 in all 3 RPMI CM samples are also called P in all 3 RPMI ML cell RNA samples. When only gene transcripts with a P call and an expression value over 100 in all 3 of the RPMI ML samples are overlapped with the RPMI ML cell RNA samples, the percentage of gene transcripts common to both is 69%. Of the remaining 31% of gene transcripts not common to both sample types, 9% (29 gene transcripts) are called P in the 3 CM samples with an expression value over 100 and A in all 3 of the RPMI ML cell RNA samples. 22 % of the gene transcripts called P with an expression value over 100 all 3 CM samples had different calls between the RPMI ML cell RNA samples e.g. P in 1 or 2 of the cell RNA samples and A in the remaining sample.

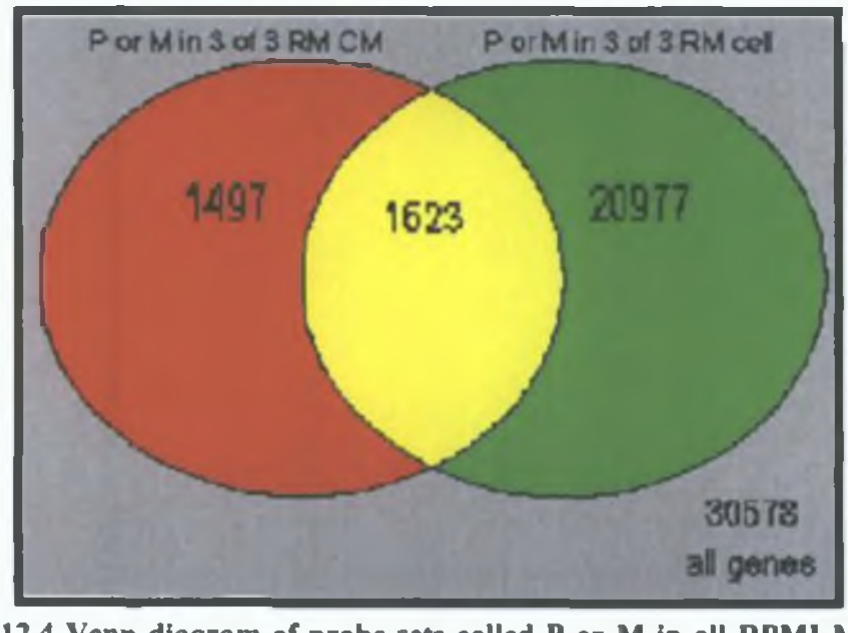


Figure 3.8.12.4 Venn diagram of probe sets called P or M in all RPMI ML CM and Cell samples. RM= RPMI ML.

RPMI TAX CM vs. Cell

Fig. 3.8.12.5 shows that 58% of the gene transcripts called P or M in all 3 RPMI TAX CM samples are also called P or M in all three RPMI TAX cell RNA samples. The percentage of gene transcripts with an expression value over 100 in the RPMI TAX CM samples that were called P in the RPMI TAX cell RNA samples was 47%. However 75% of gene transcripts were common to both CM and cell RNA samples when the gene transcripts called P in all 3 RPMI TAX cell RNA samples were overlapped with the gene transcripts called P with an expression value over 100 in all 3 RPMI TAX CM samples. 10% (39) of the gene transcripts called P with an expression value over 100 in all 3 RPMI TAX CM samples are called A in all 3 RPMI TAX cell RNA samples. The remaining 15% of gene transcripts gene transcripts called P with an expression value over 100 in all 3 RPMI TAX CM samples had different P/A/M calls between the 3 RPMI TAX cell RNA samples e.g. P in 1 or 2 of the cell RNA samples and A in the remaining sample.

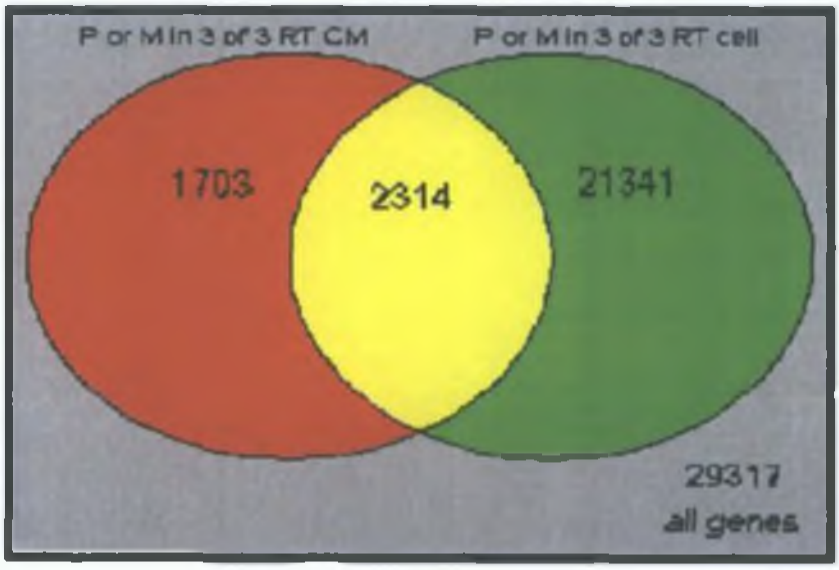


Figure 3.8.12.5 Venn diagram of probe sets called P or M in all RPMI TAX CM and Cell samples. RT= RPMI TAX.

MDA-F CM vs. Cell

Fig. 3.8.12.6 shows that 57% of the gene transcripts called P or M in all 3 MDA-F CM samples are also called P or M in all three MDA-F cell RNA samples. The percentage of gene transcripts with an expression value over 100 in the MDA-F CM samples that were called P in the MDA-F cell RNA samples was 47%. However 77% of gene transcripts were common to both CM and cell RNA samples when the gene transcripts called P in all 3 MDA-F cell RNA samples were overlapped with the gene transcripts called P with an expression value over 100 in all 3 MDA-F CM samples. 6% (20) of the gene transcripts called P with an expression value over 100 in all 3 MDA-F CM samples are called A in all 3 MDA-F cell RNA samples. The remaining 17% of gene transcripts called P with an expression value over 100 in all 3 MDA-F CM samples had different P/A/M calls between the 3 MDA-F cell RNA samples e.g. P in 1 or 2 of the cell RNA samples and A in the remaining sample.

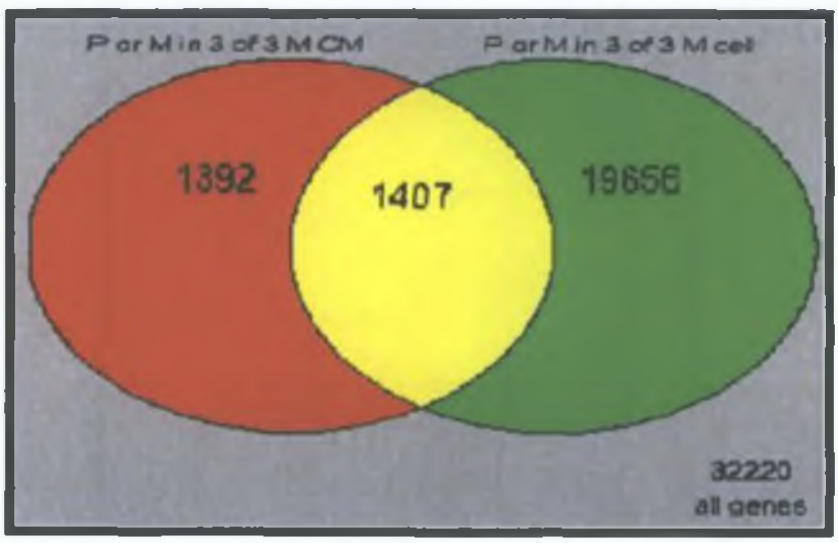


Figure 3.8.12.6 Venn diagram of probe sets called P or M in all MDA-F CM and Cell samples. M= MDA-F

MDA-F-ADR-SI CM vs. Cell

Fig. 3.8.12.7 shows that 62% of the gene transcripts called P or M in all 3 MDA-F-ADR-SI CM samples are also called P or M in all three MDA-F-ADR-SI cell RNA samples. The percentage of gene transcripts with an expression value over 100 in the MDA-F-ADR-SI CM samples that were called P in the MDA-F-ADR-SI cell RNA samples was 49%. However 75% of gene transcripts were common to both CM and cell RNA samples when the gene transcripts called P in all 3 MDA-F-ADR-SI cell RNA samples were overlapped with the gene transcripts called P with an expression value over 100 in all 3 MDA-F-ADR-SI CM samples. 7% (33) of the gene transcripts called P with an expression value over 100 in all 3 MDA-F-ADR-SI CM samples are called A in all 3 MDA-F-ADR-SI cell RNA samples. The remaining 18% of gene transcripts called P with an expression value over 100 in all 3 MDA-F-ADR-SI CM samples had different P/A/M calls between the 3 MDA-F-ADR-SI cell RNA samples e.g. P in 1 or 2 of the cell RNA samples and A in the remaining sample.

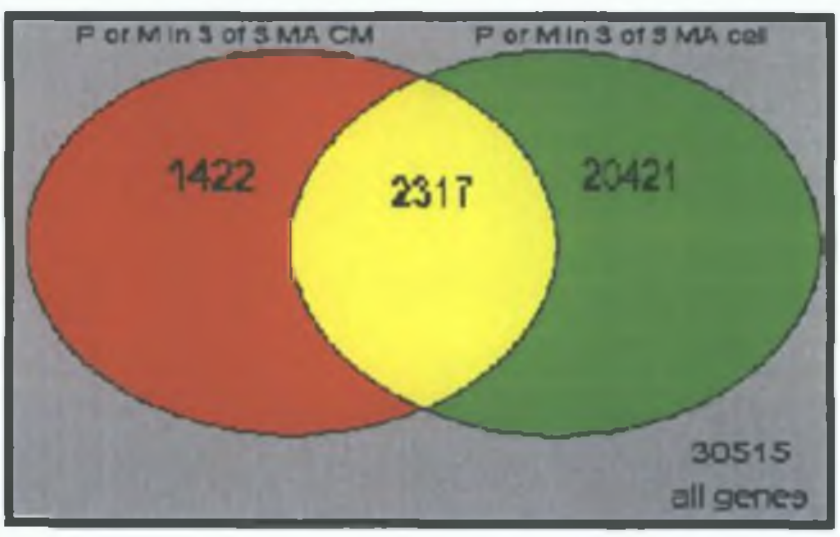


Figure 3.8.12.7 Venn diagram of probe sets called P or M in all MDA-F-ADR-SI CM and Cell samples. MA= MDA-F-ADR-SI

MDA-F- TAX-SI CELL vs. CM

Fig. 3.8.12.8 shows that 49% of the gene transcripts called P or M both MDA-F-TAX-SI CM samples are also called P or M in all three MDA-F-TAX-SI cell RNA samples. The percentage of gene transcripts with an expression value over 100 in the MDA-F-TAX-SI CM samples that were called P in the MDA-F-TAX-SI cell RNA samples was 40%. However 62% of gene transcripts were common to both CM and cell RNA samples when the gene transcripts called P in all 3 MDA-F-TAX-SI cell RNA samples were overlapped with the gene transcripts called P with an expression value over 100 in both MDA-F-TAX-SI CM samples. 10% (41) of the gene transcripts called P with an expression value over 100 in both MDA-F -TAX-SI CM samples are called A in all 3 MDA-F-TAX-SI cell RNA samples. The remaining 28% of gene transcripts gene transcripts called P with an expression value over 100 in both MDA-F-TAX-SI CM samples had different P/A/M calls between the 3 MDA-F-TAX-SI cell RNA samples e.g. P in 1 or 2 of the cell RNA samples and A in the remaining sample.

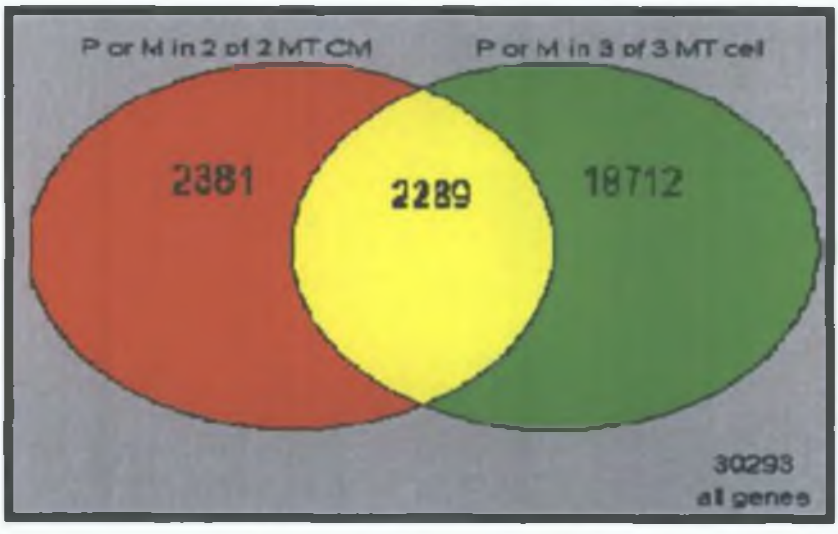


Figure 3.8.12.8 Venn diagram of probe sets called P or M in all MDA-F-TAX-SI CM and Cell samples. MT= MDA-F-TAX-SI.

Table 3.8.12.1 summarises the results in this Section. It highlights the fact that the correlation between the gene transcripts common to both the cell and CM samples is much higher when a criteria of expression over 100 is applied to the CM samples. This indicates that the data for the CM samples with an expression under 100 may be attributable to noise possibly due to the 2 cycle amplification procedure.

	Criteria for all CM samples	Criteria for all cell samples	CM	Overlapping 3/3 cell samples
H1299	P or M call	P or M call	3269	1858 (57%)
	Exp >100	P call	1260	610 (48%)
	P call + exp >100	P call	371	312 (84%)
H1299 TAX	P or M call	P or M call	4135	2690 (65%)
	Exp >100	P call	1199	610 (51%)
	P call + exp >100	P call	394	312 (79%)
RPMI	P or M call	P or M call	3462	1863 (54%)
	Exp >100	P call	1208	516 (43%)
	P call + exp >100	P call	335	249 (74%)
RPMI ML	P or M call	P or M call	3120	1623 (52%)
	Exp >100	P call	1208	516 (43%)
	P call + exp >100	P call	335	249 (74%)
RPMI TAX	P or M call	P or M call	3017	2314 (58%)
	Exp >100	P call	1150	543 (47%)
	P call + exp >100	P call	375	281 (75%)
MDA-F	P or M call	P or M call	2799	1407 (50%)
	Exp >100	P call	1155	543 (47%)
	P call + exp >100	P call	350	270 (77%)
MDA-F-ADR-SI	P or M call	P or M call	3793	2317 (62%)
	Exp >100	P call	1261	612 (49%)
	P call + exp >100	P call	456	344 (75%)
MDA-F-TAX-SI	P or M call	P or M call	4670	2289 (49%)
	Exp >100	P call	1232	496 (40%)
	P call + exp >100	P call	390	241 (62%)

Table 3.8.12.1 A summary of the number of probe sets common in both CM and cell RNA microarray results with different criteria for the probe sets. P or M = Present or Marginal, Exp> 100 = expression over 100.

3.8.13 Probe sets significantly changed in CM and Cell groups

H1299 vs. H1299TAX CM and Cells

Table 3.8.13.1 indicates that probe sets corresponding to four gene transcripts are up-regulated in both H1299 CM and cell RNA in relation to the H1299 CM and cell RNA samples. Two of these, adenylate kinase 3 and ataxin 7-like 1, also showed both a fold change greater than 1.2 and expression difference greater than 50 in the H1299 vs. H1299 CM samples.

Probe sets corresponding to five gene transcripts are down-regulated in both H1299 TAX CM and cell RNA compared to H1299 CM and cell RNA. The remaining four gene transcripts that were deemed to be significantly changed were differentially regulated in both the CM and cell RNA samples, for example nucleosome assembly protein 1-like 2 (219368_at) was down-regulated in H1299 TAX CM by 3.8 fold in comparison to H1299 CM but was 2.5 fold up-regulated in H1299 TAX cell RNA compared to H1299 cell RNA.

			CM			CELLS		
probe set	gene	Accession	fold change	diff of means	P value	fold change	diff of means	P value
238670_at	CDNA FLJ41972 fis, clone SKNMC2003987	AI144156	10.24	14.97	0.047459	1.92	50	0.015726
204347_at	adenylate kinase 3 /// adenylate kinase 3	AI653169	4.32	104.05	0.001076	1.27	133.97	0.027701
214257_s_at	gb:AA890010 /DB_XREF=gi:3016889 /DB_XI	AA890010	2.25	46.98	0.039985	1.34	438.07	0.029288
227732_at	ataxin 7-like 1	AB033044	2.23	83.79	0.009627	1.75	48.23	0.021087
224908_s_at	tubulin tyrosine ligase	AI888594	-8.76	-25.67	0.02224	2.25	212.78	0.015902
233461_x_at	zinc finger protein 226	AK023091	-4.38	-27.06	0.013664	-2.33	-222.23	0.036992
211810_s_at	galactosylceramidase (Krabbe disease)	D25284	-1.91	-16.13	0.020374	4.92	31.43	0.013773
235940_at	chromosome 9 open reading frame 64	AW983691	-1.64	-34.72	0.046958	19.27	281.17	0.031986
223989_s_at	small fragment nuclease	BC003502	-1.27	-25.07	0.026115	1.65	154.35	0.007588
224472_x_at	calcium binding protein Cab45 precursor /// c	BC006211	4.34	14.35	0.033531	-1.23	-99.54	0.030905
219368_at	nucleosome assembly protein 1-like 2	NM_021963	-3.78	-13.48	0.034204	2.47	60.58	0.021605
213532_at	a disintegrin and metalloproteinase domain 1	AI797833	-3.39	-16.99	0.040954	1.60	222.34	0.030285
203846_at	tripartite motif-containing 32	BC003154	-1.45	-31.89	0.048034	1.27	56.94	0.035028

Table 3.8.13.1 Probe sets significantly changed in both CM and Cell RNA H1299 vs. H1299 TAX groups. The replicate samples in each CM sample group were averaged and compared to their averaged drug selected CM variant samples. Gene transcripts which had a P value <0.05 were included on the list regardless of fold change or

difference in expression. The same steps were carried out with the data for the cell line RNA. The two lists were then cross compared to generate a list of gene transcripts that were called significantly changed (by criteria of P value <0.05 alone) in both CM and cell RNA samples. Gene transcripts shaded light grey are up- regulated in both samples, dark grey indicates gene transcripts down- regulated in both groups and no shading indicates gene transcripts that were not uniformly up/ down-regulated in both samples. The gene transcripts within each category are sorted based on fold change in the CM samples. Gene transcripts highlighted in bold also passed the criteria of difference in fold change> 1.2, difference of means> 50 in the CM samples and are also listed elsewhere in section 3.8.8.

RPMI vs. RPMI ML CM and Cells

Table 3.8.13.2 indicates that probe sets corresponding to 19 gene transcripts were similarly up-regulated in the RPMI ML CM and cell RNA samples compared to the RPMI CM and cell RNA samples. A further seven gene transcripts were down-regulated in both RPMI ML CM and cell samples compared to the RPMI CM and cell samples. 10 of the 26 gene transcripts that showed similar trends of regulation of expression in the CM and cell RNA samples were also further classified as significant in the CM samples based on the fact that they showed a fold change greater than 1.2 with an expression difference greater than 50 between the RPMI and RPMI CM samples.

The majority of the gene transcripts determined to be significantly differentially expressed in the RPMI ML CM and cell RNA samples compared to the RPMI CM and cell samples did not show the same trend of change in expression in the CM and cell RNA samples. 15 of the 46 gene transcripts showing different but significant trends of expression change between the sample groups, were also further classified as significant in reference to the CM samples because both the fold change and difference in expression were greater than 1.2 and 50 respectively in the RPMI CM group compared to the RPMI ML CM group. An example is manosidase, alpha, class 2A, member 1 (226538_at), this gene is down-regulated 3.7 fold with a difference in expression of 80 in the RPMI ML CM samples compared to the RPMI CM samples, however in the cell RNA samples it is up-regulated 7 fold in the RPMI ML cell samples compared to the RPMI cell samples.

			CM			CELLS		
probe set	gene	Accession	fold change	diff of means	P value	fold change	diff of means	P value
200991_s_at	sorting nexin 17	NM_014748	24.27	23.27	0.034216	1.38	291.82	0.033211
223732_at	solute carrier family 23 (nucleobase transport	AF170911	17.02	19.7	0.045717	22.72	97.1	0.005526
244026_at	gb:BF063657 /DB_XREF=gi:10822567 /DB_X	BF063657	16.98	17.26	0.028985	3.77	22.72	0.007817
222266_at	**gb:BF796940 /DB_XREF=gi:12101994 /DB_**	**BF796940**	**14.85**	**66.62**	**0.005071**	**1.3**	**81.57**	**0.037723**
1552390_a_a	hypothetical protein FLJ39553	NM_173549	14.12	13.12	0.028815	7.03	61.41	0.000545
211945_s_at	integrin, beta 1 (fibronectin receptor, beta pol	BG500301	11.24	45.66	0.023639	12.46	5126.4	0.000539
204755_x_at	hepatic leukemia factor	M95585	8.03	38.93	0.013058	5.18	22.06	0.011187
212195_at	**interleukin 6 signal transducer (gp130, on**	**AL049265**	**5.67**	**59.29**	**0.040941**	**3.06**	**876.22**	**0.001289**
1554512_a_a	hypothetical protein FLJ14640	BC032307	5.1	39.97	0.026487	2.98	60.54	0.004958
241904_at	**gb:BE068893 /DB_XREF=gi:8413543 /DB_X**	**BE068893**	**4.41**	**58.98**	**0.045711**	**5.43**	**52.64**	**0.004657**
229871_at	**hypothetical protein FLJ10211**	**W74622**	**3.88**	**153.62**	**0.00284**	**2.04**	**50.84**	**0.015523**
213931_at	**inhibitor of DNA binding 2, dominant neg**	**AI819238**	**3.03**	**82.59**	**0.012924**	**24.87**	**33.42**	**0.037183**
1553162_x_a	hypothetical protein BC017947	NM_144692	3.01	36	0.044058	1.51	65.84	0.016061
232058_at	Actinin, alpha 4	AU158358	2.93	47.7	0.020915	3.48	37.97	0.018973
224617_at	ROD1 regulator of differentiation 1 (S. pombe	AI735576	2.9	25.92	0.043128	6.03	1640.07	0.002189
200833_s_at	RAP1B, member of RAS oncogene family	NM_015646	2.5	44.07	0.040241	1.5	755.81	0.008097
213005_s_at	ankyrin repeat domain 15	D79994	2.4	29.34	0.048473	2.74	184.8	0.003042
227952_at	**Zinc finger protein 595**	**AI580142**	**2.26**	**302.8**	**0.044858**	**10.75**	**134.8**	**0.001031**
212854_x_at	**AG1**	**AB051480**	**1.74**	**57.58**	**0.043794**	**1.36**	**178.25**	**0.04228**
1555517_at	Transcribed locus, highly similar to NP_1500	BC045709	-8.9	-22.75	0.027167	-93.24	-101.46	0.001712
210236_at	protein tyrosine phosphatase, receptor type,	U22815	-5.63	-12.46	0.038176	-1.36	-17.77	0.033831
220808_at	**Theg homolog (mouse)**	**NM_016585**	**-4.75**	**-119.27**	**0.004057**	**-24.29**	**-55.2**	**0.024016**
218698_at	**likely ortholog of mouse monocyte macr**	**NM_015957**	**-3.47**	**-141.77**	**0.043591**	**-1.22**	**-112.97**	**0.020009**
225276_at	**G1 to S phase transition 1**	**AA143579**	**-1.85**	**-50.93**	**0.048889**	**-3.32**	**-445.68**	**0.001061**
223367_at	Williams Beuren syndrome chromosome regi	BC005056	-1.34	-25.66	0.034059	-1.3	-52.86	0.045176
212113_at	Similar to microtubule-associated proteins 1A	AI927479	-1.32	-27.79	0.015358	-4.05	-134.49	0.021727
224609_at	CTL2 gene	AI264216	26.75	25.75	0.045236	-3.12	-60.57	0.003005
204675_at	steroid-5-alpha-reductase, alpha polypeptide	NM_001047	21.65	20.86	0.047673	-3.26	-434.04	0.000102
225957_at	adult retina protein	AI307750	16.92	25.63	0.046647	-1.85	-28.62	0.049982
235048_at	KIAA0888 protein	AV720650	14.08	20.79	0.030609	-2.16	-329.98	0.000295
233019_at	CCR4-NOT transcription complex, subunit 7	AU145061	13.88	38	0.043313	-3.24	-43.92	0.015836
205867_at	protein tyrosine phosphatase, non-receptor ty	NM_002834	13.25	12.25	0.033844	-1.71	-42.01	0.011227
232194_at	methyltransferase like 4	AA764787	11.64	12.58	0.044119	-2.43	-69.59	0.001529
236520_at	gb:AW972380 /DB_XREF=gi:8162226 /DB_X	AW972380	11.03	41.73	0.014949	-1.81	-20.51	0.041727
204347_at	**adenylate kinase 3 /// adenylate kinase 3**	**AI653169**	**11.01**	**75.31**	**0.00913**	**-7.45**	**-577.7**	**0.019387**
243801_x_at	gb:AA971709 /DB_XREF=gi:3146999 /DB_XF	AA971709	10.35	17.21	0.036532	-1.46	-80.93	0.030205
218464_s_at	hypothetical protein FLJ10700	NM_018182	10.29	54.7	0.018546	-1.34	-63.69	0.006886
226846_at	Hypothetical protein FLJ20171	BF001941	10.13	17.26	0.043196	-116.18	-342.09	0.000315
230479_at	gb:AI872374 /DB_XREF=gi:5546423 /DB_XR	AI872374	7.68	27.07	0.029527	-6.91	-186.47	0.009381
1557756_a_a	Chromosome 14 open reading frame 145	AW028337	7.02	19.74	0.046678	-2.61	-15.63	0.024204
226374_at	gb:BG260087 /DB_XREF=gi:12769903 /DB_X	BG260087	6.07	30.23	0.036605	-2.1	-327.07	0.016084
244007_at	gb:BE889301 /DB_XREF=gi:10346478 /DB_X	BE889301	3.63	33.15	0.023578	-4.93	-270.12	0.030517
238590_x_at	hypothetical protein MGC10744	BF981428	3.6	48.36	0.031261	-1.36	-138.56	0.028727
221580_s_at	hypothetical protein MGC5306	BC001972	3.58	28.74	0.030651	-1.54	-284.71	0.002007
231215_at	gb:AI733145 /DB_XREF=gi:5054258 /DB_XR	AI733145	3.36	21.97	0.047739	-6.77	-54.75	0.003056
1569181_x_a	**gb:BC017896.1 /DB_XREF=gi:22450843 /TI**	**BC017896**	**3.12**	**61.21**	**0.037658**	**-2.26**	**-31.42**	**0.033741**
208905_at	**cytochrome c, somatic**	**BC005299**	**3**	**52.9**	**0.043279**	**-1.41**	**-1626.58**	**0.000564**
206055_s_at	small nuclear ribonucleoprotein polypeptide A	NM_003090	2.97	25.58	0.048927	-1.23	-327.99	0.014491
242856_at	gb:AI291804 /DB_XREF=gi:3934578 /DB_XR	AI291804	2.94	24.84	0.03847	-28.31	-310.83	0.048618
238156_at	gb:AW205632 /DB_XREF=gi:6505106 /DB_X	AW205632	2.9	26.83	0.028741	-1.51	-54.54	0.016997
231101_at	gb:AI689803 /DB_XREF=gi:4901097 /DB_XR	AI689803	2.6	46.59	0.045497	-1.51	-146.26	0.024271
214323_s_at	**UPF3 regulator of nonsense transcripts h**	**N36842**	**2.36**	**132.87**	**0.0098**	**-2.66**	**-931.26**	**0.000066**
208668_x_at	**high-mobility group nucleosomal binding**	**BC003689**	**2.25**	**71.87**	**0.047983**	**-1.33**	**-1422.73**	**0.005078**
201658_at	ADP-ribosylation factor-like 1	AU151560	1.88	41.84	0.042816	-2.04	-508.34	0.002845
200817_x_at	**ribosomal protein S10**	**NM_001014**	**1.78**	**80.62**	**0.049138**	**-1.14**	**-1070.7**	**0.025345**
212153_at	pogo transposable element with ZNF domain	AB007930	1.44	31.14	0.015645	-2.2	-576.5	0.002132
210334_x_at	**baculoviral IAP repeat-containing 5 (surv**	**AB028869**	**1.34**	**76.11**	**0.010615**	**-1.21**	**-267.03**	**0.02108**
52005_at	**Widely-interspaced zinc finger motifs**	**AA422049**	**1.23**	**52.47**	**0.02343**	**-2.33**	**-108.84**	**0.001071**
233816_at	**Solute carrier family 8 (sodium/calcium e**	**AW668617**	**1.21**	**52.85**	**0.034609**	**-9.15**	**-28.12**	**0.029041**
207732_s_at	discs, large homolog 3 (neuroendocrine-dlg, [	NM_021120	1.13	122.1	0.022506	-3.1	-88.12	0.043847
201671_x_at	ubiquitin specific protease 14 (tRNA-guanine	BC003556	1.1	181.47	0.028099	-1.42	-245.9	0.013316
241223_x_at	gb:AI821721 /DB_XREF=gi:5440800 /DB_XR	AI821721	1.07	243.58	0.035408	-1.55	-69.88	0.022032
214614_at	**homeo box HB9**	**AI738662**	**-1.26**	**-176.96**	**0.033232**	**1.66**	**199.84**	**0.037287**
210201_x_at	**bridging integrator 1**	**AF001383**	**-1.27**	**-173.45**	**0.014539**	**13.91**	**206.2**	**0.00078**
208817_at	**catechol-O-methyltransferase**	**BC000419**	**-1.49**	**-338.75**	**0.010668**	**4.72**	**349.52**	**0.005855**
215482_s_at	eukaryotic translation initiation factor 2B, sut	AJ011307	-2.47	-20.68	0.044077	1.37	192.78	0.011284
206491_s_at	**N-ethylmaleimide-sensitive factor attachi**	**NM_003827**	**-2.58**	**-66.53**	**0.014097**	**1.6**	**186.95**	**0.019195**
226784_at	TWIST neighbor	AA121481	-2.71	-36.75	0.009507	1.62	124.71	0.002578
240304_s_at	**Transmembrane channel-like 5**	**BG484769**	**-3.32**	**-138.84**	**0.014492**	**80.46**	**229.63**	**0.005727**
226538_at	**mannosidase, alpha, class 2A, member 1**	**AV700323**	**-3.66**	**-80.41**	**0.015963**	**7.06**	**226.11**	**0.00222**
231274_s_at	gb:R92925 /DB_XREF=gi:965279 /DB_XREF	R92925	-5.09	-24.73	0.020705	1.52	61.2	0.019781
212419_at	chromosome 10 open reading frame 56	AA131324	-10.54	-30.05	0.011622	4.65	82.18	0.023895

Table 3.8.13.2 Probe sets significantly changed in both CM and Cell RNA RPMI vs. RPMI ML groups. Gene transcripts shaded light grey are up- regulated in both samples, dark grey indicates gene transcripts down- regulated in both groups and no shading indicates gene transcripts that were not uniformly up/ down-regulated in both samples. The gene transcripts within each category are sorted based on fold change in the CM samples. Gene transcripts highlighted in bold also passed the criteria of difference in fold change> 1.2, difference of means> 50 in the CM samples and are also listed elsewhere in section 3.8.8.

RPMI vs. RPMI TAX CM and Cells

A large amount of gene transcripts (130) found to be significantly differentially expressed in CM samples were also found to be significantly differentially expressed in the cell samples when the sample groups tested were RPMI against RPMI TAX and criteria for classification as significant was a p-value of less than 0.05 (see Tables 3.8.13.3 (a) and (b)). The majority of these gene transcripts (62) were up-regulated in both RPMI TAX CM and cell RNA samples compared to RPMI CM and cell samples. Just 13% (10 gene transcripts), were down- regulated in both CM and cell RPMI TAX samples compared to the CM and cell RPMI samples. Of these 72 that were similarly up/ down-regulated in both the CM and cell samples, 20 were further classified as significant in regard to the CM samples on the basis that both fold change and difference in expression was greater than 1.2 fold and 50 respectively.

18 of the 52 gene transcripts that showed different trends of expression change in the CM and cell sample groups were also further classified as significant with regard to the CM samples because of their fold change and difference in expression status. Interestingly three of these gene transcripts, showed the same expression trends in the RPMI CM and cell samples compared to the RPMI ML CM and cell sample groups (see Table 3.8.11.2). Adenylate kinase 3 (204347_at) was up-regulated by 11 fold and down-regulated by 7.5 fold in the RPMI ML CM and cell RNA samples respectively when compared to the RPMI CM and cell RNA. It was similarly up-regulated by 10.7 fold and down-regulated by 7 fold in the RPMI TAX CM and cell RNA samples respectively in comparison to the RPMI CM and cell sample groups.

Catechol-O-methyltransferase (208817_at) was down-regulated by 1.4 fold and 1.8 fold in the RPMI ML and TAX CM samples respectively in comparison to the parental RPMI CM samples. It was also up-regulated by 4 fold and 4.7 fold in the RPMI ML and RPMI TX

cell samples respectively in comparison to the RPMI parental cell RNA samples. Mannosidase, alpha, class 2A, member 1 gene is down-regulated by 3.7 fold and 3 fold in the RPMI ML and TAX CM sample groups respectively in comparison to the RPMI CM samples and is up-regulated by 7 fold and 6.2 fold respectively in the RPMI ML and TAX cell RNA sample groups respectively in comparison to the RPMI cell RNA samples.

			CM			CELLS		
probe set	gene	Accession	fold change	diff of means	P value	fold change	diff of means	P value
243410_at	gb:AI340002 /DB_XREF=gi:4076929 /DB_XR	AI340002	33.55	32.55	0.024651	2.98	22.96	0.019938
225280_x_at	arylsulfatase D	N51673	20.17	19.17	0.028927	3.22	51.18	0.004316
227326_at	FLJ46603 protein	BE966768	15.66	14.66	0.037998	1.73	76.8	0.041453
230793_at	gb:BE671038 /DB_XREF=gi:10031579 /DB_>	BE671038	15.06	32.62	0.031075	5.25	609.34	0.005796
237864_at	gb:AI990122 /DB_XREF=gi:5837003 /DB_XR	AI990122	12.96	37.54	0.008567	4.33	46.43	0.032064
215148_s_at	amyloid beta (A4) precursor protein-binding, f	AI141541	11.26	16.41	0.021855	2.15	79.17	0.024466
225142_at	KIAA1718 protein	AW294022	10.34	20.54	0.047348	1.90	15.38	0.043093
63825_at	Abhydrolase domain containing 2	AI557319	10.23	9.23	0.038724	2.83	445.82	0.013117
213844_at	homeo box A5	NM_019102	9.68	41.15	0.025493	54.07	3035.1	0.005721
200681_at	glyoxalase I	NM_006708	9.48	40.94	0.024752	1.81	3044.03	0.001201
1555392_at	Testis derived transcript (3 LIM domains)	AY143171	9.32	24.78	0.046628	38.86	128.71	0.046896
218757_s_at	UPF3 regulator of nonsense transcripts hom	NM_023010	8.31	12.36	0.047165	1.99	290.1	0.003452
240939_x_at	gb:T97999 /DB_XREF=gi:747344 /DB_XREF=	T97999	8.12	41.63	0.026269	3.07	38.72	0.015051
1552309_a_at	nexilin (F actin binding protein)	NM_144573	7.70	20.17	0.038591	7.77	189.53	0.003345
64486_at	coronin, actin binding protein, 1B	AI341234	7.00	26.22	0.035041	1.23	61.65	0.039208
227227_at	**Transcribed locus, moderately similar to**	**AI344332**	**7.03**	**136.32**	**0.014621**	**1.64**	**44.22**	**0.043011**
1563467_at	Platelet derived growth factor C	AL833266	6.44	24.32	0.045599	10.38	23.07	0.035595
204653_at	transcription factor AP-2 alpha (activating enh	BF343007	6.03	19.12	0.042088	24.53	549.05	0.002267
201924_at	myeloid/lymphoid or mixed-lineage leukemia	NM_005936	5.99	49.86	0.009319	9.11	372.84	0.002993
230621_at	hypothetical protein LOC285148	AA502936	5.23	34.31	0.022221	1.59	788.93	0.04011
230466_s_at	**gb:AI092770 /DB_XREF=gi:3431746 /DB_XI**	**AI092770**	**5.07**	**66.62**	**0.013971**	**1.74**	**525.56**	**0.009844**
236251_at	gb:AA226366 /DB_XREF=gi:1849976 /DB_XF	AA226366	5.06	23.89	0.017469	28.60	1374.93	0.006584
240594_at	gb:W86659 /DB_XREF=gi:1400535 /DB_XRE	W86659	5.01	37.49	0.044713	2.87	44.14	0.03351
212608_s_at	Nudix (nucleoside diphosphate linked moiety	W85912	4.57	45.56	0.035479	1.61	251.66	0.003557
244625_at	gb:AW629478 /DB_XREF=gi:7376268 /DB_X	AW629478	4.53	45.85	0.028591	5.05	50.57	0.016209
219108_x_at	**DEAD (Asp-Glu-Ala-Asp) box polypeptide**	**NM_017895**	**4.51**	**68.77**	**0.029036**	**3.09**	**794.9**	**0.000231**
237330_at	**gb:AA603494 /DB_XREF=gi:2437355 /DB_>**	**AA603494**	**4.43**	**50.61**	**0.030548**	**2.50**	**22.76**	**0.01352**
213101_s_at	ARP3 actin-related protein 3 homolog (yeast)	Z78330	4.38	15.82	0.032696	1.63	1384.57	0.000076
221580_s_at	hypothetical protein MGC5306	BC001972	4.05	33.97	0.030329	1.53	434.08	0.001932
238549_at	**core-binding factor, runt domain, alpha s**	**AI420611**	**3.77**	**115.41**	**0.028758**	**1.77**	**59.65**	**0.029817**
239811_at	gb:BF954306 /DB_XREF=gi:12371581 /DB_>	BF954306	3.64	43.07	0.03211	2.09	25.15	0.038278
213221_s_at	salt-inducible serine/threonine kinase 2	AB018324	3.60	41.24	0.03322	1.56	55.38	0.028011
202600_s_at	nuclear receptor interacting protein 1	AI824012	3.58	32.84	0.043991	4.67	313.16	0.004754
226332_at	**hypothetical protein MGC40405**	**AI569932**	**3.40**	**58.41**	**0.039815**	**1.67**	**382.4**	**0.00359**
212195_at	Interleukin 6 signal transducer (gp130, oncos	AL049265	3.36	30.03	0.021098	1.63	267.94	0.01027
205609_at	angiopoietin 1	NM_001146	3.29	38.66	0.020532	116.85	1046.47	0.008277
225412_at	hypothetical protein FLJ14681	AA761169	3.06	28.56	0.036369	2.84	185.45	0.046019
215693_x_at	DEAD (Asp-Glu-Ala-Asp) box polypeptide 27	AL512707	3.03	48.08	0.021788	3.11	780.36	0.000231
1554402_a_at	SAR1a gene homolog 2 (S. cerevisiae)	BC002847	2.81	20.59	0.048319	1.31	63.6	0.021741
200833_s_at	**RAP1B, member of RAS oncogene family**	**NM_015646**	**2.78**	**52.3**	**0.034405**	**1.89**	**1346.01**	**0.001562**
239131_at	gb:AI304862 /DB_XREF=gi:3988551 /DB_XR	AI304862	2.70	41.11	0.03911	3.02	88.41	0.001365
242759_at	**gb:AI821726 /DB_XREF=gi:5440805 /DB_XI**	**AI821726**	**2.69**	**83.32**	**0.005837**	**2.87**	**36.95**	**0.010399**
218858_at	**DEP domain containing 6**	**NM_022783**	**2.58**	**51.64**	**0.035523**	**338.32**	**3906.22**	**0.005662**
236501_at	gb:AI279296 /DB_XREF=gi:3917530 /DB_XR	AI279296	2.52	25.88	0.046949	18.60	114.72	0.028721
208641_s_at	ras-related C3 botulinum toxin substrate 1 (rh	BC004247	2.51	31.86	0.04712	1.66	1073.91	0.000279
1554730_at	multiple C2-domains with two transmembran	BC030005	2.38	27.23	0.02666	1.81	27.7	0.010337
241786_at	gb:AI380514 /DB_XREF=gi:4190367 /DB_XR	AI380514	2.37	38.51	0.049774	2.99	60.09	0.008349
232256_s_at	**Hypothetical gene supported by AK09271**	**AF143329**	**2.23**	**96.03**	**0.031818**	**3.59**	**31.19**	**0.01374**
202845_s_at	ralA binding protein 1	NM_006786	2.20	35.27	0.04603	1.59	378.34	0.000673
212854_x_at	**AG1**	**AB051480**	**2.09**	**85.15**	**0.040579**	**2.66**	**825.51**	**0.003601**
200817_x_at	**ribosomal protein S10**	**NM_001014**	**2.08**	**111.67**	**0.040589**	**1.23**	**1978.48**	**0.008517**
214665_s_at	calcium binding protein P22	AK000095	2.07	29.83	0.045841	2.10	1739.54	0.005456
242920_at	**gb:AW590838 /DB_XREF=gi:7277980 /DB_**	**AW590838**	**2.05**	**71.43**	**0.039764**	**4.45**	**28.02**	**0.009525**
235429_at	gb:AW965494 /DB_XREF=gi:8156330 /DB_X	AW965494	2.00	28.31	0.047878	1.36	58.38	0.030405
200807_s_at	heat shock 60kDa protein 1 (chaperonin)	NM_002156	1.97	30.05	0.048458	1.32	1985.77	0.002192
203819_s_at	IGF-II mRNA-binding protein 3	AU160004	1.96	24.96	0.046039	3.69	762.54	0.005703
200801_x_at	**actin, beta**	**NM_001101**	**1.73**	**59.12**	**0.027541**	**1.46**	**3110.75**	**0.004302**
1560562_a_a	**Hypothetical protein MGC48625**	**AK026366**	**1.56**	**83.28**	**0.026495**	**8.27**	**295.22**	**0.000462**
231886_at	**Hypothetical LOC388572**	**AL137655**	**1.47**	**51.74**	**0.040335**	**3.33**	**54.69**	**0.024656**
35150_at	tumor necrosis factor receptor superfamily, m	X60592	1.33	41.46	0.007928	2.23	126.3	0.007043
235081_x_at	hypothetical protein LOC201292	AI739057	1.06	347	0.045768	1.76	159.62	0.012831
215868_x_at	TGF-betaIIR beta	AK026238	1.06	304.37	0.024296	1.19	663.47	0.029022

Table 3.8.13.3 Probe sets significantly upregulated in both CM and Cell RNA in RPMI vs. RPMI TAX groups.

			CM			CELLS		
probe set	gene	Accession	fold change	diff of means	P value	fold change	diff of means	P value
1555517_at	Transcribed locus, highly similar to NP_1500	BC045709	-6.87	-22.74	0.034106	-102.56	-101.56	0.002681
206491_s_at	N-ethylmaleimide-sensitive factor attachr	NM_003827	-5.89	-90.17	0.016714	-1.78	-135.82	0.04968
227446_s_at	chromosome 14 open reading frame 167	BF445127	-4.38	-21.63	0.044575	-1.58	-73.19	0.028613
218698_at	likely ortholog of mouse monocyte macr	NM_015957	-3.89	-147.93	0.007462	-1.19	-99.15	0.040887
225276_at	G1 to S phase transition 1	AA143579	-3.38	-77.98	0.027384	-3.90	-474.33	0.0015
231274_s_at	gb:R92925 /DB_XREF=gi:965279 /DB_XREF	R92925	-2.87	-20.05	0.038728	-1.47	-37.38	0.04574
226137_at	High mobility group nucleosomal binding dor	AI288759	-2.82	-37.6	0.036898	-2.79	-260.23	0.007224
233878_s_at	5'-3' exoribonuclease 2	BE536170	-2.23	-24.26	0.031227	-1.67	-297.06	0.009721
208316_s_at	oculocerebrorenal syndrome of Lowe	M88162	-2.02	-52.76	0.027333	-1.69	-81.97	0.016098
221096_s_at	hypothetical protein PRO1580	NM_018502	-1.59	-34.22	0.048443	-1.86	-94.26	0.00196
212046_x_at	mitogen-activated protein kinase 3	X60188	-1.56	-49.54	0.04765	-1.86	-131.74	0.011983
206018_at	forkhead box G1B	NM_005249	31.09	30.09	0.006249	-753.85	-752.85	0.004867
236115_at	gb:AA035771 /DB_XREF=gi:1507599 /DB_XF	AA035771	19.50	25.53	0.043478	-7.98	-272.15	0.007092
201498_at	ubiquitin specific protease 7 (herpes virus-ass	AJ160440	17.16	16.16	0.039074	-1.79	-420.3	0.013539
212440_at	putative nucleic acid binding protein RY-1	X76302	16.13	21.34	0.044901	-1.17	-143.52	0.012264
226602_s_at	Breakpoint cluster region /// Similar to breakp	T30183	15.73	14.73	0.018282	-1.73	-62.31	0.034776
232099_at	protocadherin beta 16	AB046841	14.11	19.01	0.015752	-12.59	-46.92	0.004482
229533_x_at	hypothetical protein FLJ90430	AI963028	11.89	11.65	0.048161	-1.47	-95.54	0.027066
232365_at	CDNA FLJ10247 fis, clone HEMBB1000705	AU146318	11.63	18.07	0.04078	-96.95	-95.95	0.001418
204347_at	adenylate kinase 3 /// adenylate kinase 3	AI653169	10.65	72.6	0.010142	-7.15	-573.84	0.017783
213191_at	TIR domain containing adaptor inducing interf	AF070530	10.06	9.06	0.049502	-1.58	-41.95	0.008951
233019_at	CCR4-NOT transcription complex, subunit 7	AU145061	8.18	21.18	0.046858	-2.66	-39.67	0.016758
201038_s_at	Acidic (leucine-rich) nuclear phosphoprotein 3	T67821	8.16	18.82	0.027986	-1.26	-273.38	0.010899
214363_s_at	matrin 3	AA129420	8.03	24.04	0.024793	-1.54	-1367	0.000578
222266_at	gb:BF796940 /DB_XREF=gi:12101994 /DB_X	BF796940	7.97	33.52	0.040458	-1.35	-70.6	0.026309
224851_at	cyclin-dependent kinase 6	AW274756	7.42	58.46	0.01727	-115.25	-2568.31	0.017487
230479_at	gb:AI872374 /DB_XREF=gi:5546423 /DB_XR	AI872374	7.33	25.63	0.033418	-2.41	-127.59	0.022553
223504_at	Ras-associated protein Rap1	AL117490	5.89	24.03	0.031577	-3.10	-56.55	0.005639
202850_at	ATP-binding cassette, sub-family D (ALD)	NM_002858	5.29	64.55	0.004703	-3.34	-1741.8	0.000318
204712_at	WNT inhibitory factor 1	NM_007191	4.82	76.27	0.037368	-20.87	-19.87	0.031208
235381_at	Hepatitis B virus x associated protein	BF033855	4.73	51.6	0.032579	-1.55	-331.28	0.018527
239376_at	CDNA clone IMAGE:4333081, partial cds	AA489041	4.38	18.09	0.035471	-2.40	-396.57	0.001206
202960_s_at	methylmalonyl Coenzyme A mutase	NM_000255	3.87	29.24	0.04914	-1.57	-179.9	0.033708
218412_s_at	GTF2I repeat domain containing 1	NM_016328	3.86	52.09	0.048293	-1.81	-475.34	0.005026
228462_at	hequois homeobox protein 2	AI928035	3.74	204.57	0.007047	-34.60	-34.27	0.034861
230713_at	gb:BF115786 /DB_XREF=gi:10985262 /DB_X	BF115786	3.47	50.75	0.02118	-2.08	-43.92	0.006795
239082_at	Homo sapiens, clone IMAGE:5311370, mRN	BF437161	3.41	59.27	0.028984	-4.96	-677.82	0.008546
232113_at	Hypothetical gene supported by BX64760	N90870	3.14	92.51	0.010047	-31.56	-216.96	0.045021
1555370_a_at	calmodulin binding transcription activator 1	AF111804	2.99	42.34	0.047786	-2.73	-272.48	0.000465
226603_at	chromosome 7 open reading frame 6	BE966604	2.91	18.53	0.028849	-3.47	-39.06	0.007781
208905_at	cytochrome c, somatic	BC005299	2.59	42.1	0.030713	-1.32	-1351.65	0.001339
232097_at	Chromosome 14 open reading frame 92	AA657818	2.58	35.66	0.045155	-2.11	-28.54	0.009995
232156_at	CDNA: FLJ21268 fis, clone COL01718	AK024921	2.53	24.51	0.031658	-1.56	-34.72	0.028544
212153_at	pogo transposable element with ZNF domain	AB007930	1.53	38.01	0.011103	-1.93	-510.26	0.00525
236267_at	zinc finger protein 346	BG178775	1.52	68.65	0.030424	-1.72	-47.38	0.011077
221270_s_at	queuine tRNA-ribosyltransferase 1 (tRNA-	NM_031209	1.44	81.11	0.036975	-2.26	-230.35	0.005677
214329_x_at	gb:AW474434 /DB_XREF=gi:7044540 /DB_X	AW474434	-1.09	-381.58	0.016468	381.26	665.46	0.035813
228541_x_at	Hypothetical gene supported by AK098314	AI049608	-1.14	-266.88	0.025512	2.37	188.92	0.022134
214643_x_at	bridging integrator 1	BG034080	-1.19	-132.94	0.037713	11.02	47.09	0.046574
219113_x_at	dehydrogenase/reductase (SDR family) n	NM_016246	-1.22	-459.85	0.033826	4.33	181.57	0.005791
218836_at	ribonuclease P 21kDa subunit	NM_024839	-1.26	-255.48	0.025326	1.15	159.28	0.049646
207138_at	PHD finger protein 2	NM_005392	-1.29	-925.98	0.0104	20.29	29.52	0.017393
203565_s_at	menage a trois 1 (CAK assembly factor)	NM_002431	-1.39	-98.06	0.009458	1.59	128.89	0.022958
230423_at	Hypothetical LOC388638	AI554075	-1.59	-42.23	0.041907	6.06	113.9	0.021205
219358_s_at	centaurin, alpha 2	NM_018404	-1.62	-166.05	0.012815	3.32	71.3	0.008742
208817_at	catechol-O-methyltransferase	BC000419	-1.81	-461.62	0.003261	4.76	353.02	0.001387
202223_at	integral membrane protein 1	NM_002219	-2.33	-101.41	0.011746	1.90	485.37	0.003391
226784_at	TWIST neighbor	AA121481	-2.78	-37.3	0.013352	2.12	224.7	0.000347
234797_at	gb:AL049794 /DB_XREF=gi:10280528 /FEA=	AL049794	-2.96	-37.5	0.038123	1.93	119.08	0.021669
226538_at	mannosidase, alpha, class 2A, member 1	AV700323	-2.97	-73.43	0.033573	6.20	194.23	0.001568
201422_at	interferon, gamma-inducible protein 30	NM_006332	-3.07	-67.49	0.046016	129.21	456.41	0.000142
212419_at	chromosome 10 open reading frame 56	AA131324	-4.45	-25.74	0.027484	6.80	130.77	0.008499
203530_s_at	syntaxin 4A (placental)	NM_004604	-6.36	-173.14	0.020926	3.51	267.65	0.001315
201301_s_at	annexin A4	BC000182	-7.94	-23.04	0.040717	2.47	677.47	0.000194

Table 3.8.13.3 (contd) Probe sets significantly downregulated/ changed in both CM and Cell RNA in RPMI vs. RPMI TAX groups. Gene transcripts shaded light grey are up- regulated in both samples, dark grey indicates gene transcripts down- regulated in both groups and no shading indicates gene transcripts that were not uniformly up/ down-regulated in both samples. The gene transcripts within each category are sorted based on fold change in the CM samples. Gene transcripts highlighted in bold also passed the criteria of difference in fold change> 1.2, difference of means> 50 in the CM samples and are also listed elsewhere in section 3.8.7.

MDA F vs. MDA F ADR CM and Cells

The results listed in Table 3.8.13.4 show that of the 44 gene transcripts found to be differentially expressed in the MDA-F-ADR-SI CM and Cell RNA samples compared to the MDA-F CM and cell samples, the majority of these (31) are differentially regulated in the CM samples compared to the cell sample groups.

11 gene transcripts are similarly up-regulated in both MDA-F-ADR-SI CM and cell sample groups compared to the MDA-F CM and sample groups respectively. One of these gene transcripts, catechol-O-methyltransferase (208817_at), was found previously to be differentially regulated in CM compared to cell samples (see Tables 3.8.13.3 (b) and 3.8.13.2).

Only 2 of the 44 gene transcripts listed in Table 3.8.13.4 are down-regulated in both MDA-F-ADR-SI CM and cell groups in comparison to the MDA-F CM and cell sample groups respectively.

Overall, of the 44 gene transcripts found to be significant in both MDA-F-ADR-SI CM and cell groups compared to the MDA-F CM and cell groups, 10 of them were further classified as significant in the CM samples due to the fold change and difference of expression being greater than 1.2 and 50 respectively between the parental and drug selected variant CM sample groups.

			CM			CELLS		
probe set	gene	Accession	fold change	diff of means	P value	fold change	diff of means	P value
202102_s_at	bromodomain containing 4	BF718610	11.79	41.98	0.005651	1.76	134.16	0.049284
200004_at	eukaryotic translation initiation factor 4 gamn	NM_001418	5.26	24.56	0.038482	1.11	608.03	0.041439
219762_s_at	**ribosomal protein L36**	**NM_015414**	**4.50**	**50.99**	**0.042117**	**1.31**	**1350.88**	**0.048064**
202505_at	small nuclear ribonucleoprotein polypeptide E	NM_003092	3.56	34.54	0.03698	1.53	1570.72	0.00813
212284_x_at	**tumor protein, translationally-controlled**	**BG498776**	**3.09**	**127.31**	**0.018554**	**1.18**	**1413.62**	**0.028621**
1558173_a_a	**leucine zipper protein 1**	**AK093016**	**2.71**	**56.98**	**0.012152**	**3.06**	**118.94**	**0.003801**
211943_x_at	tumor protein, translationally-controlled 1	AL565449	2.32	34.93	0.042745	1.19	1452.59	0.018735
214327_x_at	tumor protein, translationally-controlled 1	AI888178	2.05	28.45	0.044854	1.14	832.03	0.041186
1553830_s_at	melanoma antigen, family A, 2B /// melanom	NM_005361	2.03	14.79	0.047465	2.05	491.23	0.002197
1567139_at	gb:AJ297963.1 /DB_XREF=gi:14475538 /TID=	AJ297963	1.41	18.78	0.046961	1.61	17.95	0.029331
208817_at	**catechol-O-methyltransferase**	**BC000419**	**1.37**	**190.71**	**0.000717**	**1.75**	**727.96**	**0.012579**
226636_at	phospholipase D1, phosphatidylcholine-specifi	AI378587	-1.97	-20.74	0.043564	-2.28	-125.38	0.015622
1554663_a_a	nuclear mitotic apparatus protein 1	BC043499	-1.40	-118.72	0.048052	-3.75	-27.02	0.010018
209375_at	xeroderma pigmentosum, complementation g	D21089	12.97	11.97	0.040173	-2.82	-91.29	0.003691
1553570_x_a	**cytochrome c oxidase II**	**NM_173705**	**12.86**	**223.41**	**0.002722**	**-1.11**	**-1222.55**	**0.008342**
216321_s_at	nuclear receptor subfamily 3, group C, memb	X03348	11.63	16.05	0.033238	-1.70	-173.2	0.001617
211969_at	**heat shock 90kDa protein 1, alpha**	**BG420237**	**11.07**	**209.55**	**0.000107**	**-1.27**	**-1663.78**	**0.014906**
226707_at	similar to CG3714 gene product	AI925574	9.24	19.7	0.02122	-1.36	-109.17	0.01812
200610_s_at	nucleolin	NM_005381	7.19	56.06	0.008283	-1.10	-453.19	0.046023
232341_x_at	hyaluronan binding protein 4	AK025144	6.80	23.24	0.027273	-1.84	-38.87	0.0324
200006_at	Parkinson disease (autosomal recessive, ear	NM_007262	5.98	57.39	0.03086	-1.40	-1755.76	0.000106
214114_x_at	FAST kinase	AK023141	5.92	50.16	0.043193	-1.34	-217.68	0.025542
225726_s_at	pleckstrin homology domain containing, famil	AB033026	5.10	15.9	0.046458	-1.59	-63.2	0.010472
221637_s_at	hypothetical protein MGC2477	BC001434	5.04	26.49	0.022142	-1.58	-642.96	0.005023
225601_at	gb:AI806853 /DB_XREF=gi:5393419 /DB_XR	AI806853	4.98	17.55	0.03389	-1.58	-72.5	0.027233
201812_s_at	hypothetical protein LOC201725 /// transloca	NM_019059	4.70	36.61	0.005587	-1.14	-638.32	0.026942
200031_s_at	**ribosomal protein S11 /// ribosomal prote**	**NM_001015**	**4.24**	**94.47**	**0.041127**	**-1.45**	**-3210.34**	**0.047139**
201091_s_at	chromobox homolog 3 (HP1 gamma homoloc	BE748755	4.01	26.82	0.023565	-1.49	-1235.6	0.000586
222390_at	WW domain containing adaptor with coiled-c	AL135461	3.88	35.46	0.013863	-1.70	-591.59	0.006782
212333_at	DKFZP564F0522 protein	AL049943	3.66	30.55	0.037507	-1.27	-266.45	0.015118
202233_s_at	ubiquinol-cytochrome c reductase hinge prote	NM_006004	3.61	16.73	0.048895	-1.22	-1167.96	0.015139
204949_at	intercellular adhesion molecule 3	NM_002162	3.55	14.17	0.040598	-1.25	-54.88	0.039596
210213_s_at	integrin beta 4 binding protein	AF022229	3.47	24.53	0.014301	-1.28	-685.72	0.006994
203113_s_at	eukaryotic translation elongation factor 1 delt	NM_001960	3.43	34.41	0.034992	-1.20	-602.69	0.016292
221700_s_at	ubiquitin A-52 residue ribosomal protein fusio	AF348700	3.32	47.97	0.010711	-1.33	-1565.5	0.022225
212593_s_at	programmed cell death 4 (neoplastic transfor	N92498	2.84	13.79	0.033609	-1.88	-222.69	0.005819
201306_s_at	**acidic (leucine-rich) nuclear phosphoprot**	**NM_006401**	**2.76**	**131.08**	**0.027315**	**-1.34**	**-930.68**	**0.023772**
211025_x_at	cytochrome c oxidase subunit Vb /// cytochr	BC006229	1.87	19.7	0.048589	-1.42	-1285.9	0.001017
1554577_a_at	proteasome (prosome, macropain) 26S subui	AY057056	1.86	19.7	0.049893	-1.31	-285.56	0.030427
221488_s_at	**chromosome 6 open reading frame 82**	**AF230924**	**1.74**	**72.47**	**0.017006**	**-1.35**	**-609.24**	**0.003306**
201829_at	neuroepithelial cell transforming gene 1	AW263232	1.58	26	0.010455	-1.34	-123.82	0.01589
223463_at	RAB23, member RAS oncogene family	AF161486	1.58	29.43	0.039727	-1.52	-97.86	0.015161
231913_s_at	c6.1A	X64643	-1.23	-23.89	0.04263	1.47	78.56	0.009711
206491_s_at	N-ethylmaleimide-sensitive factor attachment	NM_003827	-2.89	-46.61	0.048118	2.04	344.53	0.01416

Table 3.8.13.4 Probe sets significantly changed in both CM and Cell RNA in MDA-F vs. MDA-F-ADR groups. Gene transcripts shaded light grey are up- regulated in both samples, dark grey indicates gene transcripts down- regulated in both groups and no shading indicates gene transcripts that were not uniformly up/ down-regulated in both samples. The gene transcripts within each category are sorted based on fold change in the CM samples. Gene transcripts highlighted in bold also passed the criteria of difference in fold change> 1.2, difference of means> 50 in the CM samples and are also listed elsewhere in section 3.8.7.

MDA-F vs. MDA-F-TAX-SI CM and Cells

The number of gene transcripts significantly changed in both MDA-F-TAX-SI CM and cell RNA samples in comparison to the parental MDA-F CM and cell RNA groups respectively is the smallest of the sample groups tested. None of the 9 gene transcripts listed in Table 3.8.13.5 are similarly up-regulated in both CM and cell RNA sample groups, 2 are similarly down-regulated in both groups. 7 gene transcripts show different trends of change in expression in the two groups with 6 gene transcripts being up-regulated in the CM samples and down-regulated in the cell samples when the MDA-F sample groups are compared to the MDA-F-TAX-SI groups respectively.

			CM			CELLS		
probe set	gene	Accession	fold change	diff of means	P value	fold change	diff of means	P value
242136_x_at	Hypothetical LOC403340	T66145	-2.48	-55.05	0.039278	-1.41065	-111.22	0.02193
226544_x_at	muted homolog (mouse)	AV734582	-1.28	-170.81	0.035781	-1.38184	-70.59	0.023166
201498_at	ubiquitin specific protease 7 (herpes virus-as	AI160440	3.98	24.43	0.033032	-1.48768	-175.36	0.013647
228452_at	chromosome 17 open reading frame 39	AA827865	3.38	56.63	0.019761	-2.59255	-129.57	0.001584
222380_s_at	Similar to Microneme antigen	AI907083	1.96	287.96	0.013562	-2.40008	-33.84	0.011535
242539_at	gb:AW665509 /DB_XREF=gi:7458056 /DB_X	AW665509	1.95	46.29	0.007488	-2.68856	-50.84	0.003534
225845_at	HSPC063 protein	BG253884	1.55	46.26	0.022492	-2.03083	-204.94	0.002313
213064_at	nuclear protein UKp68	N64802	1.38	35.3	0.026678	-1.58451	-126.81	0.006956
211921_x_at	prothymosin, alpha (gene sequence 28) /	AF348514	-2.54	-302.11	0.045783	1.28063	1509.09	0.013685

Table 3.8.13.5 Probe sets significantly changed in both CM and Cell RNA in MDA-F vs. MDA-F-TAX-SI groups. Gene transcripts shaded dark grey indicate gene transcripts down- regulated in both groups and no shading indicates gene transcripts that were not uniformly up/ down-regulated in both samples. No gene transcripts were up-regulated in both CM and cell RNA samples. The gene transcripts within each category are sorted based on fold change in the CM samples. Gene transcripts highlighted in bold also passed the criteria of difference in fold change> 1.2, difference of means> 50 in the CM samples and are also listed elsewhere in section 3.8.7.

The aim of Section 3.9 was to asses whether mRNA transcripts could be amplified from the CM of cultured normal mammary breast cells, and also whether the mRNA could be amplified using the 2 cycle amplification procedure employed in Sections 3.7 and 3.8 and hybridised to an Affymetrix human genome Plus 2.0 microarray chip.

3.9 Pilot Study analysis of CM from normal breast cells.

The pilot DLKP and larger cancer cell line CM array studies yielded valuable information about the mRNA gene expression profiles of extracellular mRNA. However, the study only looked at the difference in gene expression between cancer cell lines. The gene transcripts confirmed to be up or down- regulated in the different drug selected CM samples compared to the parental CM samples or the gene transcripts expressed in all cancer cell line CM samples, whilst they may be candidates for cell- specific markers, it could not be determined whether they could be possible cancer- specific biomarkers.

For this reason it was decided to culture HMEC- 1001 normal human breast cells (HMEC) and extract RNA from both the CM and cells and perform microarray analysis of both. Before the samples were hybridised to the microarray chips however, RT-PCR was carried out to test if a few mRNA transcripts were amplifiable from the normal breast cell CM sample.

3.9.1 RT-PCR of HMEC CM

The results shown in Figs. 3.9.1.1-3 indicate that it is possible to amplify mRNA gene transcripts using the CM from normal cultured breast cells. As a result of this we decided to use the and process both the HMEC CM and cell RNA simultaneously using the 2 cycle amplification procedure for comparison with the CM RNA profile obtained for the cancer cell line CM samples.

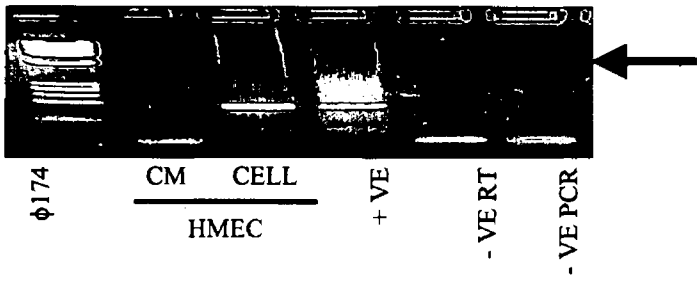


Figure. 3.9.1.1 RT-PCR of HMEC cell and CM RNA for β-actin (171bp). β-actin gene transcripts (arrow) were expressed HMEC cell RNA with a faint band in the HMEC

CM sample. Amplified product was detected in the positive control, RPMI TAX cell RNA, and was undetected in the negative control samples (-VE RT = RT reaction with = H_20 instead of RNA, -VE PCR = PCR reaction with H_20 instead of cDNA as control).

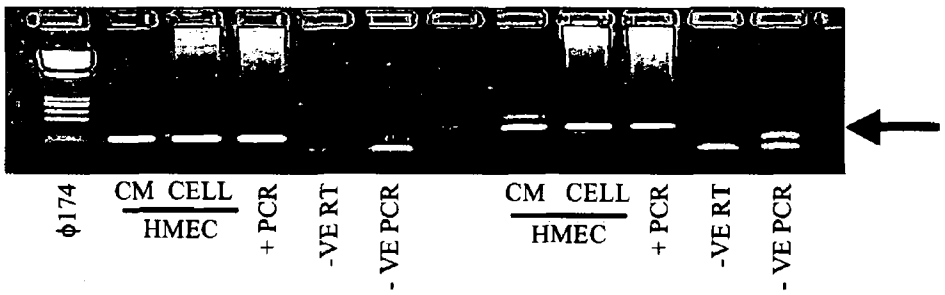


Figure 3.9.1.2 RT-PCR for APIP mRNA (117bp) and Stearoyl-CoA desaturase 5 mRNA (151bp). APIP (red arrow) and Stearoyl- CoA desaturase 5 (black arrow) mRNA gene transcripts are expressed in HMEC CM and cell RNA samples. Amplified product was detected in the positive control (DLKP cell RNA) and was undetected in the negative control samples (-VE RT = RT reaction with = H_20 instead of RNA, -VE PCR = PCR reaction with H_20 instead of cDNA as control).

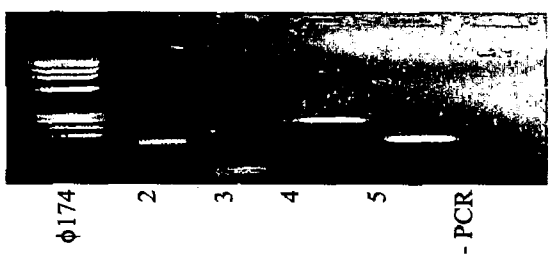


Figure 3.9.1.3 RT-PCR of HMEC CM for various transcripts detectable on Affymetrix microarray chip, lane 2 SMAF1 mRNA 152bp, lane 3 Clone IMAGE 5528716 mRNA 152bp, lane 4 ABCB4 mRNA 239bp, lane 5 Clone HLUNG2003061 mRNA 154bp. Gene transcripts were amplifiable for all transcripts except Clone IMAGE 5528716 mRNA. Amplified products was undetected in the negative control sample (-VE PCR = PCR reaction with H_20 instead of cDNA as control).

3.9.2 Yield of HMEC CM and Cell cRNA after 2 rounds of IVT

There was almost five times more cRNA generated from the HMEC cell RNA than the CM RNA using the 2 cycle amplification procedure (see Table 3.9.2.1). Both samples yielded the required 15μg required for hybridisation to the Affymetrix microarray chip.

Sample	After 1st IVT (ng)	After 2nd IVT (μg)
HMEC CM	131.4	21.12
HMEC cell	465.3	97.65

Table 3.9.2.1 Quantity of RNA before and during 2 cycle amplification protocol, IVT= in vitro transcription

3.9.3 QA parameters of Affymetrix Plus 2.0 chips

Both the HMEC CM and cell samples had acceptable QA parameters as indicated by the results shown in Table 3.9.3.1. However the large difference in scaling factor between the two samples means that the two cannot be directly compared against each other. The % Present call for the cell RNA sample is six times greater than the % Present call in the CM sample.

Parameter	HMEC CM	HMEC Cell
Scaling Factor	25.6	0.939
Noise	1.49	1.850
Background	48.9	61.04
% Present	7	42.4
3'/ 5' GAPDH ratio	11.3	1.48
3'/M GAPDH ratio	3.93	1.12

Table 3.9.3.1 QA results of Affymetrix U133A Plus 2.0 chips for HMEC samples

3.9.4 Probe sets called P or M in HMEC CM and Cell RNA

3,393 probe sets are called P or M in HMEC CM and A in HMEC cell RNA (see Fig. 3.9.4.1). 22,527 probe sets are called P or M in HMEC cell RNA and A in HMEC CM RNA. 6,906 probe sets are called P or M in both samples, which is 67% of the gene transcripts called P or M in HMEC CM. Only 10% of these gene transcripts are expressed at levels over 100 in the HMEC CM RNA samples. The % correlation of the probe sets called P or M in both the CM and cell RNA is increased in comparison to the cancer cell line microarray study (Section 3.8.12). The average correlation of the probe sets called P or M in the CM and cell samples was 56%, this indicates that the use of two different amplification protocols for the CM and cell samples may have impacted slightly on the results obtained and perhaps if the 2 cycle amplification protocol was used for both sets of samples, the correlation of gene transcripts called P or M in CM and cell RNA samples would have been higher.

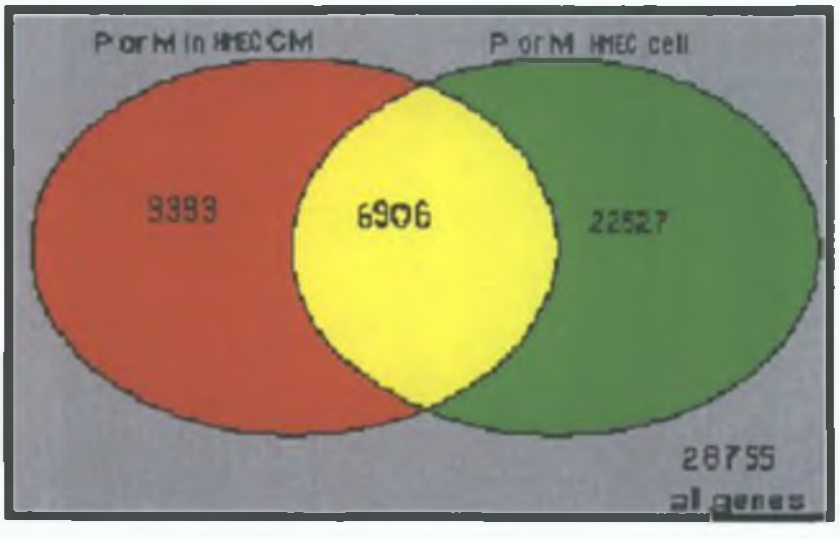


Figure 3.9.4.1 Gene transcripts called P or M in HMEC CM vs. Gene transcripts called P or M in HMEC cell RNA. The numbers of probe sets called P or M in HMEC CM RNA sample (left circle) is overlapped with the number of probe sets called P or M in HMEC cell RNA (right circle).

3.9.5 Comparison of normal breast CM with cancer cell line CM

The gene transcripts that were called Present in all CM samples (including the DLKP CM samples from Section 3.7) were compared to the gene transcripts that were called Absent in the HMEC CM sample. The 23 gene transcripts listed in Table 3.9.5.1 were the resulted from that comparison.

probe set	gene
1316_at	thyroid hormone receptor, alpha (erythroblastic leukemia viral (v-erb-a) oncogene homolog, avian)
213605_s_at	FLJ40092 protein
216508_x_at	gb:AC007277 /DB_XREF=gi:5091647 /FEA=DNA /CNT=1 /TID=Hs.283906.0 /TIER=ConsEnd /STK=0 /UG=Hs
217586_x_at	gb:N35922 /DB_XREF=gi:1157064 /DB_XREF=yy28g05.s1 /CLONE=IMAGE:272600 /FEA=EST /CNT=3 /TID=
81811_at	Similar to Microneme antigen
227338_at	Hypothetical gene supported by BC066916
227917_at	MRNA: cDNA DKFZp434N2116 (from clone DKFZp434N2116)
230296_at	Hypothetical protein LOC146174
235327_x_at	UBX domain containing 4
235368_at	A disintegrin-like and metalloprotease (reprolysin type) with thrombospondin type 1 motif, 5 (aggrecanase-2)
236841_at	gb:BE464132 /DB_XREF=gi:9509907 /DB_XREF=hy21b07.x1 /CLONE=IMAGE:3197941 /FEA=EST /CNT=6 /T
237180_at	gb:T97717 /DB_XREF=gi:747062 /DB_XREF=ye54e08.s1 /CLONE=IMAGE:121574 /FEA=EST /CNT=5 /TID=H
237483_at	gb:AI990790 /DB_XREF=gi:5837671 /DB_XREF=ws23g11.x1 /CLONE=IMAGE:2498084 /FEA=EST /CNT=7 /TI
238156_at	gb:AW205632 /DB_XREF=gi:6505106 /DB_XREF=UI-H-BI1-afr-e-09-0-UI.s1 /CLONE=IMAGE:2722673 /FEA=E
238348_x_at	gb:AW390231 /DB_XREF=gi:6894890 /DB_XREF=CM3-ST0181-111199-033-a12 /FEA=EST /CNT=25 /TID=Hs
238431_at	gb:W68845 /DB_XREF=gi:1376076 /DB_XREF=zd43e07.s1 /CLONE=IMAGE:343380 /FEA=EST /CNT=13 /TID
241835_at	Hypothetical gene supported by BC036435
241996_at	gb:AI669591 /DB_XREF=gi:4834365 /DB_XREF=tw34b09.x1 /CLONE=IMAGE:2261561 /FEA=EST /CNT=5 /TI
242171_at	gb:AA693730 /DB_XREF=gi:2694668 /DB_XREF=zi55d08.s1 /CLONE=IMAGE:434703 /FEA=EST /CNT=4 /TID
242961_x_at	DEAD (Asp-Glu-Ala-Asp) box polypeptide 58
1553703_at	hypothetical protein FLJ90396
1555977_at	LOC440476
1558982_at	FLJ00310 protein

Table 3.9.5.1 Probe sets/ gene transcripts called P in all human cancer cell line CM samples and A in HMEC CM.

The gene transcripts that were differentially expressed in the HMEC CM sample in comparison to the panel of cancer cell line CM samples are listed in Table 3.9.5.2. As only one HMEC CM sample was available, statistical significance could not be ascertained. Because of this, a stricter criterion for selection was employed; in order to be included on the list the gene transcripts had to show a fold change greater than 2 and a difference in expression greater than 100.

143 gene transcripts passed this criterion with 55 of them up-regulated in the HMEC CM compared to the panel of cancer cell line CM samples and the remaining 88 were down-regulated in the HMEC CM compared to the panel of cancer cell line CM samples.

probe set	gene	Accession	All CM mean	HMEC	fold change	difference of means
204971_at	cystatin A (stefin A)	NM_005213	4.03	195.9	48.62	191.87
204268_at	S100 calcium binding protein A2	NM_005978	12.17	524.42	43.08	512.25
209351_at	keratin 14 (epidermolysis bullosa simplex, Dow	BC002690	7.11	175.7	24.72	168.59
203324_s_at	caveolin 2	NM_001233	8.79	152.73	17.37	143.94
209814_at	zinc finger protein 330	BC004421	6.58	106.59	16.2	100.01
207425_s_at	MLL septin-like fusion	NM_006640	10.28	165.92	16.14	155.64
218179_s_at	FLJ12716 protein	NM_021942	7.59	120.38	15.85	112.79
214211_at	ferritin, heavy polypeptide 1	AA083483	14.27	191.85	13.44	177.58
1561938_at	gb:AL832704.1 /DB_XREF=gi:21733283 /TID=I	AL832704	10.47	136.25	12.92	124.78
212185_x_at	gb:NM_005953.1 /DB_XREF=gi:5174763 /GEN	NM_005953	55.02	697.82	12.68	642.8
213680_at	keratin 6B	AI831452	9.57	114.73	11.99	105.16
33322_i_at	stratifin	X57348	31.47	317.68	10.1	286.21
200748_s_at	ferritin, heavy polypeptide 1	NM_002032	94.23	906.55	9.62	812.32
33323_r_at	stratifin	X57348	30.44	285.54	9.38	255.1
211628_x_at	ferritin, heavy polypeptide pseudogene 1 /// ferr	J04755	179.95	1618.42	8.99	1438.47
212859_x_at	metallothionein 1E (functional)	BF217861	13.85	124.01	8.95	110.15
215955_x_at	Rho GTPase activating protein 26	Y10388	225.28	1731.22	7.68	1505.94
228578_at	developmentally regulated RNA-binding protein	AI471723	118.4	900.16	7.6	781.76
210986_s_at	tropomyosin 1 (alpha)	Z24727	16.46	119.7	7.27	103.24
206461_x_at	metallothionein 1H	NM_005951	18.92	127.44	6.74	108.52
241662_x_at	gb:AW063472 /DB_XREF=gi:6887409 /DB_XRI	AW063472	24	153.79	6.41	129.79
1552853_at	hypothetical protein FLJ32784	NM_144623	254.97	1547.65	6.07	1292.68
222359_x_at	Follicular lymphoma variant translocation 1	BF573849	140.24	837.77	5.97	697.53
206286_s_at	teratocarcinoma-derived growth factor 1	NM_003212	26.41	153.72	5.82	127.3
242264_at	gb:AI973099 /DB_XREF=gi:5769925 /DB_XREI	AI973099	24.21	136.47	5.64	112.26
225961_at	KIAA1340 protein	AI334297	21.97	123.45	5.62	101.48
200912_s_at	eukaryotic translation initiation factor 4A, isofo	NM_001967	51.11	247.82	4.85	196.71
231821_x_at	Transcribed locus, weakly similar to XP_37388	AK024248	346.81	1597.48	4.61	1250.67
231288_at	Similar to 0610010D24Rik protein	BF475370	43.63	183.71	4.21	140.08
233690_at	Chromosome 21 open reading frame 96	AK026743	38.28	152.77	3.99	114.49
241613_at	gb:AW296081 /DB_XREF=gi:6702717 /DB_XRI	AW296081	43.05	156.57	3.64	113.52
243965_at	Galactose-3-O-sulfotransferase 3	AI860920	51.91	181.61	3.5	129.7
216438_s_at	thymosin, beta 4, X-linked /// thymosin-like 3	AL133228	99.97	344.12	3.44	244.15
201533_at	catenin (cadherin-associated protein), beta 1, 8	NM_001904	50.69	167.6	3.31	116.91
228934_x_at	gb:AI792511 /DB_XREF=gi:5340227 /DB_XREI	AI792511	49.16	161.01	3.28	111.85
211956_s_at	putative translation initiation factor	BF246436	199.17	648.74	3.26	449.56
212284_x_at	tumor protein, translationally-controlled 1	BG498776	390.76	1220.54	3.12	829.79
229830_at	gb:AW205919 /DB_XREF=gi:6505393 /DB_XRI	AW205919	116.12	347.77	2.99	231.64
203455_s_at	spermidine/spermine N1-acetyltransferase	NM_002970	53.14	155.51	2.93	102.37
211943_x_at	tumor protein, translationally-controlled 1	AL565449	293.34	840.72	2.87	547.38
207783_x_at	gb:NM_017627.1 /DB_XREF=gi:8923031 /GEN	NM_017627	308.17	880.49	2.86	572.32
214327_x_at	tumor protein, translationally-controlled 1	AI868178	251.55	717.03	2.85	465.48
239175_at	gb:AW203986 /DB_XREF=gi:6503458 /DB_XRI	AW203986	59.73	168.42	2.82	108.7
212869_x_at	tumor protein, translationally-controlled 1	AI721229	328.13	919.1	2.8	590.97
216520_s_at	tumor protein, translationally-controlled 1	AF072098	99.86	267.58	2.68	167.72
242405_at	gb:BF358386 /DB_XREF=gi:11317458 /DB_XR	BF358386	81.59	214.47	2.63	132.88
213288_at	O-acyltransferase (membrane bound) domain c	AI761250	81.82	199.71	2.44	117.89
241593_x_at	gb:BF476913 /DB_XREF=gi:11547740 /DB_XR	BF476913	196.17	474.97	2.42	278.8
212788_x_at	ferritin, light polypeptide	BG537190	83.44	201.37	2.41	117.94
1554375_a_at	nuclear receptor subfamily 1, group H, member	AF478446	98.83	234.03	2.37	135.2
210103_s_at	forkhead box A2	AB028021	526.42	1198.09	2.28	671.68
228253_at	lysyl oxidase-like 3	AI917716	87.63	198.09	2.26	110.46
232150_at	gb:AA134418 /DB_XREF=gi:1691911 /DB_XRE	AA134418	123.84	272.38	2.2	148.54
1563754_at	glutamate receptor, ionotropic, kainate 2	AJ252246	135.16	297.89	2.2	162.73
208110_x_at	ARC/mediator transcriptional coactivator subun	NM_030973	126.51	271.34	2.14	144.83
208668_x_at	high-mobility group nucleosomal binding domai	BC003689	244.06	121.51	-2.01	-122.55
211792_s_at	cyclin-dependent kinase inhibitor 2C (p18, inhit	U17074	277.87	137.24	-2.02	-140.63
200680_x_at	high-mobility group box 1	NM_002128	282.29	137.81	-2.05	-144.48
220960_x_at	ribosomal protein L22	NM_000983	395.96	193.47	-2.05	-202.49
1558698_at	zinc finger protein 264	BF971035	485.88	236.16	-2.06	-249.72
211378_x_at	peptidylprolyl isomerase A (cyclophilin A)	BC001224	375.18	180.65	-2.08	-194.53
211765_x_at	peptidylprolyl isomerase A (cyclophilin A) /// pe	BC005982	327.36	157.64	-2.08	-169.72
201293_x_at	peptidylprolyl isomerase A (cyclophilin A)	NM_021130	332.97	159.38	-2.09	-173.59
211978_x_at	peptidylprolyl isomerase A (cyclophilin A)	AI708767	487.62	230.15	-2.12	-257.47
221775_x_at	ribosomal protein L22	BG152979	388.68	182.05	-2.14	-206.64
213757_at	gb:AA393940 /DB_XREF=gi:2046909 /DB_XRE	AA393940	500.59	232.81	-2.15	-267.77
1555373_at	gb:AF304443.1 /DB_XREF=gi:16588391 /TID=I	AF304443	1537.68	711.67	-2.16	-826.02
220106_at	NPC1 (Niemann-Pick disease, type C1, gene)-	NM_013389	239.66	108.67	-2.21	-130.99

Table 3.9.5.2 Gene transcripts up/ down-regulated in HMEC CM compared to all cancer cell line CM (n=26). The list is sorted by fold change, no shading = up-regulated, grey shading = down-regulated. Criteria for inclusion on the list were a difference in expression of 100 and a fold change greater than 2.

probe set	gene	Accession	All CM mean	HMEC	fold change	difference of means
200021_at	cofilin 1 (non-muscle) /// cofilin 1 (non-muscle)	NM_005507	199.77	89.92	-2.22	-109.85
212661_x_at	peptidylprolyl isomerase A (cyclophilin A)	BE731738	468.25	208.86	-2.24	-259.39
213699_s_at	tyrosine 3-monooxygenase/tryptophan 5-mono	AA854017	224.26	97.4	-2.3	-126.87
200981_x_at	GNAS complex locus	NM_016592	192.26	78.89	-2.44	-113.37
215911_x_at	ATPase, Ca++ transporting, plasma membrane	AW615612	431.25	170.01	-2.54	-261.24
206248_at	protein kinase C, epsilon	NM_005400	236.13	89.46	-2.64	-146.67
200780_x_at	GNAS complex locus	NM_000516	194.08	72.62	-2.67	-121.46
225413_at	upregulated during skeletal muscle growth 5	BG291685	193.92	70.55	-2.75	-123.36
217673_x_at	GNAS complex locus	AA650558	154.05	53.68	-2.87	-100.38
202233_s_at	ubiquinol-cytochrome c reductase hinge protein	NM_006004	273.75	93.8	-2.92	-179.95
209251_x_at	tubulin alpha 6	BC004949	194.32	65.35	-2.97	-128.96
203606_at	NADH dehydrogenase (ubiquinone) Fe-S protei	NM_004553	206.47	68.38	-3.02	-138.09
231628_s_at	gb:AW262311 /DB_XREF=gi:6639127 /DB_XRI	AW262311	213.44	70.31	-3.04	-143.12
1553064_at	H1 histone family, member O, oocyte-specific	NM_153833	148.84	48.75	-3.05	-100.09
220232_at	stearoyl-CoA desaturase 4	NM_024906	879.38	269.12	-3.27	-610.27
1570033_at	WIPI49-like protein 2	BC016912	174.35	53.14	-3.28	-121.21
211931_s_at	heterogeneous nuclear ribonucleoprotein A3	BG505670	175.06	52.24	-3.35	-122.82
217719_at	eukaryotic translation initiation factor 3, subuni	NM_016091	189.35	56.21	-3.37	-133.15
1553588_at	gb:NM_173710.1 /DB_XREF=gi:27754195 /GE	NM_173710	950.57	270.71	-3.51	-679.86
205967_at	histone 1, H4c	NM_003542	282.04	79.72	-3.54	-202.32
201318_s_at	myosin regulatory light chain MRLC2 /// myosi	NM_006471	156.22	43.95	-3.55	-112.27
1555653_at	NADH dehydrogenase 5	BC012090	214.28	59.17	-3.62	-155.11
211969_at	heat shock 90kDa protein 1, alpha	BG420237	208.9	54.03	-3.87	-154.87
238199_x_at	Similar to OK/SW-CL.16	AI708524	784.46	198.16	-3.96	-586.31
217801_at	ATP synthase, H+ transporting, mitochondrial I	NM_006886	329.82	83.07	-3.97	-246.75
224187_x_at	heat shock 70kDa protein 8	AF217511	310.84	77.93	-3.99	-232.91
213911_s_at	H2A histone family, member Z	BF718636	179.43	44.41	-4.04	-135.02
230466_s_at	gb:AI092770 /DB_XREF=gi:3431746 /DB_XRE	AI092770	163.59	40.47	-4.04	-123.12
202428_x_at	diazepam binding inhibitor (GABA receptor mo	NM_020548	136.1	33.2	-4.1	-102.9
237031_at	Full length insert cDNA clone YP08F12	AI743452	167.63	39.44	-4.25	-128.18
1553567_s_at	ATP synthase 6	NM_173702	523	122.98	-4.25	-400.03
208549_x_at	prothymosin, alpha (gene sequence 28)	NM_016171	1032.53	240.5	-4.29	-792.03
205644_s_at	small nuclear ribonucleoprotein polypeptide G	NM_003096	240.78	55.07	-4.37	-185.71
224375_at	gb:AF271776.1 /DB_XREF=gi:12006208 /FEA=	AF271776	205.95	46.97	-4.38	-158.98
211921_x_at	prothymosin, alpha (gene sequence 28) /// prot	AF348514	1113.15	253.53	-4.39	-859.61
200006_at	Parkinson disease (autosomal recessive, early	NM_007262	188.31	42.29	-4.45	-146.02
1553570_x_at	cytochrome c oxidase II	NM_173705	671.15	147.46	-4.55	-523.69
216384_x_at	gb:AF257099 /DB_XREF=gi:8037944 /FEA=DI	AF257099	386.58	84.2	-4.59	-302.38
200772_x_at	prothymosin, alpha (gene sequence 28)	BF686442	883.08	191.09	-4.62	-691.99
221891_x_at	heat shock 70kDa protein 8	AA704004	348.93	73.99	-4.72	-274.95
200943_at	high-mobility group nucleosome binding domai	NM_004965	128.31	25.5	-5.03	-102.81
200807_s_at	heat shock 60kDa protein 1 (chaperonin)	NM_002156	200.98	39.74	-5.06	-161.23
207127_s_at	heterogeneous nuclear ribonucleoprotein H3 (2I	NM_021644	125.05	22.39	-5.58	-102.66
200642_at	superoxide dismutase 1, soluble (amyotrophic	NM_000454	163.31	29.19	-5.6	-134.13
211072_x_at	tubulin, alpha, ubiquitous /// tubulin, alpha, ubic	BC006481	252.51	44.86	-5.63	-207.65
211058_x_at	tubulin, alpha, ubiquitous /// tubulin, alpha, ubic	BC006379	268.97	46.88	-5.74	-222.09
213507_s_at	karyopherin (importin) beta 1	BG249565	131.54	21.82	-6.03	-109.72
1553569_at	cytochrome c oxidase II	NM_173705	415.68	65.33	-6.36	-350.34
208845_at	voltage-dependent anion channel 3	BC002456	138.76	21.76	-6.38	-117
233634_at	MARVEL domain containing 3	BC005052	139.72	21	-6.65	-118.72
224372_at	NADH dehydrogenase 4 /// NADH dehydrogens	AF253979	1040.52	139.47	-7.46	-901.06
237483_at	gb:AI990790 /DB_XREF=gi:5837671 /DB_XRE	AI990790	127.83	17.01	-7.52	-110.82
229353_s_at	nuclear ubiquitous casein kinase and cyclin-de	AW515443	355.79	45.25	-7.86	-310.54
212639_x_at	tubulin, alpha 3 /// tubulin, alpha, ubiquitous	AL581768	139.7	17.74	-7.87	-121.96
200633_at	ubiquitin B /// ubiquitin B	NM_018955	307.97	37.22	-8.27	-270.75
237180_at	gb:T97717 /DB_XREF=gi:747062 /DB_XREF=y	T97717	119.76	13.3	-9.01	-106.46
1553538_s_at	gb:NM_173704.1 /DB_XREF=gi:27754203 /GE	NM_173704	1153.33	125.55	-9.19	-1027.77
213646_x_at	tubulin, alpha, ubiquitous	BE300252	162.23	15.79	-10.28	-146.44
224616_at	dynein, cytoplasmic, light intermediate polype	BG110975	135.63	13.16	-10.31	-122.47
208687_x_at	heat shock 70kDa protein 8	AF352832	148.45	14.15	-10.49	-134.29
214288_s_at	proteasome (prosome, macropain) subunit, bet	W86293	215.8	18.16	-11.88	-197.64
201090_x_at	tubulin, alpha, ubiquitous	NM_006082	174.26	13.85	-12.58	-160.41
214328_s_at	heat shock 90kDa protein 1, alpha	R01140	287.88	22.44	-12.83	-265.43
201426_s_at	vimentin	AI922599	123.85	7.83	-15.82	-116.02
202043_s_at	spermine synthase	NM_004595	130.63	7.58	-17.23	-123.04
224373_s_at	NADH dehydrogenase 4 /// NADH dehydrogens	AF253979	649.36	34.93	-18.59	-614.42
222608_s_at	anillin, actin binding protein (scraps homolog, [	AK023208	157.68	8.26	-19.08	-149.42
202276_at	split hand/foot malformation (ectrodactyly) type	NM_006304	107.36	5.59	-19.22	-101.77
201024_x_at	eukaryotic translation initiation factor 5B	BG261322	114.41	5.62	-20.35	-108.78
1553551_s_at	NADH dehydrogenase 2	NM_173709	444.69	14.33	-31.04	-430.36
200873_s_at	chaperonin containing TCP1, subunit 8 (theta)	NM_006585	105.39	3.38	-31.2	-102.01
1554128_at	hypothetical protein MGC39724	BC029594	205.68	6.37	-32.3	-199.31
211600_at	NADH dehydrogenase 5 /// NADH dehydrogens	U20489	134.85	4.16	-32.39	-130.69
212857_x_at	activated RNA polymerase II transcription cofac	BG231551	113.47	2.08	-54.58	-111.39
200738_s_at	phosphoglycerate kinase 1	NM_000291	112.64	1.97	-57.18	-110.67

Table 3.9.5.2 (contd) Gene transcripts up/ down-regulated in HMEC CM compared to all cancer cell line CM (n=26). The list is sorted by fold change, no shading = up-regulated, grey shading = down-regulated. Criteria for inclusion on the list were a difference in expression of 100 and a fold change greater than 2.

The aim of Section 3.10 was to apply the methods developed for CM samples to clinical serum specimens and evaluate the usefulness of microarray technology for these samples. In Section 3.11 transcripts were randomly selected identified by microarray technology to be either:

- Present in all of the CM RNA samples and A in all of the cell RNA samples or
- Significantly differentially expressed between the sample groups

3.10 Microarray analysis of Breast cancer serum specimens

Ultimately the goal of our work is to be able to use the technology available in a clinical environment to help in either the detection or monitoring of disease. To this end, a study was designed to see if it was possible to apply the RNA extraction and microarray techniques developed to cell- free serum RNA extracted from breast cancer patients (see Section 2.7.10) for study design. Briefly, serum was taken from four patients recently diagnosed with breast cancer and before they had undergone any therapy treatments/ lumpectomy surgery (Pre serum). Serum specimens were then taken from the same four patients 2- 4 months after removal of the breast tumour (Post serum) and also from six normal volunteers (Normal serum). Gene expression profiles of the sample groups were evaluated and compared to the gene expression profiles obtained from tumour and matched normal tissue from the four breast cancer patients mentioned above.

3.10.1 QA of serum specimens

The QA results listed in Table 3.10.1.1 show that the serum and tissue samples, as separate groups, are acceptable to be compared to each other, however they cannot be normalised as one group as the scaling factors are too far apart.

	Pre Serum				Post Serum				Tumour Tissue				Matched Normal Tissue			
Patient #	3	7	11	14	3	7	11	14	3	7	11	14	3	7	11	14
Scaling Factor	18.11	15.967	20	20.6	21.09	20.047	20	22.9	0.831	1.255	0.757	2.3	1.2	1.33	2	2.74
Noise	1.31	1.59	1.4	1.28	1.1	1.27	1.22	1.15	1.91	1.7	1.95	1.24	1.7	1.31	1.37	1.78
Background	44.08	51.3	44	42.46	36.03	40.51	40	37.56	59.43	53.09	62.55	40.8	55.31	41.05	43	55.91
% Present	10.5	8	7.6	8.5	9.7	9	8.3	8.1	48.9	44.1	45.6	43.5	46.6	48.5	44	38.7
GAPDH 3'/ 5' *									2.67	3.43	4.37	4.83	3.34	4.1	6.01	13.34
GAPDH 3'/ M *									1.375	1.48	0.497	1.39	0.877	1.514	1.85	4

	Normal Serum					
Volunteer #	FN1	FN2	FN3	FN4	FN5	FN6
Scaling Factor	21.721	20.4	21.17	23.986	26.053	24.821
Noise	1.19	1.32	1.18	1.23	1.08	1.24
Background	38.41	42.12	37.03	40.85	37.04	39.92
% Present	8	8.8	9.2	7.3	7.2	6.5
GAPDH 3'/ 5' *						
GAPDH 3'/ M *						

* GAPDH level not recorded for sera samples as at least one of 5'/ M or 3' probe sets called A in all samples

Table 3.10.1.1 QA parameters measured for Affymetrix Array chips.

3.10.2 Expression of Two-Cycle Target Hybridisation and Labelling Controls

All hybridisation controls graphed in Fig.3.10.2.1 were called Present in all of the clinical samples. Figure 3.10.2.1 shows that the 5' and 3' probes are also expressed at similar levels across all 22 samples, which indicates that the hybridisation, washing and staining procedures were without incidence. The targets were detected according to the ratio of their concentration in the initial hybridisation cocktail i.e. *cre* was the most abundant followed by *bioD, bioC* and *bioB.*

The Poly- A controls were called Present in all of the clinical samples. Figure 3.10.2.2 shows that the expression of the 3' target probes (yellow lines) is higher in all samples than the expression of the 5' (dark blue lines) or middle probes (pink lines). This indicates that there was a 3' bias in the amplification procedure similar to that seen with the cell line CM samples (Fig. 3.8.3.2). At first glance it also looks as though the serum specimens 11 Pre, 7 Post and 2 Normal did not amplify successfully but it's actually that the amplification of samples 14 Pre and Normals 1, 3 and 6 were more efficient and so are skewing the results. This is possibly due to smaller amounts of endogenous mRNA in these samples leading to high amplification of the spiked- in Poly-A control RNAs.

The trend lines in Fig 3.10.2.3 suggest that the ratios of the Poly-A controls to each other match the ratios in the original pre-mixed control, but again like the cell line CM samples the actual relative expression of each Poly-A control to each other does not match the original pre-mixed ratios.

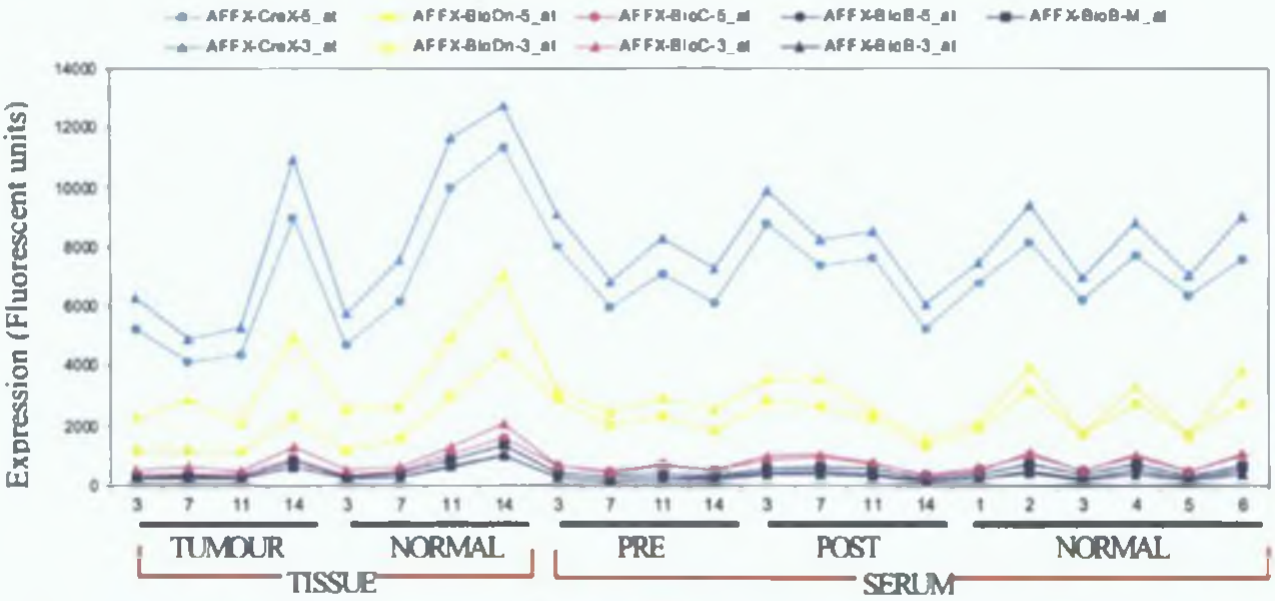


Figure 3.10.2.1 Expression levels of Hybridisation controls in serum specimens. Clinical samples are plotted on the X-axis with the expression level of each hybridisation control plotted on the Y-axis. Cre, BioD, BioC and BioB are the hybridisation spike-in controls and the suffix denotes the region targeted by the probe set i.e. 3_at targets the 3' end, M_at targets the middle and 5_at targets the 5' end.

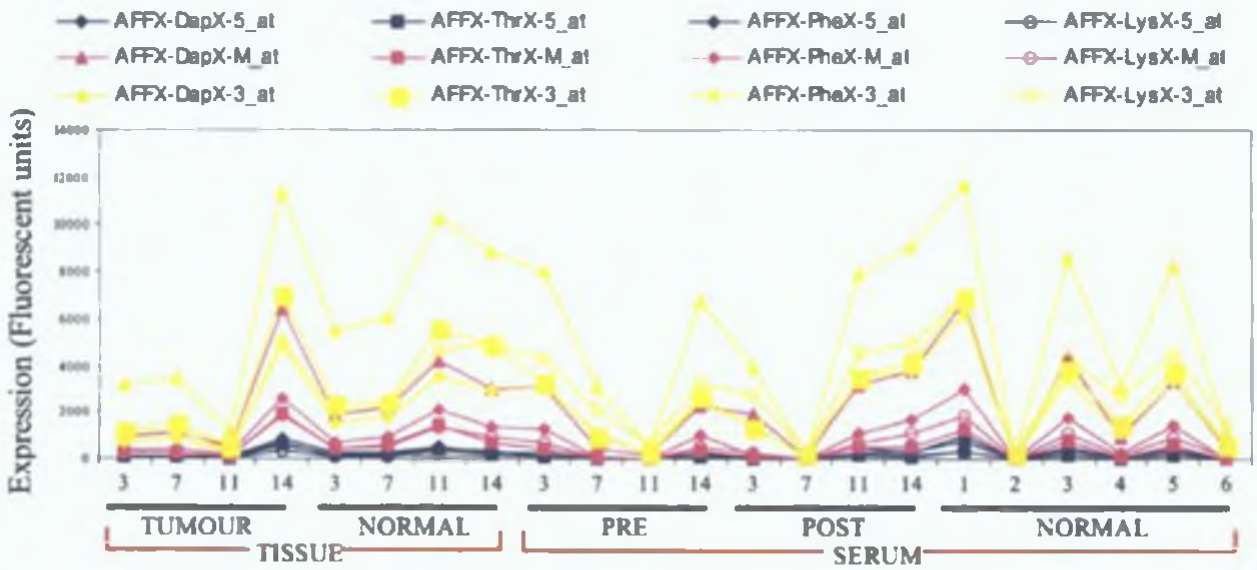


Figure 3.10.2.2 Expression of Poly A controls in serum specimens. Clinical samples are labelled on the X-axis with the expression level of each Poly A- control plotted on the Y-axis. Dap, Thr, Phe, Lys are the Poly-A spike in controls and the suffix denotes the region targeted by the probe set i.e. 3_at targets the 3' end, M targets the middle *etc.*

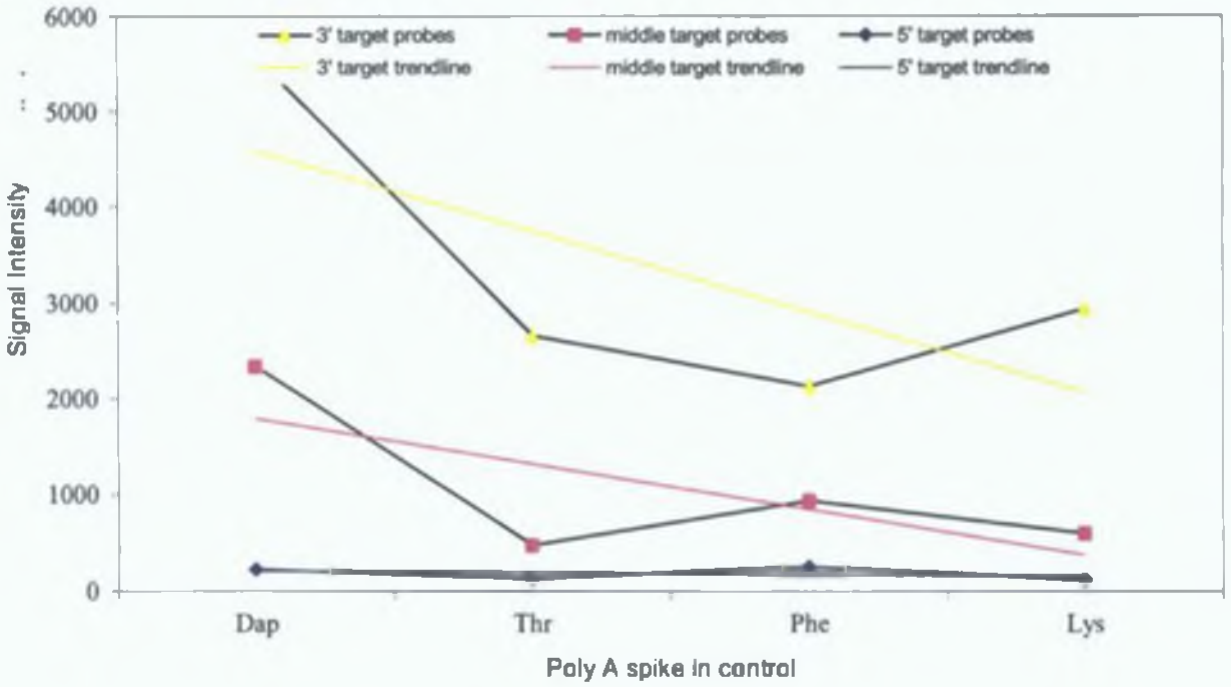


Figure 3.10.2.3 Average expression of Poly-A controls in the 22 clinical samples. Affymetrix Poly-A controls are plotted on the X-axis and the average expression for the controls plotted on the Y axis. The three linear lines (yellow, pink, and dark blue) are the trend lines for each data set and the black lines connect the actual relative expressions for the 5'Δ, middle■ ,and 3'◆ targeted probe sets.

3.10.3 Expression of "housekeeping" gene transcripts

The housekeeping gene transcripts GAPDH and β-actin are called Present in all of the tissue specimens (see Tables 3.10.3.1 and 3.10.3.2). Figs. 3.10.3.1 and 3.10.3.2 also reflect this status; there is a large difference in the expression of these two gene transcripts in the tissue samples compared to the serum specimens.

For β-actin, in the serum specimens, very similar to the CM samples, the 3' target probe had a Present call in all serum specimens (with the exception of one normal serum sample). Apart from this all other target regions with one exception were called A in all serum specimens. The graph in Fig. 3.10.3.1also shows that again, similar to the CM samples, expression levels of the 3' target probes are higher than the middle or 5' target probes in all clinical samples (expression levels in serum specimens more noticeable in the log2 transformed graph inset into Fig. 3.10.3.1).

For GAPDH, the Present/ Absent call assigned to the target probes also matched those achieved with the CM samples i.e. the 5' and 3' target regions were assigned as Present in the majority of samples and the middle target region was assigned as Absent.

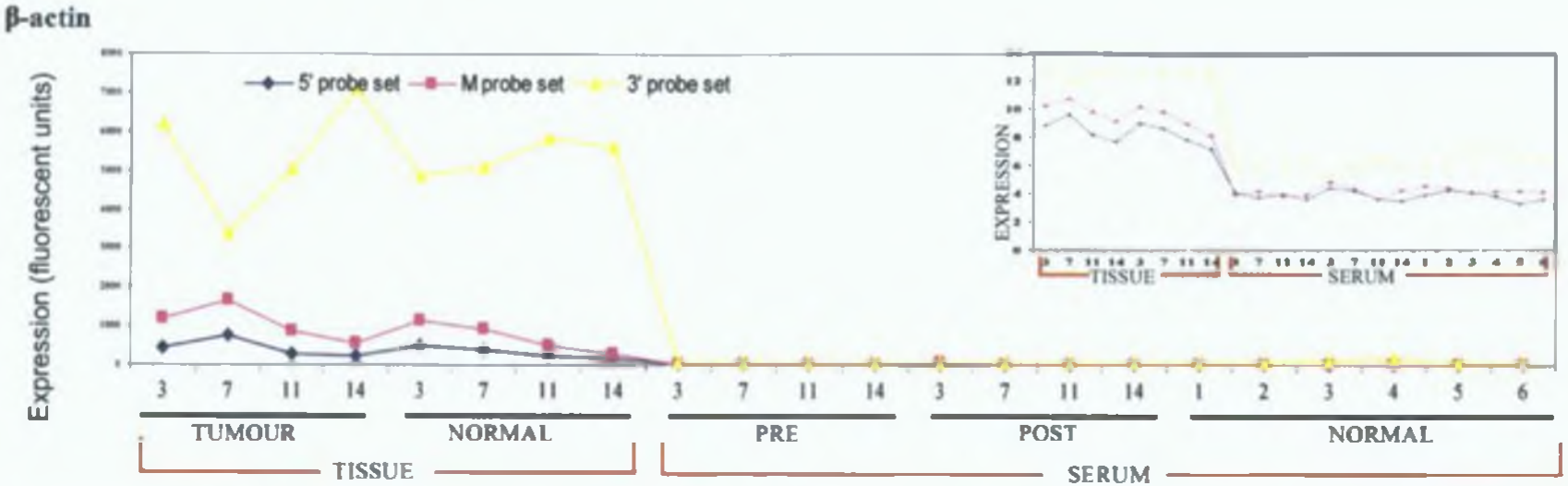


Figure 3.10.3.1 Expression levels of β-actin probe sets in clinical samples. Clinical samples are listed on the X-axis with the expression levels plotted on the Y- axis, the blue line = 5' targeted probes, pink line = middle region targeted probes, yellow = 3' targeted probes. Inset; expression data log2 transformed so large and small expression values can be more easily compared.

	TUMOUR				NORMAL				PRE SERUM				POST SERUM				NORMAL SERUM					
	3	7	11	14	3	7	11	14	3	7	11	14	3	7	11	14	1	2	3	4	5	6
5'	P	P	P	P	P	P	P	P	A	A	A	A	A	P	A	A	A	A	A	A	A	A
M	P	P	P	P	P	P	P	P	A	A	A	A	A	A	A	A	A	A	A	A	A	A
3'	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	A	P	P	P

Table 3.10.3.1 Presence/ Absent call assigned to β-actin probe sets in serum specimens.

GAPDH

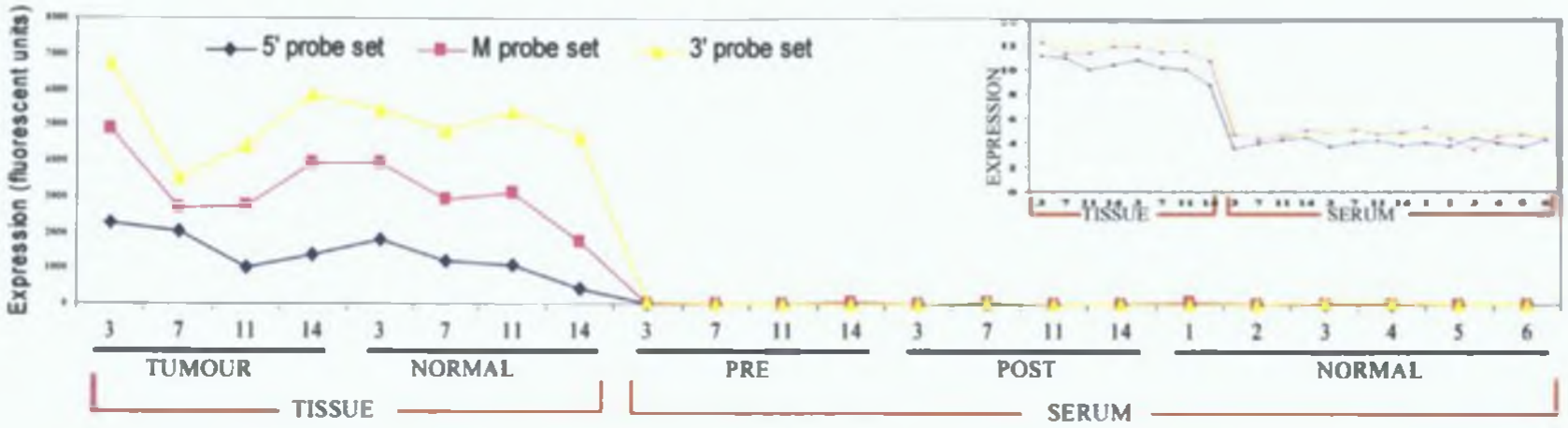


Figure 3.10.3.2 Expression levels of GAPDH probe sets in clinical samples. Clinical samples are listed on the X-axis with the expression levels plotted on the Y- axis, the blue line = 5' targeted probes, pink line = middle region targeted probes, yellow = 3' targeted probes. Inset; expression data log2 transformed so large and small expression values can be more easily compared.

	TUMOUR				NORMAL				PRE SERUM				POST SERUM				NORMAL SERUM					
	3	7	11	14	3	7	11	14	3	7	11	14	3	7	11	14	1	2	3	4	5	6
5'	P	P	P	P	P	P	P	P	P	P	P	P	P	A	P	P	P	P	P	P	P	A
M	P	P	P	P	P	P	P	P	A	A	A	A	A	A	A	A	A	A	A	A	A	A
3'	P	P	P	P	P	P	P	P	P	A	A	P	P	P	P	P	P	P	P	P	P	P

Table 3.10.3.2 Present/ Absent call assigned to GAPDH probe sets.

3.10.4 Condition tree and hierarchical clustering of Breast clinical samples

Hierarchical clustering of the breast tissue and serum specimens was carried out to see if the different sample groups clustered together. Figs. 3.10.4.1-3 show that whilst the tissue groups clustered separately from the serum specimens, the Pre, Post and Normal serum specimens did not cluster distinctly separate from each other. When hierarchical clustering was performed on the serum specimens on their own using dChip, two separate groups were found. In one all of the Post serum specimens and three Normal samples are grouped with one of the Pre serum specimens, in the other group, the remaining three Pre serum specimens are grouped with two Normal serum specimens (Fig. 3.10.4.3).

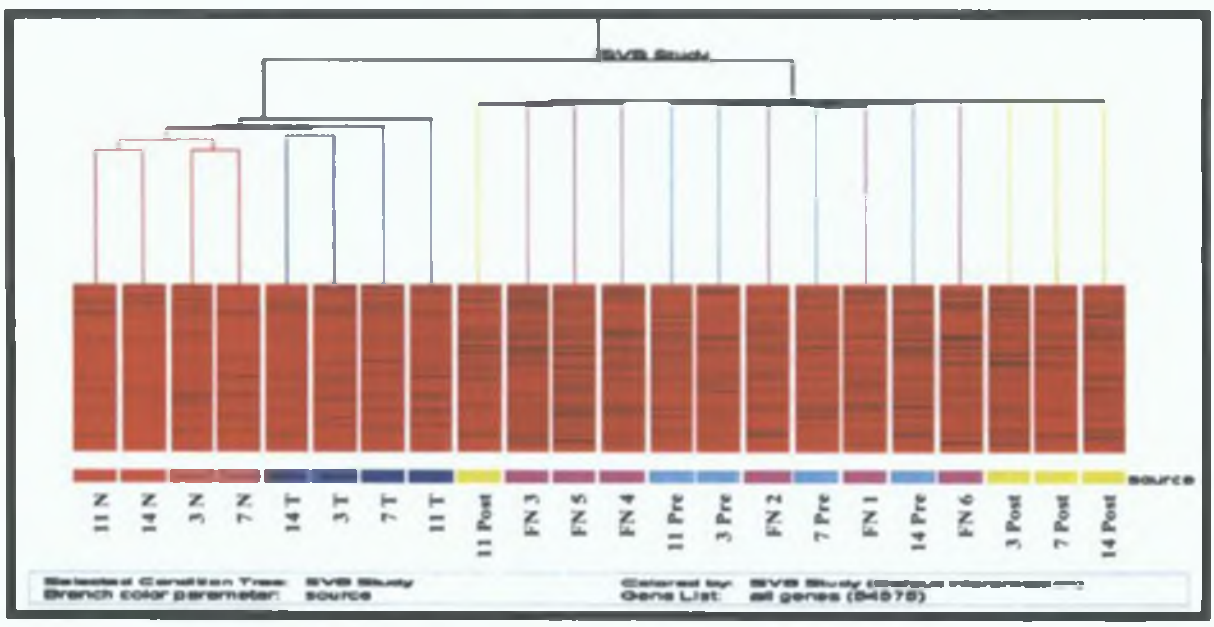


Figure 3.10.4.1 Hierarchical clustering of the clinical samples, tissue and serum using Genespring. The condition tree was performed using a Pearson correlation. The coloured boxes on the bottom of the figure indicate the sample type, red= normal tissue, blue= Tumour tissue, yellow= Post serum specimens, pink= Normal serum specimens and light blue= Pre serum specimens.

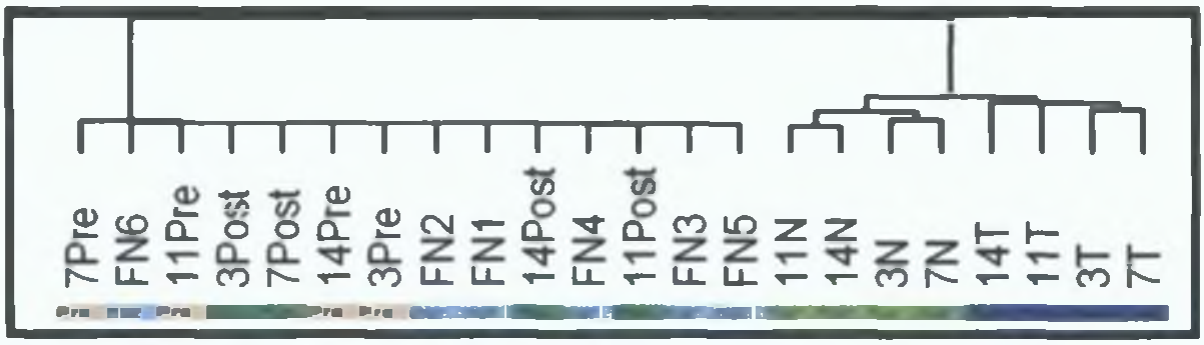


Figure 3.10.4.2 Hierarchical clustering of all the clinical samples using dChip, this clustering was carried out using a similarity measure. Gene transcripts that did not change in expression across all the CM samples were removed and the similarity of the samples was computed on the rest. Pre= Pre serum (pink box), Post= Post op (green box), FN= Female Normal serum (light blue box), N= Normal tissue (light green box), T= Tumour tissue (blue box). The numbers denote the patient/ volunteer number.

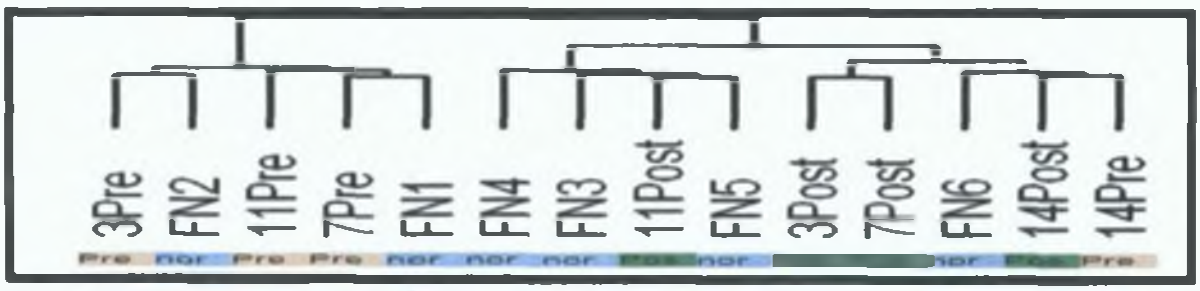


Figure 3.10.4.3 Hierarchical clustering of all the serum clinical samples using dChip, this clustering was carried out using a similarity measure. Gene transcripts that did not change in expression across all the CM samples were removed the similarity of the samples was computed on the rest. Pre= Pre serum (pink box), Post= Post serum (green box), FN= Female Normal serum (light blue box), the numbers denote the patient/ volunteer number.

It is possible that if the patient sample numbers were increased, along with the number of normal serum specimens, the separation of the serum specimens using dChip would become more apparent.

3.10.5 Probe sets called P or M in Pre serum and A in Post serum

Figures 3.10.5.1 and 3.10.5.2 show that of the 2501 gene transcripts called Present or Marginal in all of the Pre serum specimens, 51 of those were also called absent in all of the Post serum specimens. 6 of these 51 gene transcripts were also called Absent in all normal samples. It was decided to take a closer look at the expression values and patterns for these gene transcripts. Table 3.10.5.1 shows that the difference in expression values of the probe sets are not significant.

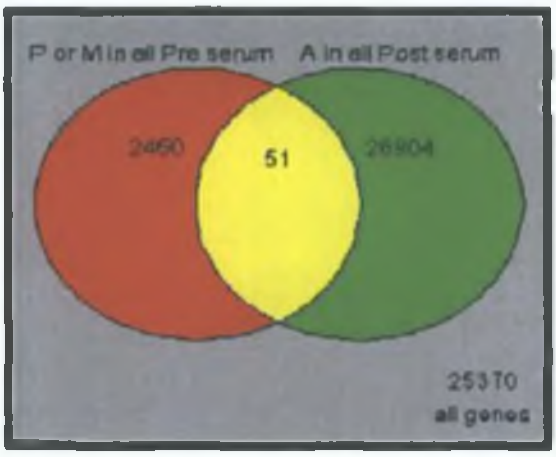


Figure 3.10.5.1 Venn diagram of probe sets called P or M in Pre serum and A in Post serum

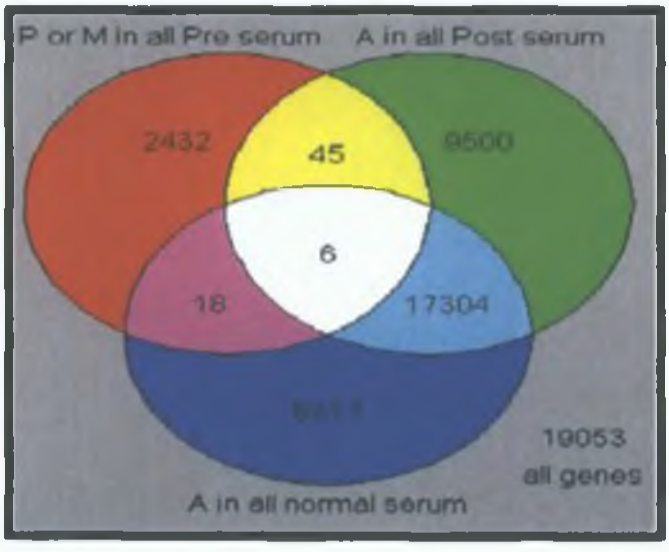


Figure 3.10.5.2 Venn diagram of probe sets called P or M in all Pre serum vs. A in all Post serum vs. A in all normal serum.

probe set	gene	11Pre		14Pre		3Pre		7Pre		11Post		14Post		3Post		7Post	
1556436_at	hypothetical protein LOC3	30.93	P	36.21	P	41.08	P	41.93	P	34.66	A	27.59	A	34.64	A	32.36	A
1558859_at	hypothetical protein LOC2	14.36	P	12.09	P	13.16	P	14.52	P	12.06	A	15.04	A	11.97	A	11.55	A
1559307_s_at	retinoblastoma-like 1 (p10	13.64	P	16.81	P	15.70	P	18.09	P	21.88	A	21.41	A	20.63	A	20.27	A
1561430_s_at	chromosome 3 open readi	32.42	P	27.78	P	27.49	P	40.65	P	28.64	A	25.61	A	18.72	A	26.41	A
1565934_at	Prematurely terminated m	16.65	P	19.24	P	14.98	P	14.84	P	15.88	A	12.16	A	11.55	A	14.55	A
1569608_x_at	ankyrin repeat domain 20A	26.35	P	27.38	P	30.21	P	27.93	P	21.84	A	24.01	A	33.10	A	32.24	A
1569917_at	Cyclin-dependent kinase E	23.97	P	25.66	P	23.23	P	27.86	P	31.15	A	26.44	A	24.55	A	22.84	A
201111_at	CSE1 chromosome segre	31.77	P	23.42	P	27.00	P	24.19	P	14.76	A	17.59	A	18.62	A	24.81	A
201339_s_at	sterol carrier protein 2	20.82	P	20.57	P	16.65	P	15.22	P	14.34	A	18.40	A	15.03	A	10.28	A
205338_s_at	dopachrome tautomerase	25.62	P	26.53	P	54.67	P	27.25	P	30.53	A	28.45	A	26.85	A	21.02	A
208068_x_at	chorionic somatomammot	38.51	P	35.41	P	33.86	P	40.16	P	39.19	A	33.77	A	32.95	A	36.31	A
209761_s_at	SP110 nuclear body prote	17.74	P	14.00	P	20.96	P	18.13	P	14.65	A	14.98	A	18.01	A	21.17	A
210616_s_at	SEC31-like 1 (S. cerevisia	37.50	P	36.91	P	36.57	P	38.46	P	31.97	A	36.67	A	28.70	A	36.86	A
213086_s_at	casein kinase 1, alpha 1	16.15	P	23.42	P	22.21	P	20.15	P	14.91	A	15.63	A	19.65	A	16.66	A
213825_at	oligodendrocyte lineage tr	41.40	P	58.24	P	44.39	P	45.41	P	40.16	A	47.97	A	38.07	A	43.05	A
214451_at	transcription factor AP-2 b	26.65	P	27.04	P	30.79	P	22.47	P	32.39	A	23.95	A	28.33	A	27.30	A
214989_x_at	Pleckstrin homology dom	6668.21	P	6751.03	P	6867.62	P	6844.02	P	6897.30	A	6735.69	A	6754.09	A	6836.73	A
215412_x_at	postmeiotic segregation in	23.42	P	23.11	P	24.45	P	20.91	P	21.32	A	16.64	A	17.97	A	16.59	A
216203_at	serine palmitoyltransferas	24.02	P	29.36	P	29.99	P	26.43	P	31.50	A	32.72	A	36.82	A	36.68	A
216349_at	Similar to bA215B13.2 (fu	98.04	P	81.70	P	77.79	P	69.19	P	42.51	A	62.53	A	74.64	A	68.81	A
216994_s_at	runt-related transcription fa	40.61	P	41.14	P	43.90	P	43.09	P	35.58	A	36.68	A	30.86	A	34.71	A
217416_x_at	gb:AK000918.1 /DB_XREF	19.41	P	16.69	P	19.68	P	16.93	P	14.89	A	14.14	A	16.23	A	18.03	A
219498_s_at	B-cell CLL/lymphoma 11A	41.04	P	41.72	P	35.61	P	41.55	P	34.10	A	38.70	A	35.39	A	38.27	A
222072_at	gb:AW450360 /DB_XREF=	76.48	P	74.39	P	62.88	P	81.86	P	51.01	A	72.14	A	70.01	A	67.83	A
223866_at	armadillo repeat containing	18.31	P	14.27	P	13.52	P	21.09	P	8.40	A	15.92	A	13.03	A	13.21	A
224494_x_at	dehydrogenase/reductase	47.00	P	44.82	P	41.43	P	39.75	P	40.86	A	35.04	A	37.96	A	32.59	A
226824_at	carboxypeptidase X (M14	54.34	P	67.08	P	66.05	P	57.43	P	63.83	A	77.43	A	56.26	A	61.00	A
229083_at	gb:AI672356 /DB_XREF=g	37.65	P	30.38	P	36.10	P	32.42	P	31.60	A	32.71	A	23.28	A	37.43	A
231130_at	gb:AA683602 /DB_XREF=	15.84	P	13.63	P	11.06	P	11.01	P	11.90	A	15.39	A	11.45	A	11.03	A
235772_at	gb:AI076941 /DB_XREF=g	28.58	P	21.76	P	24.32	P	25.64	P	14.70	A	20.87	A	22.65	A	26.87	A
237822_at	gb:AI962701 /DB_XREF=g	38.38	P	33.51	P	40.43	P	28.32	P	35.74	A	35.56	A	46.94	A	42.87	A
238646_at	gb:AA747756 /DB_XREF=	19.29	P	22.44	P	20.94	P	13.03	P	27.64	A	21.04	A	20.30	A	14.42	A
238825_at	acidic repeat containing	22.74	P	25.14	P	23.86	P	20.14	P	22.22	A	16.63	A	19.43	A	15.01	A
239018_at	Williams Beuren syndrom	49.02	P	53.14	P	51.32	P	60.95	P	44.71	A	43.86	A	47.93	A	48.42	A
239234_at	gb:AW452419 /DB_XREF=	66.62	P	61.65	P	63.56	P	52.88	P	37.42	A	46.64	A	57.50	A	45.41	A
239856_at	gb:AI701798 /DB_XREF=g	46.67	P	48.77	P	41.77	P	51.00	P	57.81	A	51.00	A	40.85	A	48.23	A
240757_at	gb:H49240 /DB_XREF=gi:	20.85	P	24.72	P	21.54	P	27.52	P	25.93	A	25.57	A	20.53	A	27.23	A
241294_at	gb:AI939422 /DB_XREF=g	28.60	P	34.76	P	26.05	P	25.82	P	25.86	A	31.48	A	26.33	A	26.29	A
241411_at	gb:AA449838 /DB_XREF=	60.46	P	75.99	P	61.18	P	57.25	P	72.86	A	78.20	A	72.33	A	61.35	A
242283_at	hypothetical protein MGC2	14.61	P	13.42	P	13.05	P	14.43	P	10.03	A	14.61	A	12.87	A	12.35	A
242303_at	gb:AI271427 /DB_XREF=g	60.58	P	61.75	P	61.18	P	52.50	P	62.38	A	71.94	A	66.55	A	79.84	A
242326_at	collagen, type XXII, alpha	52.36	P	44.32	P	31.11	P	43.57	P	34.22	A	40.31	A	37.71	A	40.03	A
243303_at	Enoyl Coenzyme A hydrat	20.50	P	16.62	P	16.68	P	15.77	P	11.29	A	11.75	A	10.62	A	15.58	A
244567_at	gb:BG165613 /DB_XREF=	7.92	P	7.63	P	8.18	P	5.31	P	3.91	A	7.27	A	7.55	A	7.87	A
1554172_a_at	zinc finger protein 261	36.32	P	34.47	P	28.42	M	30.84	P	31.47	A	29.06	A	34.20	A	32.90	A
215219_at	chromosome 21 open reac	30.43	P	33.46	P	35.01	M	32.24	P	35.08	A	33.53	A	26.07	A	25.93	A
216748_at	interferon-inducible protein	38.70	M	37.57	P	43.61	P	32.90	P	49.04	A	41.33	A	29.17	A	34.73	A
221311_x_at	hypothetical protein dJ122	7.50	M	10.29	P	10.85	P	8.18	P	14.85	A	7.97	A	9.18	A	8.10	A
231031_at	gb:AI761573 /DB_XREF=g	32.85	M	31.66	P	58.74	P	41.13	P	25.48	A	26.19	A	36.70	A	30.15	A
234099_at	Laminin, alpha 2 (merosin	29.94	P	28.39	P	27.73	M	26.63	P	35.25	A	23.53	A	25.09	A	23.14	A
235927_at	gb:BE350122 /DB_XREF=	31.39	P	43.75	P	39.79	M	33.45	P	28.64	A	32.60	A	38.83	A	31.08	A

Table 3.10.5.1 Expression values for probe sets called P or M in pre serum and A in post serum. Pre = Pre serum sample, Post = Post serum sample, P = Present, A = Absent.

Both Table 3.10.5.1 and Fig. 3.10.5.3 show that despite the 51 gene transcripts being assigned as Present or Marginal in the Pre serum specimens and Absent in the Post serum and Normal serum specimens, the actual difference in expression is negligible. To find gene transcripts significantly differentially regulated dChip software was used.

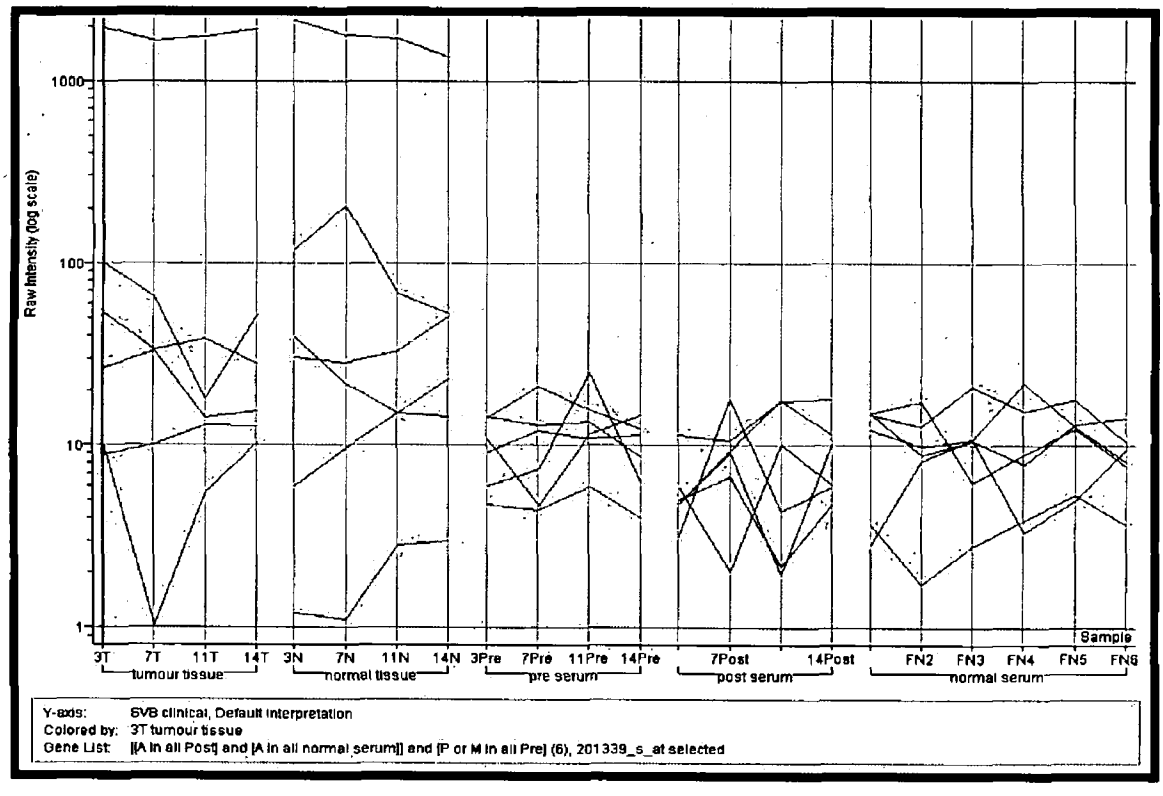


Figure 3.10.5.3 Expression profile of probe sets called P or M in Pre serum and A in Post and Normal serum specimens. Sample groups are on the X-axis (from left) tumour tissue, normal tissue, pre serum, post serum and normal serum. The expression in plotted on the Y-axis (log scale). The expression for one probe set, 201339_s_at, is highlighted in black showing similar levels of expression in tumour and normal tissue and in the different serum specimens.

3.10.6 Probe sets significantly changed in Pre vs. Post serum

A gene list was generated examining the number of gene transcripts significantly changed between the Pre and Post serum specimens. The criteria used for classification as significant was that the gene transcripts had to have a fold change >1.2, a difference in expression >20 and a paired p-value <0.05. The paired p-value refers to the significance of changes between individual pairs (e.g. Pre and Post serum specimens from the same patient) and also the p- value of the group pairings e.g. all Pre vs. all Post samples.

The results presented in Tables 3.10.6.1 (a), (b) and (c) show that of the 114 gene transcripts deemed to be significantly changed in the Post serum specimens compared to the Pre serum specimens, only 5 are down-regulated in the Post serum specimens compared to the Pre serum specimens. This seems like an illogical result, we would have expected to see more mRNA transcripts either in the Pre serum specimens or at least expressed at higher levels in the Pre serum specimens than in the Post serum specimens.

Analysis of the gene expression profile of the matched normal and tumour tissue samples however shows a similar trend. Of the 3,535 gene transcripts differentially expressed between the tumours and matched normal tissues (selection criteria; fold change>1.2, expression difference>20 and p-value<0.05), 80% (2,765) are down-regulated in the tumour tissue compared to the matched normal.

To assure that the results presented in Section 3.10.6 were not due to sample mis- labelling, the Pre serum specimens were compared to the normal serum specimens (which were isolated at a different time to the Pre and Post serum specimens, and obviously, are samples originating from different people).

Of the 542 gene transcripts that were differentially expressed between the Pre serum specimens and the Normal serum specimens using the same criteria as before i.e. difference in expression and fold change greater than 20, respectively and a p-value less than 0.05(paired p-value not used) 97% (525 gene transcripts) are up-regulated in the Normal serum compared to the Pre serum specimens, with only 3% (17 gene transcripts) down-regulated in the Normal serum compared to the Pre serum specimens (see Tables 3.10.6 (a) and (b)).

probe set	gene	Accession	fold change	diff of means	paired P value
217839_at	TRK-fused gene	NM_006070	-5.8	-42.26	0.018743
1554691_a_at	protein kinase C and casein kinase sut	BC008037	-2.08	-25.34	0.027104
204166_at	KIAA0963	NM_014963	-1.7	-200.05	0.02681
201481_s_at	phosphorylase, glycogen; brain	NM_002862	-1.64	-232.61	0.049552
1554977_at	Transcribed locus, moderately similar t	BC036837	-1.48	-128.11	0.010071

Table 3.10.6.1 (a) Probe sets significantly down-regulated in the Post serum specimens compared to the Pre serum specimens. The list is sorted by fold change the grey shading means they are down-regulated in the Post serum compared to the Pre serum. Criteria for selection as significant; difference in expression >20, fold change >1.2 and p-value <0.05.

probe set	gene	Accession	fold change	diff of means	paired P value
1562062_at	gb:BM041211 /DB_XREF=gi:16770478	BM041211	1.23	125.76	0.03516
232481_s_at	SLIT and NTRK-like family, member 6	AL137517	1.24	21.91	0.03572
207953_at	gb:AF010144.1 /DB_XREF=gi:3002526	AF010144	1.28	65.81	0.026971
233270_x_at	Pericentriolar material 1	AU146939	1.29	232.72	0.022563
214722_at	similar to NOTCH2 protein	AW516297	1.31	101.41	0.034593
1560199_x_at	Zinc finger, SWIM domain containing 6	AU120130	1.32	75.91	0.013756
234251_at	gb:AL157423.1 /DB_XREF=gi:7018452	AL157423	1.35	24.1	0.031673
230182_at	gb:AI951640 /DB_XREF=gi:5743950 /D	AI951640	1.39	49.74	0.043457
235601_at	gb:AA907029 /DB_XREF=gi:3042489 /I	AA907029	1.39	53.84	0.015268
1558426_x_at	chromosome 7 open reading frame 19	BC016797	1.4	91.2	0.022997
1563088_a_at	hypothetical protein LOC284837	BC034348	1.4	25.37	0.022096
1555268_a_at	glutamate receptor, ionotropic, delta 1	BC039263	1.41	24.29	0.022917
206293_at	sulfotransferase family, cytosolic, 2A, c	U08024	1.44	32.31	0.044106
230731_x_at	Zinc finger, DHHC domain containing 8	AI453548	1.45	32.58	0.036652
1563077_at	Homo sapiens, clone IMAGE:4042988,	BC014346	1.45	24.56	0.011801
221155_x_at	gb:NM_018603.1 /DB_XREF=gi:89240	NM_018603	1.47	90.56	0.03594
1558438_a_at	Rearranged immunoglobulin VH-CE in	S55277	1.47	29.24	0.010116
230582_at	gb:BE676408 /DB_XREF=gi:10036949	BE676408	1.48	24.72	0.026937
1563571_at	hypothetical protein LOC285463	AK092548	1.48	24.55	0.048144
232034_at	hypothetical protein LOC203274	AL117607	1.49	44.5	0.032216
210397_at	defensin, beta 1	U73945	1.5	23.67	0.009022
232511_at	RAN binding protein 2-like 1	AK022838	1.5	28.31	0.040478
1562861_at	Homo sapiens, clone IMAGE:5195119,	BC041456	1.5	21.82	0.049088
1555881_s_at	leucine zipper, putative tumor suppress	AK097997	1.51	40.47	0.009687
1570329_at	gb:BC014497.1 /DB_XREF=gi:1793960	BC014497	1.51	20.67	0.044096
232834_at	Transcribed locus, moderately similar t	AU146764	1.52	32.19	0.026994
233380_s_at	RUN and FYVE domain containing 1	AA429145	1.52	40.75	0.03046
1553614_a_at	Olfactory receptor, family 7, subfamily I	NM_173604	1.52	26.65	0.02064
1567622_at	gb:AJ431609.1 /DB_XREF=gi:18873710	AJ431609	1.52	77.73	0.010042
218143_s_at	secretory carrier membrane protein 2	NM_005697	1.53	32.21	0.03529
220633_s_at	HP1-BP74	NM_016287	1.53	46.75	0.005848
1565829_at	KIAA1731 protein	BU619319	1.55	25.82	0.049403
217888_s_at	ADP-ribosylation factor GTPase activat	NM_018209	1.56	28.52	0.017196
236794_at	gb:AW629436 /DB_XREF=gi:7376226 /	AW629436	1.56	34.09	0.044758
233319_x_at	Phosphatase and actin regulator 4	AK023907	1.57	110.19	0.044141
211416_x_at	gamma-glutamyltransferase-like activity	L20492	1.58	22.33	0.048573
214032_at	zeta-chain (TCR) associated protein kir	AI817942	1.58	24.86	0.031465
239097_at	gb:AW296997 /DB_XREF=gi:6703633 /	AW296997	1.58	25.75	0.044805
207628_s_at	Williams Beuren syndrome chromosom	NM_017528	1.59	30.63	0.008715
212109_at	chromosome 16 open reading frame 34	AI590869	1.59	59.49	0.015363
214122_at	gb:AA086229 /DB_XREF=gi:1628852 /I	AA086229	1.6	22.6	0.036004
231306_at	lysozyme-like 4	AI989932	1.6	36.82	0.014697
1565818_s_at	zinc finger protein, subfamily 1A, 1 (Ika	BF794111	1.61	22.3	0.035552
205367_at	adaptor protein with pleckstrin homolog	NM_020979	1.62	167.03	0.02477
212430_at	RNA-binding region (RNP1, RRM) cont	AL109955	1.62	31.67	0.007406
225394_s_at	MADP-1 protein	AI146850	1.62	70.99	0.024932
211058_x_at	tubulin, alpha, ubiquitous /// tubulin, alp	BC006379	1.63	22.7	0.037428
220696_at	PRO0478 protein	NM_014129	1.63	84.25	0.009298

Table 3.10.6.1 (b) Probe sets significantly up-regulated in the Post serum compared to the Pre serum. The list is sorted by fold change. Criteria for selection as significant; difference in expression >20, fold change >1.2 and a paired p-value <0.05.

probe set	gene	Accession	fold change	diff of means	paired P value
233895_at	Ankyrin repeat domain 24	AC005578	1.63	62.38	0.042981
1566763_at	CDNA clone IMAGE:5277680, partial c	BC034915	1.63	22.44	0.022409
232516_x_at	YY1 associated protein	AU150385	1.64	30.06	0.029117
233912_x_at	ELMO domain containing 2	AK021525	1.64	46.02	0.031677
1556248_at	Myosin VIIA and Rab interacting protein	BC031275	1.64	51.3	0.016559
1557996_at	gb:AK091784.1 /DB_XREF=gi:2175023	AK091784	1.64	64.38	0.049877
219013_at	UDP-N-acetyl-alpha-D-galactosamine:p	NM_022087	1.65	24.28	0.013128
223871_x_at	inhibitor of growth family, member 5	BC005370	1.65	66.27	0.032576
1562946_at	Transcribed locus, moderately similar t	BC007626	1.65	24.24	0.047108
242377_x_at	gb:AI051976 /DB_XREF=gi:3307967 /D	AI051976	1.66	90.56	0.041761
1570210_x_at	gb:BC022346.1 /DB_XREF=gi:1849062	BC022346	1.66	144.93	0.024853
220828_s_at	hypothetical protein FLJ11292	NM_018382	1.67	54.73	0.014091
1569380_a_at	gb:BC015447.1 /DB_XREF=gi:2195539	BC015447	1.67	29.63	0.01552
210981_s_at	G protein-coupled receptor kinase 6	AF040751	1.69	20.9	0.020073
231638_at	hypothetical protein MGC52282	AA400057	1.69	20.22	0.048351
232731_x_at	Transcribed locus, moderately similar t	BC001607	1.69	31.53	0.019219
1554487_a_at	cAMP responsive element binding prote	BC008394	1.69	22.77	0.04341
225986_x_at	cleavage and polyadenylation specific f	AB037788	1.7	40.69	0.023977
1568702_a_at	hypothetical protein FLJ11848	AI131457	1.7	66.45	0.038734
1569345_at	gb:BC026007.1 /DB_XREF=gi:2213567	BC026007	1.7	24.51	0.025699
207131_x_at	gamma-glutamyltransferase 1	NM_013430	1.71	20.05	0.015017
231602_at	gb:AI809150 /DB_XREF=gi:5395716 /D	AI809150	1.71	39.43	0.019132
216924_s_at	dopamine receptor D2	S62137	1.72	50.19	0.00321
220352_x_at	hypothetical protein MGC4278	NM_024305	1.72	90.08	0.040646
233513_at	Transcribed locus, highly similar to XP	AU152161	1.73	45.63	0.014733
1558792_x_at	Adaptor-related protein complex 2, alph	AK090661	1.73	72.36	0.030287
215861_at	hypothetical protein FLJ14031	AK024093	1.74	111.97	0.023972
220725_x_at	Dynein, axonemal, heavy polypeptide 3	NM_025095	1.74	40.39	0.010836
232145_at	Hypothetical LOC388969	AU147777	1.74	25.08	0.041851
222339_x_at	gb:AI054381 /DB_XREF=gi:3322168 /D	AI054381	1.75	54.91	0.009331
244193_at	hypothetical protein FLJ13236	BF981786	1.75	45.96	0.023815
243178_at	gb:AW969703 /DB_XREF=gi:8159547 /	AW969703	1.76	74.58	0.043607
236499_at	Hypothetical protein FLJ31031 /// LOC4	AI911410	1.79	28.76	0.019165
1566716_at	gb:AL050033.1 /DB_XREF=gi:4884273	AL050033	1.79	39.63	0.02707
228576_s_at	gb:BE219549 /DB_XREF=gi:8906867 /(	BE219549	1.81	92.88	0.005963
32837_at	1-acylglycerol-3-phosphate O-acyltrans	U56418	1.82	20.88	0.01468
1554183_s_at	tripartite motif-containing 50C	BC033871	1.85	24.02	0.006475
215067_x_at	peroxiredoxin 2	AU147942	1.86	75.01	0.024907
202624_s_at	calcineurin binding protein 1	NM_012295	1.87	24.48	0.016945
207914_x_at	gb:NM_001989.1 /DB_XREF=gi:450361	NM_001989	1.87	73	0.003911
1562885_at	Homo sapiens, clone IMAGE:5750288,	BC042084	1.89	29.93	0.040361
233362_at	zinc finger protein 341	AL523697	1.93	22.27	0.021026
1562771_at	Transcribed locus, moderately similar t	BC042087	1.94	24.23	0.020998
1558569_at	Mastermind-like 2 (Drosophila)	AL832308	1.96	73.69	0.042645
237561_x_at	gb:AV736334 /DB_XREF=gi:10853915	AV736334	2	25.87	0.049274
1556149_at	armadillo repeat gene deletes in velocai	AW139431	2	20.62	0.015864
1565661_x_at	Fucosyltransferase 6 (alpha (1,3) fucos	BC040472	2	23.57	0.029249
1562487_at	Transcribed locus, moderately similar t	BC033379	2.02	21.42	0.0326
1558407_at	Likely ortholog of mouse common-site l	AK095957	2.03	36.68	0.017047
1558048_x_at	gb:BG389789 /DB_XREF=gi:13283225	BG389789	2.07	175.57	0.026641
232570_s_at	a disintegrin and metalloproteinase don	AL356755	2.09	67.09	0.021287
239826_at	gb:AI357143 /DB_XREF=gi:4108764 /D	AI357143	2.11	23.09	0.035571
244544_at	gb:AA812915 /DB_XREF=gi:2882979 /(	AA812915	2.16	28.54	0.00499
242369_x_at	gb:AI561070 /DB_XREF=gi:4511411 /D	AI561070	2.19	110.94	0.009483
211627_x_at	Estrogen receptor 1 /// Estrogen recept	M69297	2.2	76.83	0.015976
220232_at	stearoyl-CoA desaturase 4	NM_024906	2.25	33.92	0.038158
1558275_at	Mitochondria-associated protein involve	BC006120	2.26	42.57	0.044397
229650_s_at	hypothetical protein MGC2747	BG538931	2.27	26.45	0.004752
217430_x_at	gb:Y15916.1 /DB_XREF=gi:3288488 /F	Y15916	2.31	27.34	0.043388
232640_at	COMM domain containing 5	AK023070	2.88	22.61	0.025067
239333_x_at	Transcribed locus, weakly similar to NF	BF525633	3.06	28.93	0.027716

Table 3.10.6.1 (b contd) Probe sets significantly up-regulated in the Post serum compared to the Pre serum. The list is sorted by fold change. Criteria for selection as significant; difference in expression >20, fold change >1.2 and a paired p-value <0.05.

probe set	gene	Accession	fold change	difference of means	P value
202449_s_at	retinoid X receptor, alpha	NM_002957	-7.65	-31.36	0.04394
217839_at	TRK-fused gene	NM_006070	-6.25	-75.38	0.003377
1554741_s_at	fibroblast growth factor 7 (keratinocyte	AF523265	-2.16	-27.61	0.002145
201024_x_at	eukaryotic translation initiation factor	BG261322	-2.14	-61.33	0.013979
227952_at	Zinc finger protein 595	AI580142	-1.97	-96.45	0.004769
204166_at	KIAA0963	NM_014963	-1.89	-401.48	0.028461
1553186_x_at	RAS and EF hand domain containing	NM_152573	-1.76	-1757.33	0.041952
1553185_at	RAS and EF hand domain containing	NM_152573	-1.75	-1772.65	0.046216
221260_s_at	chromosome 12 open reading frame 2	NM_030809	-1.69	-21.42	0.045999
1562226_at	hypothetical protein FLJ14712	AK075525	-1.68	-23.8	0.00788
1554977_at	Transcribed locus, moderately similar	BC036837	-1.45	-214.87	0.000899
1561539_at	Homo sapiens, clone IMAGE:5303541	BC041984	-1.44	-20.13	0.021455
213229_at	Dicer1, Dcr-1 homolog (Drosophila)	BF590131	-1.43	-34.48	0.049771
1555373_at	gb:AF304443.1 /DB_XREF=gi:16588	AF304443	-1.41	-1572.24	0.022969
224549_x_at	gb:AF194537.1 /DB_XREF=gi:11037	AF194537	-1.38	-2049.44	0.02895
224254_x_at	gb:AF116695.1 /DB_XREF=gi:79598	AF116695	-1.34	-586.77	0.048843
234723_x_at	CDNA FLJ21228 fis, clone COL0073	AK024881	-1.23	-889.69	0.022233

Table 3.10.6.2 (a) 17 gene transcripts were down-regulated in the Normal serum compared to the Pre serum specimens. The list is sorted by fold change and the grey shading indicates that the samples were down-regulated in the Normal serum compared to the Pre serum. Criteria for selection as significant; difference in expression >20, fold change >1.2 and paired p-value <0.05.

probe set	gene	Accession	fold change	difference of means	P value
231278_at	Hypothetical LOC338731	BF063966	5.11	32.12	0.016627
215926_x_at	small nuclear RNA activating complex	AK023513	4.46	25.58	0.025904
218684_at	leucine rich repeat containing 5	NM_018103	4.22	20.23	0.023683
212844_at	KIAA0179	AI400490	4.08	20.91	0.023571
209208_at	mannose-P-dolichol utilization defect	AF059752	3.41	20.3	0.011084
225767_at	Hypothetical protein LOC284801	AI825833	3.19	93.16	0.015115
1560741_at	Ubiquitin protein ligase E3A (human p	AL832250	3.18	41.16	0.009267
239060_at	gb:AI222295 /DB_XREF=gi:3804498 /	AI222295	3.11	35.37	0.027211
1562553_at	Homo sapiens, clone IMAGE:5287123	BC042964	3.01	29.69	0.023424
234084_x_at	Transmembrane protein induced by tu	AU147104	2.94	27.01	0.01986
244156_at	gb:AW296286 /DB_XREF=gi:6702932	AW296286	2.83	28.71	0.018482
242982_x_at	integrin, beta 8	AW131039	2.64	23.47	0.013663
227192_at	similar to lymphocyte antigen 6 comp	BF060707	2.58	30.31	0.021446
226598_s_at	GTP binding protein 5 (putative)	AK001603	2.45	34	0.030836
1563426_a_at	Homo sapiens, clone IMAGE:4177475	BC025792	2.44	55.04	0.008232
216053_x_at	hypothetical protein FLJ38374	AI143919	2.42	30.64	0.004069
216692_at	Zinc finger protein 337	AL137428	2.42	23.27	0.044249
201541_s_at	zinc finger, HIT domain containing 1	NM_006349	2.41	42.75	0.004914
205300_s_at	U11/U12 snRNP 35K	NM_022717	2.41	30.86	0.049484
219091_s_at	multimerin 2	NM_024756	2.41	22.59	0.021393
224067_at	gb:AF132206.1 /DB_XREF=gi:114935	AF132206	2.41	28.65	0.043147
211074_at	gb:AF000381.1 /DB_XREF=gi:256519	AF000381	2.4	57.41	0.034717
64440_at	interleukin 17 receptor C	AI560217	2.38	56.44	0.007526
232966_at	Similar to LPIN3	AL031667	2.37	34.48	0.012229
1557350_at	Ras-GTPase-activating protein SH3-d	AA026297	2.37	29.02	0.004281
233009_at	gb:AK024333.1 /DB_XREF=gi:10436	AK024333	2.36	21.99	0.009434
1564435_a_at	keratin protein K6irs	AK093060	2.36	35.19	0.009615
244469_at	gb:AA534603 /DB_XREF=gi:2278856	AA534603	2.34	25.13	0.036282
238405_at	gb:AI792696 /DB_XREF=gi:5340612 /	AI792896	2.31	35.19	0.039556
203398_s_at	UDP-N-acetyl-alpha-D-galactosamine:	NM_004482	2.29	24.39	0.021979
216891_at	gb:U00956.1 /DB_XREF=gi:405058 /F	U00956	2.29	33.28	0.004795
221419_s_at	gb:NM_013307.1 /DB_XREF=gi:7019	NM_013307	2.29	70	0.006794
1563108_at	KIAA1904 protein	BC033336	2.29	33.25	0.003431
240988_x_at	gb:AA207241 /DB_XREF=gi:1802734	AA207241	2.27	51.47	0.028504
241396_at	gb:AL572553 /DB_XREF=gi:1293093	AL572553	2.25	20.35	0.010682
201255_x_at	HLA-B associated transcript 3	NM_004639	2.24	21	0.016131
244350_at	gb:N33403 /DB_XREF=gi:1153802 /D	N33403	2.23	68.13	0.011083
1558275_at	Mitochondria-associated protein involv	BC006120	2.22	72.66	0.042835
236801_at	gb:T61953 /DB_XREF=gi:665196 /DB_	T61953	2.19	114.86	0.008494
237768_x_at	gb:AA825925 /DB_XREF=gi:2899237	AA825925	2.19	32.38	0.012784
214928_at	KIAA0657 protein	AI915513	2.18	67.62	0.009657
225059_at	angiotensin II receptor-associated pro	BE875567	2.17	82.25	0.011135
223422_s_at	Rho GTPase activating protein 24	AI743534	2.15	25.91	0.034577
219534_x_at	cyclin-dependent kinase inhibitor 1C (	NM_000076	2.14	41.11	0.002954
1566848_x_at	Homo sapiens, clone IMAGE:4015585	BC034050	2.13	285.96	0.004829
234439_at	gb:AK001097.1 /DB_XREF=gi:702214	AK001097	2.12	33.04	0.005486
1566778_at	gb:BC043203.1 /DB_XREF=gi:276932	BC043203	2.12	20.16	0.033337
201276_at	RAB5B, member RAS oncogene fami	AF267863	2.1	35.62	0.031321
227513_s_at	gb:AW027170 /DB_XREF=gi:5885926	AW027170	2.09	23.49	0.047504
221193_s_at	zinc finger, CCHC domain containing	NM_017665	2.08	33.6	0.018798
241767_at	gb:AI184581 /DB_XREF=gi:3735219 /	AI184581	2.08	29.17	0.046266
238845_at	solute carrier family 30 (zinc transport	AA485220	2.07	28.96	0.013832
217818_s_at	actin related protein 2/3 complex, sub	NM_005718	2.06	23.05	0.017293
232626_at	CDNA FLJ14143 fis, clone MAMMA1	AK024205	2.05	36.17	0.029293
209038_s_at	EH-domain containing 1	AL579035	2.04	84.1	0.017186
243263_at	hypothetical protein FLJ39421	AW008416	2.04	69.03	0.019604
224460_s_at	chromosome 14 open reading frame 1	BC006117	2.03	31.52	0.011831
231878_at	Major vault protein	AK026112	2.03	34.13	0.001875
205910_s_at	carboxyl ester lipase (bile salt-stimula	NM_001807	2.02	34.92	0.010456
219088_s_at	zinc finger protein 576	AA639585	2.02	20.14	0.012184
241851_x_at	CDNA FLJ45580 fis, clone BRTHA301	BE676261	2.02	30.83	0.001237
234826_at	Hypothetical gene supported by BC01	AL137596	2.01	21.53	0.043431
243325_at	gb:AV722006 /DB_XREF=gi:1082406	AV722006	2.01	21.75	0.021131
220339_s_at	tryptase gamma 1	NM_012467	2	23.5	0.008567
234072_at	sema domain, immunoglobulin domai	AU147993	2	60.71	0.004691

Table 3.10.6.3 (b) 525 gene transcripts were up-regulated in the Normal serum specimens compare to the Pre serum specimens. The gene transcripts were sorted by fold change and the 64 gene transcripts with a fold change >2 are shown above. Criteria for selection as significant; difference in expression >20, fold change >1.2 and paired p-value <0.05.

3.10.7 Study of randomly changing gene transcripts

The analysis of probe sets called P or M in Pre serum and A in Post serum yielded no leads of potential biomarkers as the expression difference between probe sets called Present in the Pre serum and Absent in the Post serum was negligible and not significant. The analysis of the gene transcripts determined to be significantly changed yielded results that seemed illogical, in so far as more transcripts were upregulated in the serum months after lumpectomy, than in the serum before the tumour was removed. In addition to these results, the hierarchical clustering did not show any tight clustering of sample groups. To test whether the gene transcripts found to be significantly different were related to a cancer-state and not simply a change in expression that has occurred by chance, various combinations of samples were compared against each other to see how many changes are found when a random selection of samples is tested against each other.

dChip software has a false discovery rate (FDR) application whereby, it the randomly mixes the samples using 50 permutations to find out the FDR. The FDR is computed by applying the original comparison criteria (e.g. fold change, p-value) to the sample-wise permuted datasets and recording the number of obtained gene transcripts at each such permutation. If the number of gene transcripts on your list of interest is similar or close to the FDR, it indicates that your list may not be meaningful, and that there is a high chance that the changes occurred by chance. A list with a median FDR<5% or 10% is considered to be meaningful (dChip User's manual 2005).

In order to test the system random samples were manually selected using combinations of the serum specimens, to see if the gene transcripts lists that we are interested in, i.e. Pre vs. Post or Pre vs. normal or Pre vs Post and Normal were deemed to be meaningful and stand out from any other selections tested.

Pre serum and Post serum FDR permutations

The FDR of each of the permutations are listed in table 3.10.7.1. The three permutations highlighted in red were deemed to be meaningful i.e. a median FDR <10%. Two out of a possible 10 Pre vs. Post serum permutations were deemed to be meaningful when the number of samples in each group was 2.

When 3 Pre serum specimens were tested against 3 Post serum specimens, the FDR % median for two samples came close to 10% (15.1 and 12.8% respectively) but none were deemed to be meaningful. However, when the 4 Pre serum specimens were tested against 4 Post serum specimens, 119[1] gene transcripts were called significantly changed, with an FDR of 9% indicating that the list was also meaningful suggesting that there needs to be a minimum number of samples in each sample group to increase the chances of our sample groups of interest being called meaningful.

Of the sample groups tested where either the Pre samples were mixed with the Post samples, or the Pre or Post samples were tested against other Pre or Post samples respectively, none of the lists of significantly changed gene transcripts were deemed to be meaningful.

The results shown here in Section 3.10.7 indicate that there is a definite pattern of gene expression in the Post serum specimens similar to the Normal serum specimens and distinct from the Pre serum specimens, indicating the potential for developing these studies further to perhaps aid in monitoring of success/ failure of therapy treatments.

[1] Slightly different to number calculated in Section 3.10.6 for this comparison as it was calculated using a paired p-value which was not the case here.

2 Pre vs. 2 Post		# sig. gene transcripts	FDR % Median
3, 11	7, 14	48	8.3
3, 7	11, 14	65	15.4
3, 14	7, 11	9	100
7, 11	3, 14	43	97.7
7, 14	3, 11	29	24.1
3, 11	3, 11	17	100
3, 7	3, 7	48	16.7
3, 14	3, 14	11	100
7, 14	7, 14	14	64.3
7, 11	7, 11	20	10
2 Pre vs. 2 Pre			
11, 14	3, 7	7	100
3, 11	7, 14	11	63.6
3, 14	7, 11	6	116.7
2 Post vs. 2 Post			
3, 11	7, 14	12	91.7
11, 14	3, 7	64	18.8
3, 14	7, 11	6	200
2Pre/ 2Post vs. 2Pre/ 2Post			
3, 7 / 11, 14	11, 14 / 3, 7	8	125
3, 11/ 7, 14	7,14 / 3, 11	0	n/a
3, 14 / 7, 11	7, 11 / 3, 14	7	114
3 Pre vs. 3 Post			
3, 7, 11	3, 7, 11	62	24.1
3, 7, 14	3, 7, 14	60	18.3
7, 11, 14	7, 11, 14	93	15.1
3, 11, 14	3, 11, 14	78	12.8
4 Pre vs. 4 Post			
3, 7, 11, 14	3, 7, 11, 14	119	9

Table 3.10.7.1 FDRs of Pre and Post serum permutations. Comparisons highlighted in red were deemed to be meaningful, i.e. with a median FDR less than 10%.

Pre serum, Post serum and Normal serum FDR permutations

When the sample groups are mixed to include 2 Pre serum specimens, 2 Post serum specimens and three Normal serum specimens in one group and the opposite/ left over Normal serum specimens in the other group, none of the permutations are were found to be meaningful (see Table 3.10.7.2).

2Pre/ 2Post/ 3Normal vs. 2Pre/ 2Post 3Normal		# sig. gene transcripts	FDR % Median
3, 7 / 11, 14 / 1, 2, 3	11, 14 / 3, 7 / 4, 5, 6	2	500
3, 7 / 11, 14 / 1, 3, 5	11, 14 / 3, 7 / 2, 4, 6	38	21
3, 7 / 11, 14 / 1, 4, 6	11, 14 / 3, 7 / 2, 3, 5	8	87.5
3, 7 / 11, 14 / 1, 3, 6	11, 14 / 3, 7 / 2, 4, 5	6	116.7
3, 7 / 11, 14 / 3, 4, 5	11, 14 / 3, 7 / 1, 2, 6	38	28.9
3, 7 / 11, 14 / 2, 5, 6	11, 14 / 3, 7 / 1, 3, 4	2	400
3, 7 / 11, 14 / 1, 2, 5	11, 14 / 3, 7 / 3, 4, 6	7	114.3
3, 7 / 11, 14 / 1, 5, 6	11, 14 / 3, 7 / 2, 3, 4	9	66.7
3, 7 / 11, 14 / 3, 5, 6	11, 14 / 3, 7 / 1, 2, 4	9	122.2
3, 7 / 11, 14 / 1, 4, 5	11, 14 / 3, 7 / 2, 3, 6	16	56.3
3, 11 / 7, 14 / 1, 2, 3	7, 14 / 3, 11/ 4, 5, 6	10	80
3, 11 / 7, 14 / 1, 3, 5	7, 14 / 3, 11/ 2, 4, 6	1	700
3, 11 / 7, 14 / 1, 4, 6	7, 14 / 3, 11/ 2, 3, 5	1	600
3, 11 / 7, 14 / 1, 3, 6	7, 14 / 3, 11/ 2, 4, 5	0	n/a
3, 11 / 7, 14 / 3, 4, 5	7, 14 / 3, 11/ 1, 2, 6	0	n/a
3, 11 / 7, 14 / 2, 5, 6	7, 14 / 3, 11/ 1, 3, 4	1	800
3, 11 / 7, 14 / 1, 2, 5	7, 14 / 3, 11/ 3, 4, 6	1	1100
3, 11 / 7, 14 / 1, 5, 6	7, 14 / 3, 11/ 2, 3, 4	0	n/a
3, 11 / 7, 14 / 3, 5, 6	7, 14 / 3, 11/ 1, 2, 4	0	n/a
3, 11 / 7, 14 / 1, 4, 5	7, 14 / 3, 11/ 2, 3, 6	0	n/a
3, 14 / 7, 11 / 1, 2, 3	7, 11 / 3, 14/ 4, 5, 6	17	52.9
3, 14 / 7, 11 / 1, 3, 5	7, 11 / 3, 14/ 2, 4, 6	18	55.6
3, 14 / 7, 11 / 1, 4, 6	7, 11 / 3, 14/ 2, 3, 5	1	1000
3, 14 / 7, 11 / 1, 3, 6	7, 11 / 3, 14/ 2, 4, 5	4	200
3, 14 / 7, 11 / 3, 4, 5	7, 11 / 3, 14/ 1, 2, 6	20	50
3, 14 / 7, 11 / 2, 5, 6	7, 11 / 3, 14/ 1, 3, 4	0	n/a
3, 14 / 7, 11 / 1, 2, 5	7, 11 / 3, 14/ 3, 4, 6	4	250
3, 14 / 7, 11 / 1, 5, 6	7, 11 / 3, 14/ 2, 3, 4	1	800
3, 14 / 7, 11 / 3, 5, 6	7, 11 / 3, 14/ 1, 2, 4	0	n/a
3, 14 / 7, 11 / 1, 4, 5	7, 11 / 3, 14/ 2, 3, 6	4	250

Table 3.10.7.2 FDRs of Pre, Post and normal serum permutations.

The sample groups were then mixed so that the Pre and Post serum specimens were in the right order but the Normal samples were split up and mixed between them. In this

assessment 2 of the 10 comparisons were deemed to be meaningful indicating that there may be some ambiguity when three Normal serum specimens are added to each grouping. However, when the six Normal serum specimens are grouped together either with all of the Pre serum specimens or the Post serum specimens, the only grouping found to be meaningful was the Pre serum against the Post op and Normal serum specimens. Further to this, when the Normal serum specimens are tested against either the Pre serum specimens or the Post serum specimens, only the Pre serum vs. the Normal serum list were found to be meaningful.

4Pre/ 3Normal vs. 4Post/ 3Normal		# sig. gene transcripts	FDR % Median
3, 7, 11, 14 / 1, 2, 3	3, 7, 11, 14 / 4, 5, 6	70	44
3, 7, 11, 14 / 4, 5, 6	3, 7, 11, 14 / 1, 2, 3	196	15.3
3, 7, 11, 14 / 1, 5, 6	3, 7, 11, 14 / 2, 3, 4	125	29.6
3, 7, 11, 14 / 1, 3, 4	3, 7, 11, 14 / 2, 5, 6	18	305
3, 7, 11, 14 / 1, 3, 6	3, 7, 11, 14 / 2, 4, 5	64	54.7
3, 7, 11, 14 / 1, 2, 4	3, 7, 11, 14 / 3, 5, 6	115	29.6
3, 7, 11, 14 / 1, 2, 5	3, 7, 11, 14 / 3, 4, 6	58	65.5
3, 7, 11, 14 / 1, 2, 6	3, 7, 11, 14 / 3, 4, 5	488	6.4
3, 7, 11, 14 / 1, 3, 5	3, 7, 11, 14 / 2, 4, 6	25	180
3, 7, 11, 14 / 1, 4, 5	3, 7, 11, 14 / 2, 3, 6	27	163
3, 7, 11, 14 / 1, 4, 6	3, 7, 11, 14 / 2, 3, 5	201	20.9
3, 7, 11, 14 / 2, 3, 4	3, 7, 11, 14 / 1, 5, 6	66	48.5
3, 7, 11, 14 / 2, 5, 6	3, 7, 11, 14 / 1, 3, 4	305	9.8
3, 7, 11, 14 / 2, 4, 5	3, 7, 11, 14 / 1, 3, 6	79	39.2
3, 7, 11, 14 / 3, 5, 6	3, 7, 11, 14 / 1, 2, 4	70	47.1
3, 7, 11, 14 / 3, 4, 6	3, 7, 11, 14 / 1, 2, 5	106	38.7
3, 7, 11, 14 / 3, 4, 5	3, 7, 11, 14 / 1, 2, 6	46	104.3
4 Pre vs. 6 Normal			
3, 7, 11, 14	1, 2, 3, 4, 5, 6	542	6.3
4 Post vs. 6 Normal			
3, 7, 11, 14	1, 2, 3, 4, 5, 6	182	32.4
4 Pre / 6 Normal vs. 4 Post			
3, 7, 11, 14 / 1, 2, 3, 4, 5, 6	3, 7, 11, 14	38	36.8
4 Pre vs. 4 Post / 6 Normal			
3, 7, 11, 14	3, 7, 11, 14 / 1, 2, 3, 4, 5, 6	220	3.2

Table 3.10.7.3 FDR of Pre serum vs. Post serum with normal serum permutations. Comparisons highlighted in red were deemed to be meaningful, i.e. with a median FDR less than 10%.

The aim of Section 3.11 was to analyse the sequence of these transcripts for presence of known regulatory elements which may enhance the stability of the transcripts.

3.11 Sequence analysis of transcripts identified in CM/ serum

Throughout this study attempts were made to identify and confirm presence of gene transcripts in either the CM of cancer cell lines, normal breast cells or serum from breast cancer patients or normal volunteers. This section covers sequence analysis of some of these transcripts deemed to be either called Present in CM or significantly differentially regulated between groups of CM samples.

335 gene transcripts were called Present in all CM samples. 20 Gene transcripts from this list with an "_at" probe set designation and grade A annotation were randomly chosen for sequence analysis. The transcript assignments for each probe set were firstly aligned with the Affymetrix target probes for that transcript assignment to confirm 100% identity. The transcript assignment sequences were used instead of the reference sequence for the probe sets as in many cases the reference sequence does not align 100% with either the transcript assignment sequence or the Affymetrix target probes.

The transcript assignment sequence for each chosen probe set was then submitted to the UTRscan online analysis software (see Section 2.13.10). The UTRscan programme can scan any submitted sequence for the presence of known 5'/ 3' UTR regulatory sequences. The results of the scan are emailed back to the submitter and the report lists the known sequences and how many (if any) are present in the submitted sequence. The report also includes the exact position of the sequence. There are HTML links to the UTR regulatory sequence database were further information on the regulatory sequences can be found.

Transcripts were also randomly chosen that were called Present in all cancer cell RNA microarray samples and Absent in all cancer cell CM samples for comparison. Exact sequences detected and positions in the transcript assignment sequence for all transcripts are listed in the Appendix (7.2). Tables 3.11.1-3 list the different transcripts chosen for

sequence analysis. The sequences contain a mixture of well annotated and un-annotated sequences.

Probe Set ID	Gene Title	Gene Symbol	GO Biological Process Description	GO Molecular Function Description
1316_at	thyroid hormone receptor, alpha (	THRA	regulation of transcription, DNA-dependent /// transcription from RNA	transcription factor activity /// steroid hormone recepto
1553185_at	RAS and EF-hand domain contai	RASEF	small GTPase mediated signal transduction /// protein transport	calcium ion binding /// GTP binding
1553718_at	zinc finger protein 548	ZNF548	regulation of transcription, DNA-dependent	nucleic acid binding /// zinc ion binding
1554595_at	symplekin	SYMPK	cell adhesion	protein binding /// binding
1556373_at	chromosome 21 open reading fra	C21orf114	--	--
1557996_at	--	--	--	--
202136_at	zinc finger, MYND domain contai	ZMYND11	negative regulation of transcription from RNA polymerase II promoter	DNA binding /// protein binding /// zinc ion binding /// r
202252_at	RAB13, member RAS oncogene	RAB13	intracellular protein transport /// cell adhesion /// small GTPase medi	nucleotide binding /// GTPase activity /// GTP binding
210230_at	Chromodomain helicase DNA bin	LOC388388	--	--
212599_at	autism susceptibility candidate 2	AUTS2	--	--
214235_at	cytochrome P450, family 3, subf	CYP3A5	electron transport /// xenobiotic metabolism /// steroid metabolism	iron ion binding /// oxygen binding /// heme binding ///
220232_at	stearoyl-CoA desaturase 5	SCD5	fatty acid biosynthesis /// lipid biosynthesis	stearoyl-CoA 9-desaturase activity /// iron ion binding
220940_at	KIAA1641	KIAA1641	--	--
222835_at	Thrombospondin, type I, domain	THSD4	--	metalloendopeptidase activity /// peptidase activity
225834_at	family with sequence similarity 7	FAM72A	--	--
227227_at	Hypothetical LOC388397	LOC388397	--	--
228628_at	Similar to Formin binding protein	RP11-343N15.3	--	--
232975_at	Tripartite motif-containing 26	TRIM26	protein ubiquitination	DNA binding /// ubiquitin-protein ligase activity /// prote
233331_at	Similar to Ribosome biogenesis	LOC389953	--	--
244292_at	Full length insert cDNA clone ZD	--	--	--

Table 3.11.1 Probe sets called P in all CM samples chosen for sequence analysis. The Gene Ontology (GO) for the gene transcripts that are annotated are also listed.

Probe Set I	Gene Title	Gene Symbol	GO Biological Process Description	GO Molecular Function Description
1554128_at	small adipocyte factor 1	SMAF1	--	--
1560325_at	Clone 23786 mRNA sequenc	--	--	--
1561305_at	CDNA clone IMAGE:483751	--	--	--
1561573_at	Homo sapiens, clone IMAGE	--	--	--
1561813_at	MRNA full length insert cDN	--	--	--
1563754_at	glutamate receptor, ionotropi	GRIK2	ion transport /// potassium ion transport /// glutamate signaling pathw	receptor activity /// ionotropic glutamate receptor activit
1570505_at	ATP-binding cassette, sub-fi	ABCB4	lipid metabolism /// transport /// response to drug /// transport	nucleotide binding /// ATP binding /// ATPase activity ///
203045_at	ninjurin 1	NINJ1	cell adhesion /// nervous system development /// tissue regeneration	protein binding
207138_at	PHD finger protein 2	PHF2	regulation of transcription, DNA-dependent	transcription factor activity /// protein binding /// zinc ion
207631_at	neighbor of BRCA1 gene 2	NBR2	ER to Golgi vesicle-mediated transport /// protein transport /// transpo	nucleotide binding /// GTP binding
208817_at	catechol-O-methyltransferas	COMT	catecholamine metabolism /// neurotransmitter catabolism	magnesium ion binding /// protein binding /// catechol O
211238_at	ADAM metallopeptidase don	ADAM7	proteolysis	metalloendopeptidase activity
214614_at	homeo box HB9	HLXB9	regulation of transcription from RNA polymerase II promoter /// humor	transcription factor activity /// RNA polymerase II transc
218698_at	likely ortholog of mouse mor	MMRP19	--	identical protein binding
220808_at	Theg homolog (mouse)	THEG	spermatogenesis /// chaperone-mediated protein complex assembly	--
223815_at	Coiled-coil domain containin	CCDC45	--	protein binding
224449_at	DNA-damage inducible prote	DDI2	protein modification	--
229871_at	hypothetical protein FLJ1021	FLJ10211	--	--
233658_at	Chromosome 13 open readir	C13orf24	--	--
240882_at	MRNA; cDNA DKFZp313N0	--	--	--

Table 3.11.2 Probe sets differentially expressed in parental CM samples Vs drug selected variant CM samples chosen for sequence analysis. The Gene Ontology (GO) for the gene transcripts that are annotated are also listed.

Probe Set	Gene Title	Gene Symbol	GO Biological Process Description	GO Molecular Function Description
1553978_at	MADS box transcription enhancer factor 2	MEF2B	transcription /// regulation of transcription, DNA-depend	transcription factor activity /// transcription coa
1556301_at	Full-length cDNA clone CS0DB006YA03 c	—	—	—
203450_at	PKD2 interactor, golgi and endoplasmic re	PGEA1	protein localization /// negative regulation of Wnt recept	beta-catenin binding /// identical protein binding
203500_at	glutaryl-Coenzyme A dehydrogenase	GCDH	electron transport	glutaryl-CoA dehydrogenase activity /// oxidore
212141_at	MCM4 minichromosome maintenance defi	MCM4	DNA replication /// DNA replication initiation /// transcri	nucleotide binding /// DNA binding /// protein bi
217811_at	selenoprotein T	SELT	selenocysteine incorporation /// cell redox homeostasis	selenium binding /// selenium binding
217817_at	actin related protein 2/3 complex, subunit	ARPC4	actin filament polymerization /// actin nucleation	structural constituent of cytoskeleton /// protein
217989_at	dehydrogenase/reductase (SDR family) m	DHRS8	steroid biosynthesis /// metabolism /// lipid biosynthesi	oxidoreductase activity
218112_at	mitochondrial ribosomal protein S34	MRPS34	—	structural constituent of ribosome
218291_at	mitogen-activated protein-binding protein-i	MAPBPIP	—	polypeptide N-acetylgalactosaminyltransferase
218308_at	transforming, acidic coiled-coil containing	TACC3	—	—
218684_at	hypothetical protein FLJ21127	FLJ21127	—	—
218836_at	ribonuclease P 21kDa subunit	RPP21	tRNA processing	ribonuclease P activity /// protein binding /// hy
218893_at	isochorismatase domain containing 2	ISOC2	metabolism	protein binding /// catalytic activity
219366_at	apoptosis, caspase activation inhibitor	AVEN	apoptosis /// anti-apoptosis	protein binding
219394_at	Phosphatidylglycerophosphate synthase 1	PGS1	phospholipid biosynthesis /// metabolism	CDP-diacylglycerol-glycerol-3-phosphate 3-pho
219822_at	mitochondrial translational release factor 1	MTRF1	protein biosynthesis /// regulation of translational termi	translation release factor activity, codon specifi
222759_at	suppressor of variegation 4-20 homolog 1	SUV420H1	—	—
222626_at	pallidin homolog (mouse)	PLDN	synaptic vesicle docking during exocytosis	syntaxin-13 binding /// identical protein binding
223482_at	transmembrane protein induced by tumor	TMPIT	—	—

Table 3.11.3 Probe sets called Present in all cancer line cell RNA samples and Absent in all CM samples chosen for sequence analysis. The Gene Ontology (GO) for the gene transcripts annotated are also listed.

The results for the UTRscan of the sequences called P in all CM samples are listed in Table 3.11.4. 17/20 of the transcripts submitted came back positive for at least one 15-Lipoxygenase Differentiation Control Element (15-LOX-DICE) sequence, with just over half of those transcripts (9/16) containing multiple 15-LOX-DICE sequences. 7/20 transcripts also came back positive for other regulatory sequences; K-box (3/20), Brd-box (2/20), GY-box (1/20) or ADH_DRE (1/20) sequences.

The results of the UTRscan of the probe sets differentially expressed in parental CM samples compared to their drug selected variants are listed in Table 3.11.5. 13/20 transcripts have at least one 15-LOX-DICE sequence, 6 of these 13 have multiple 15-LOX-DICE sequences. 12/20 transcripts also possess another known 5'/ 3'UTR regulatory sequence. 6/20 transcripts have at least 1 K-box sequence (3/6 have multiple sequences), 3/20 have an IRES, 2/20 have either a GY-box or Brd-box and 1/20 have either a CPE or an

ADH_DRE sequence.

The results of the UTRscan of the probe sets called Present in all cancer cell line RNA samples and Absent in the matching cancer cell line CM RNA samples are listed in Table 3.11.6. 18/ 20 came back positive for at least one 15-LOX-DICE sequence, with just under half (8/18) containing multiple 15- LOX-DICE sequences. A further 8/20 also contained other regulatory sequences; K-Box 3/20, IRES 2/20, Brd-Box, ADH_DRE, CPE, TOP and SECIS-2 1/20, none of which were present more than once in the sequence.

Probe Set ID	Gene Symbol	Accession	15-LOX-DICE	Other
1316_at	THRA	NM_199334	5	---
1553185_at	RASEF	AK056176	4 +1 matching repeat	---
1553718_at	ZNF548	NM_152909	3	---
1554595_at	SYMRK	NM_004819	5	---
1555373_at	C21orf114	NM_001012707	---	---
1557996_at	---	AK091784	1	---
202136_at	ZMYND11	NM_00624	---	1 K-Box
202252_at	RAB13	NM_002870	1	---
210230_at	LOC388388	AK027091	2	1 GY-Box
212599_at	AUTS2	NM_015570	7	1 Brd-Box
214235_at	CYP3A5	AK055879	1	---
220232_at	SCD5	NM_024906	---	---
220940_at	KIAA1641	NM_020970	1	---
222835_at	THSD4	AK130734	2	1 K-Box
225834_at	FAM72A	NM_207418	1	1 ADH_DRE, 1Brd-Box
227227_at	LOC388397	AK057167	2	K-Box
228628_at	RP11-343N15.3	BC017972	1	1 SECIS
232975_at	TRIM26	AK023334	---	---
233331_at	LOC389953	AY010114	3	---
244292_at	---	AF086452	1	---

Table 3.11.4 Results of UTR scan for gene transcripts called Present in all CM samples.

Probe Set ID	Gene Symbol	Accession	15-LOX-DICE	Other
1554128_at	SMAF1	NM_001018082	---	1 GY-Box
1560325_at	---	AL832767	4	2 K-Box
1561305_at	---	BC040325	1	---
1561573_at	---	BC039476	---	1 K-Box
1561813_at	---	AL109711	1	---
1563754_at	GRIK2	AJ252246	1	1 K-Box
1570505_at	ABCB4	BC020618	---	---
203045_at	NINJ1	NM_004148	---	---
207138_at	PHF2	AF043725	6	---
207631_at	NBR2	NM_005821	1	---
208817_at	COMT	NM_00754	---	1 GY-Box
211238_at	ADAM7	NM_003817	1	1 K-Box
214614_at	HLXB9	NM_005515	1	---
218698_at	MMRP19	NM_015957	---	1-CPE, 1IRES, 3 K-Box
220808_at	THEG	NM_016585	2	1IRES
223815_at	CCDC45	BX641136	3	1 ADH_DRE, 2 K-Box
224449_at	DDI2	BC006011	1	1 Brd-Box
229871_at	FLJ10211	AK091238	4	1 IRES
233668_at	C13orf24	AK022413	4	---
240882_at	---	BX537871	---	1 Brd-Box

Table 3.11.5 Results of the UTRscan for probe sets differentially expressed in CM samples

Probe Set ID	Gene Symbol	Accession	15-LOX-DICE	Other
1553978_at	MEF2B	AK128256	5	1 K-Box
1556301_at	---	BC024246	1	1 IRES
203450_at	PGEA1	NM_015373	---	1 CPE
203500_at	GCDH	NM_000159	3	---
212141_at	MCM4	NM_005914	2	---
217811_at	SELT	NM_016275	1	1 SECIS-2, 1 Brd-Box
217817_at	ARPC4	NM_005718	1	---
217989_at	DHRS8	NM_016245	1	1 TOP
218112_at	MRPS34	NM_023936	2	---
218291_at	MAPBPIP	NM_014017	---	---
218308_at	TACC3	NM_006342	2	1 IRES
218584_at	FLJ21127	AK024780	1	1 K-Box
218836_at	RPP21	NM_024839	1	---
218893_at	ISOC2	NM_024710	3	---
219366_at	AVEN	NM_020371	1	---
219394_at	PGS1	NM_024419	1	---
219822_at	MTRF1	NM_004294	2	---
222759_at	SUV420H1	NM_016028	1	---
222826_at	PLDN	NM_012388	2	1 ADH_DRE, 1 K-Box
223482_at	TMPIT	NM_031925	1	---

Table 3.11.6 Results of the UTRscan for probe sets called Present in all cell RNA samples and Absent in all CM samples

4 Discussion

The finding that it was possible to amplify mRNA transcripts from the serum of cancer patients came as a surprise to many researchers. This was especially true with cancer patients as they were found to display elevated levels of RNases in their serum (Reddi *et al* 1976).

At the time that the work presented here in this thesis was started, only a few publications regarding the detection of mRNA in the serum of cancer had been published. The techniques utilised for the separation of serum/ plasma and extraction of RNA were different in all of the studies, some of them are summarised below.

Kopreski *et al* (1999) carried out one of the first studies on serum mRNA. They were able to detect tyrosinase mRNA in serum from 4/6 malignant melanoma patients with no detection in 20 healthy controls. The serum was collected by centrifugation of peripheral clotted blood and careful removal of the serum. RNA was extracted from 50µl of serum using a column-based kit for RNA isolation, Perfect RNA Total RNA Isolation kit. Nested PCR (consisting of 15 cycles, then dilution of the PCR solution 1:100 and a further 40 cycles) was then carried out on the RNA for detection of the tumour-related transcript.

In 2000, Chen *et al* isolated serum from breast cancer patients by centrifuging clotted blood to separate the serum; a second centrifugation step was then used to remove cellular debris. RNA was isolated from 100µl of serum using the Promega SV Total RNA Isolation System. RT-PCR was then carried out on 1 or 5µl of the extracted RNA for two telomerase RNA subunits using 50 PCR cycles. Detection of the mRNA transcripts were found in 25-28% of breast cancer serum specimens for the two mRNA subunits and serum from healthy volunteers was negative for both.

Silva *et al* (2000) isolated plasma from breast cancer patients by centrifugation of peripheral blood. RNA was then extracted using an RNeasy mini kit column-based protocol. Nested PCR was carried out on 5-15ng of the extracted RNA for mammaglobin

and CK-19 mRNA using 2 rounds of 40 PCR cycles. 60% and 12% of patients and healthy volunteers respectively were positive for detection of mammaglobin mRNA with 49% and 20% respectively positive for CK-19 mRNA.

These studies helped to realise the potential of using mRNA detectable in serum as possible biomarker candidates. However, despite these early results, we felt that there were limitations to the studies carried out. The numbers of patients in the studies were quite small and only a limited range of gene transcripts were analysed. In addition to this, in many of the studies, the serum was subjected to different centrifugation protocols and filtration of the serum, to ensure prevention of cell carryover, was not always performed. Also, different RT-PCR protocols were carried out using very large numbers of nested PCR cycles in some cases. Due to the fact that the reproducibility of these techniques, in general, had not been addressed this is where we started our study.

By utilising CM from cultured human cancer cell lines we were able to set up experiments in duplicate in order to have biological replicate samples. Different methods for the extraction of RNA from these samples were tested and optimised. All the necessary precautions required for working with RNA were taken i.e. pre- cleaning workspaces with RNase- zap, using only DEPC treated/ RNase-free water, changing gloves regularly *etc.* These initial tests indicated that if the volume of CM from which RNA was extracted was too small, a variable result between duplicate samples was possible. This reinforced for us the advantage of the cell culture system, because in the clinical environment biological repeat samples are not possible. Disparity of mRNA expression between patients could have been explained as the result of patient-specific mRNA expression and not due to any limits of the detection system used, as no replicate samples were available to confirm either way.

Once we had established a protocol for RNA extraction and RT-PCR that was both reliable

and reproducible, we set about testing the CM RNA for the presence of different cancer-related mRNAs. The cancer cell lines that we cultured for removal of CM had previously been studied for expression of different cancer-related mRNA transcripts e.g. MDR1, which guided us in the choice of mRNA transcripts to look for in the CM taken from these cells.

We discovered that in different cancer cell lines expressing the same mRNA transcripts, detection of these transcripts was not always possible in the corresponding CM samples. This meant that cells showed selective expression of mRNA transcripts detectable in their corresponding CM. The detection of mRNA transcripts in the CM samples was not simply a result of a cell "dumping" mechanism. Also, detection of full-length transcripts for at least some mRNA transcripts was possible.

This encouraged us to look for ways to enable us to look at the global expression of mRNA from CM samples using microarray analysis. At the time, the minimum starting RNA requirement for microarray work was 5µg. The recent addition to the laboratory of a Nanodrop enabled us to quantify the RNA extracted from CM, as previous attempts with a standard spectrophotometer had failed due to the small RNA content, although this method also was not without challenges.

Methods for concentrating the RNA using centrifugal devices were assessed. Despite concentration of RNA in the samples to a certain extent, they could not be concentrated sufficiently to yield the 5µg minimum required.

Around this stage in the work, Affymetrix developed a 2-cycle protocol for the amplification of RNA from small samples. This protocol was mainly targeted to studies where only small amounts of tissue were available/ where laser capture microdissection was used to cut single cells from tissues. RNA could be amplified from as little as 10ng to yield sufficient cRNA for hybridisation to their microarray chips, enabling the global

expression analysis of mRNA transcripts from these samples to be carried out.

We decided to perform a pilot study using three biological replicate CM samples as starting material. At this stage, or indeed since, no reports of microarray analysis of extracellular RNA extracted from CM were reported. Our pilot study was found to be successful i.e. enough cRNA was amplified from the CM samples for hybridisation to the whole genome microarray chips. Due to the success of the pilot study, it was decided to expand it to include CM obtained from other cancer cell lines in culture. CM from parental human cancer cell lines and their drug selected variants were processed so that the global mRNA expression profiles could be assessed.

Bioinformatics software including dChip and Genespring aided in the visualisation and manipulation of the data obtained. Hierarchical clustering was performed on the samples and dChip software was successfully applied to grouping most of the CM replicate samples together. This was a very important result in so far as we had proven that the CM from cultured cancer cell lines, differentiated only by their selection to a particular drug, exhibited unique gene expression patterns that distinguished them from the other cancer cell line CM sample groups analysed.

The groups of replicate parental vs. drug selected variant CM samples were then explored for the patterns of gene expression that may distinguish them from the other samples analysed. From a list of the most highly changing gene transcripts between parental and drug selected variants, 10 gene transcripts were chosen for confirmation by RT-PCR. Validation of the microarray data using RT-PCR proved problematic in that only 50% validation was achieved. However, this discordance of microarray data with RT-PCR data was not unique to our system. The MicroArray Quality Control Project has estimated that, in general, above their signal threshold (of 2-10 mRNA copies/cell) microarray measurements accurately reflect the existence and direction of expression changes in ~70-

90% of the gene transcripts (Draghici *et al* 2005).

At this time a group in California published results on microarray studies that they had carried out on extracellular RNA extracted from saliva samples of oral cancer patients and healthy volunteers (Li *et al* 2004 (a) and (b)). We were able to compare the results obtained from extracellular mRNA from CM with the results they obtained from extracellular saliva mRNA. They, too, found some discordance when confirming the microarray results with RT-PCR methods.

Due to the successful application of RNA derived from CM samples to microarray techniques, it was decided to evaluate its usefulness using serum specimens obtained from breast cancer patients. Blood specimens were collected from patients on the morning of surgery to remove a breast tumour and a follow-up sample was taken 2-4 months post surgery. Tumour and matched normal tissues were also collected from the same patients and blood specimens were obtained from healthy volunteers. All blood specimens were treated the same i.e. processed within four hours and the serum was separated from the clotted blood, centrifuged and filtered with a 0.45μm filter before being stored at -80°C.

Using the methods developed and described in this thesis (with minor modifications), RNA was extracted from the serum, and equal volumes of serum RNA were amplified and hybridised to an Affymetrix whole genome microarray chip. The corresponding tissue RNA (tumour and normal), were also amplified and hybridised to the microarray chips.

Whilst hierarchical clustering of the serum specimens did not yield distinct groupings of the specimens, further analysis of gene lists generated, e.g. with Pre-operation (Pre) serum specimens vs. post-operation (Post) serum specimens, including analysis of false discovery rates (i.e. the chance of these changes occurring by chance), indicated that the Post serum specimens were more similar to the gene expression profile of the serum specimens belonging to the healthy volunteers (Normal) than the Pre serum specimens. The

comparison of the Pre serum specimens to the Normal serum specimens yielded 542 significantly changed gene transcripts, whilst comparison of the Post serum specimens to the Normal serum specimens yielded only 182 significantly changed gene transcripts. Further analysis of the false discovery rates of these comparisons revealed that the comparison of the Pre serum to the Normal serum specimens and Pre serum to the Post serum specimens yielded the most meaningful data i.e. the chance of falsely identifying a gene changed by chance was low in these comparisons.

4.1 Identifying a suitable time point for CM collection and analysis

Cells were generally seeded at $2x10^6$ cells per $T75cm^2$ flask in duplicate. Conditioned media (CM) samples were collected at 4 time-points after cell seeding to identify a suitable time-point at which amplifiable mRNA could be routinely isolated for analysis. As indicated in Fig 3.1.2, in the case of β-actin, mRNA could be isolated and amplified at all time points evaluated. The increasing intensity of the PCR bands in the samples over the 96 hours may be due to the increasing amounts of β-actin transcripts secreted into the CM over the 96 hours assuming that the transcripts are stable in CM for that length of time. Another reason may be that as 1ml of CM was taken every 24 hours and was not replaced, the concentration of the β-actin gene transcripts increased due to the smaller volume of media in the flask with each 24 hours. As the cells were in a healthy, proliferating state at the 48 hour time-point (see Fig 3.1.2 as example) with little or no dead cells floating in the media, this time point was chosen for all further studies using this RNA extraction protocol.

4.2 Determining conditions for reproducible detection of mRNA in CM

The first RNA extraction method tested was the Promega SV Total RNA Isolation protocol. This was to mimic the extraction method employed by Chen *et al* (2000) were they used the kit to extract RNA from 100µl of serum. They found that they were able to amplify two telomerase RNA gene transcripts from the serum of breast cancer patients using RNA that was extracted with this protocol and 50 cycles of PCR.

Fig 3.2.1.1 (a) and (b) show that the PCR results obtained using RNA isolated with the Promega system, when trying to amplify the CK-19 gene transcript, were not reproducible. Duplicate samples showed little/ no concordance with each other e.g. Fig 3.2.1.1 (a) samples 24 (a) and 24 (b). The reason for this could have been that the sample size of 100µl was too small and the PCR was at its detection limit or that the RNA consisted mostly of small fragments which could not consistently be picked up by the spin columns (shorter transcripts like the 5S rRNA and tRNA molecules cannot be extracted using this technique as they are smaller than 200 nucleotides and don't bind to the spin columns).

Experiments were then set up to test both of these theories using 48 hour CM (for the reasons stated in Section 3.1 and also because these time points showed inconsistent results when using RNA extracted with the Promega kit (Fig 3.2.1.1)). RNA was extracted from two volumes of the CM, 100µl and 250µl and was amplified, in parallel, for a range of gene transcripts, to test out whether the PCR was at its limit of detection with the smaller volume. To overcome the possibility that the binding columns may have led to the inconsistent results, the RNA was extracted using the modified Tri Reagent protocol, which extracts the RNA (including small fragments/ nucleotides) by isopropanol precipitation rather than using binding columns.

The results of this experiment can be seen in Figs 3.2.2.2-3 and 3.2.2.5- 6. There was an

overall increase in the reproducibility of all the results for RPMI TAX CM, but not for the RPMI ML CM. The results noticeably show that when the smaller volume of CM was used (samples labelled 48.100), although β-actin was amplifiable for both duplicate samples (Fig 3.2.2.2), CK-19 and HnRNP B1 had conflicting results for the duplicate samples (Figs 3.2.2.3 and 3.2.2.5).

When the larger volume of CM was used for extraction of RNA (labelled in results 48.250 and equivalent to approximately 50µl CM/ RT as opposed to approximately 9µlCM/ RT with the 48.100 samples), the results show that there were no inconsistencies between duplicate samples. This proved that the PCR was previously most likely at its limit of detection with the smaller sample volume. The Tri Reagent protocol was easily adjustable so that more RNA could be re-suspended in smaller amounts of DEPC treated water and so concentrating the volume of RNA in the final RNA sample. For this reason therefore, the Tri Reagent extraction method was deemed the protocol of choice for this work. Based on this data and the results in section 3.1, which showed that at 48 hours the cells look quite healthy and were not confluent, it was decided that all future CM studies would be carried out using 250µl of 48hour cultured CM and extracted using the modified and optimised Tri Reagent protocol.

4.3 Amplification of β-actin gene using 5' and 3' specific primers

Unlike the column-based methods for extraction of RNA (e.g. Qiagen RNeasy kit, Promega SV isolation kit) where smaller fragments of RNA (~≤200nts e.g. 5S rRNA, tRNA) can be lost through the column, the precipitation method used in the Tri Reagent RNA extraction protocol would ensure that all fragments were extracted.

The decision to use the Tri Reagent protocol for extraction of the CM RNA meant that it was still possible that some of the RNA transcripts amplified in the previous results (referred to in section 3.2) could have resulted from amplification of fragmented RNA i.e. degraded mRNA but with the poly-A tail still intact. Although reverse transcription was carried out using Oligo dT as a reverse primer, not all transcripts had 5' targeted primer sets like HnRNPB1.

In some studies of human serum/plasma it has been suggested that the RNA detected is probably present as short fragments (El- Hefnawy *et al* 2004). Similarly, extracellular RNA isolated from HeLa and A431 growth medium has been suggested to be represented by 100-200 nucleotide-long fragments (Morozkin *et al* 2004)). Therefore, in order to check whether or not the RNA (at least for the β-actin gene transcript) was full-length, primers were designed to amplify 5', middle and 3'regions of the β-actin gene transcript (Fig 3.3.1 for positions of amplified regions). As the RNA was reverse transcribed using only Oligo dT with no random hexamers, any 5' specific amplified regions would indicate that, for β-actin at least, there were full length gene transcripts present.

Fig 3.3.2 shows that it was possible to amplify different regions of the β-actin gene transcript. The intensity of the amplified product from the 5' UTR region is less intense than the other three regions, but it clearly indicates the presence of full length β-actin gene transcripts and also the intensity of the 5' coding region targeted primers PCR products are

as intense as the shorter 3' coding region targeted primer PCR bands for the CM samples.

Li *et al* (2004) also assessed RNA isolated from the serum of oral cancer patients for the presence of full length transcripts. Four housekeeping gene transcripts (ACTB, B2M, GAPDH and RPS9) were assessed. Primers were designed to amplify regions that covered 56.8- 88.9% of the gene transcripts. The amplicons ranged in sized from 188bp to 1Kb. Gene transcripts for the four gene transcripts were amplifiable in all cases.

Studies carried out by Wong *et al* (2005) on extracellular placental RNA extracted from maternal plasma, suggested that there was a higher incidence of 5' mRNA fragments in maternal plasma samples. Out of 7 gene transcripts screened, 5 of them showed higher concentrations of 5' amplicons over 3' amplicons. For one gene transcript in particular, βhCG mRNA, the concentration of the 5' amplicon increased with the length of gestation.

In an assessment of the RNA extracted from saliva for the presence of full-length transcripts, Park *et al* (2006) found that only 4/5 transcripts were detectable. GAPDH was not detected in any of the saliva RNA samples. Of the other four transcripts analysed, β-actin, RPS9 and IL8, gene transcripts were detected in 2/8 samples and SAT mRNA was detected in 4/8 samples. Relative abundances of β-actin and IL8 5', middle and 3' regions of the mRNA transcripts were assessed by real-time PCR. Park *et al* found that amounts of the middle and 3' regions of the two transcripts were less abundant than the 5' region.

The β-actin RT-PCR results shown in Fig. 3.3.2 do not indicate that there are more abundant 5' amplicons than 3' amplicons in RNA extracted from CM.

4.4 Analysis of gene transcription in RNA isolated from cell lines and conditioned media

Expression of transcripts representing a number of gene products (including mdr-1, mrp-1, CK-19, HnRNP B1, GST-π, topoisomerase II, bcl-2 and β-actin; amplified products ranging between 157 – 383 bp (Table 2.11.1.1)) was evaluated in a range of human cancer cell lines (see Table 2.5.1 for the list of cell lines and their origin) and corresponding CM. Although 30 cycles of PCR were sufficient for amplification of RNA from cultured cells, 45 cycles was generally found to be necessary for CM studies. However, use of nested primers and large numbers (70-83 cycles) of PCR cycles as reported for some of the serum studies (Kopreski *et al* 1999, Silva *et al* 2001)) were not found to be necessary in our experiments.

Two different filter sizes were also tested for each CM. According to Ng *et al* (2002), filter size can have a significant effect on the amount of RNA extractable from serum specimens. They found a statistically significant reduction in hPL, βhCG and GAPDH mRNA levels detectable in the serum of pregnant women after filtration with a 0.45μm filter.

It was decided to test CM filtered with a 0.45μm filter (in addition to the 0.22μm filtered CM), which would still exclude cells from the CM, to see if there was any effect on the gene transcripts amplifiable.

As expected, the ubiquitously expressed "house-keeping" gene, β-actin, was detected following 30 cycles of PCR in all cell line extracts analysed.

For all RNA extracted from cells, β-actin was successfully co-amplified in association with all other gene transcripts of interest. However, although β-actin was routinely amplified alone in all CM, attempts to co-amplify β-actin in these samples inhibited detection of other gene products of interest. For this reason, β-actin was not co-amplified with other cDNAs when analysing CM.

β-actin was successfully amplified following 0.22/ 0.45μm filtration of the CM (Fig. 3.4.1). Amplification of CK-19 and HnRNP B1 gene transcripts were also detected in all cell and CM samples similar to β-actin (Figs 3.4.2, 3.4.3).

To investigate if specific mRNAs were detectable only in CM from cell lines expressing that product (i.e. to ensure results did not include false positives), CM from cell lines expressing versus not expressing a particular gene transcript was analysed. As shown in Fig. 3.4.4, GST π was detected in all cell lines studied, except MCF-7 and BT474. However, although GST π was undetected in all media conditioned by BT474 cells, a weak band resulted from analysis of the MCF-7 0.45 μm CM filtrate. This result was reproducibly found. A possible explanation for this unexpected finding may be the use of 45 cycles of PCR for CM analysis, with only 30 cycles routinely used for amplification of RNA from cell lines.

Topoisomerase II (Fig. 3.4.5) mRNA was expressed by all cell lines studied and was amplified following isolation from 0.45 μm filtered CM. However, with the exception of medium conditioned by the nasal carcinoma cell line variant, RPMI-TAX, and the breast cancer cell lines, MDA-F and MDA-F-ADR-SI (very weak band), topoisomerase II was generally undetected in 0.22 μm filtered CM.

MRP-1 was expressed by all cell lines (Fig 3.4.6(a). It was also detected in only 5 of the 0.45 μm filtered CM, but in only 2 of the subsequent 0.22 μm filtrates, analysed (Fig. 3.4.6 (b, c).

Bcl-2 was expressed (*albeit* at relatively low levels) in the entire 8 cell lines analysed (see Fig. 3.4.7 (a)). However, this gene transcript was apparently not secreted by these cells *i.e.* with the exception of a very weak band resulting following amplification of 0.45 μm filtered RPMI-TAX CM, bcl-2 mRNA was undetected in all CM.

Results from this study suggest that the secretion of mRNA by human cancer cells may not just be a general "dumping" mechanism, but may be somewhat selective. This is supported by the fact that although β-actin, CK-19 and HnRNP B1 are expressed by all cell lines and are detected in all CM analysed, other mRNAs - such as MRP1 and bcl-2 - were present in RNA isolated from most/all cell lines, but were not present in all corresponding CM samples. An alternative explanation to selective secretion of the gene transcripts may be differing stabilities of mRNAs secreted. However, support for the former suggestion is the fact that, for example, RPMI-ML and RPMI-TAX cells grow in the same medium type, under the same culture conditions; both cell line variants express mrp-1; but, although this product is detected in both the 0.45 μm and 0.22 μm CM from RPMI-TAX, it is undetected in all medium conditioned by RPMI-ML cells.

With one exception, an amplified product was never detected in CM that was not detected in the corresponding cell line.

To prove that the mRNA was not somehow coming from the media itself that the cells were cultured in, a sample of media, unconditioned by any cells, underwent the same filtration/ extraction protocol as the CM. Fig 3.4.1.1.1 shows that this sample was negative for β-actin which indicates that the mRNA found in the CM did not come from the media or FCS added.

4.5 Concentration of CM with Millipore concentrators

Millipore centrifugal devices are routinely used in our laboratory for concentration of proteins from CM samples. They are characterised by their ability to retain molecules above a specified molecular weight. The percent of retentate (concentrate) recovery for a membrane, according to the manufacturers, is over 90%. In addition to their obvious use in recovery of proteins from dilute samples, one of the intended uses of the devices is concentration of nucleic acids from dilute solutions. Despite this advertised intended use, the manufacturer has carried out no evaluation of the recovery yields of DNA or RNA.

Having developed/ optimised methods for the reproducible analysis of a range of mRNAs in medium conditioned by a range of cancer cell lines (results in Section 3.4), we wished to investigate the feasibility of studying the global expression of mRNAs in CM using whole genome microarrays. To achieve this and to enable results from microarray analysis to be validated by RT-PCR, increased yield of RNA may be required. For this purpose we investigated the possibility of concentrating the RNA in CM samples by using the Millipore centrifugal devices.

The results shown in Section 3.4 indicate that the centrifugal devices did aid in concentration of RNA fragments found in the CM samples (see bioanalyser traces Fig 3.5.2.1 (h, i, k and l)). However the quantity of the RNA in the concentrated samples, as measured by the Nanodrop, indicated a fold increase of less than 1.5 (from 59.8 ng/µl in the 0.45B sample to 75.55ng/µl in the 0.45C sample- see Fig 3.5.1). The fold increase was slightly better at 1.67 for the 0.45C (4X) sample compared to the 0.45B (4X) sample. In the samples taken before concentration that were filtered through a 0.45µm filter (0.45B and 0.45B (4X)) however, by extracting RNA from 4 x 250µl aliquots and re-suspending the four pellets into the one 15µl volume (0.45B (4X)), the concentration of the RNA in the

resulting solution was 3 fold higher than the sample in which only one RNA pellet (resulting from extraction of 1 x 250µl aliquot of CM), was resuspended. RNA fragments from this sample were also visible on the bioanalyser analysis of the samples (see Fig. 3.5.2.1 (h)). None of the protocols tested, despite different optimisations and an increase in the RNA quantity, were able to yield the 5µg of RNA required for microarray analysis.

4.6 Interference in quantification of small amounts of RNA

In a review of methods for real-time quantitative PCR, Ginzinger (2002) assessed the variables that might limit the use of real-time quantitative PCR. One of these variables related to nucleic acid purification methods utilised. In the review Ginzinger recommended that care should be taken if the protocol is a phenol/ chloroform-based method. Any trace amount of phenol can contribute to reduced reverse-transcription efficiencies. As traces of phenolic compounds absorb light in the UV range, any traces of phenol carried through the protocol can also result in an overestimation of the amount of RNA present when the samples are quantified using a UV spectrophotometer.

This was also confirmed in a study carried out by an Agilent Technologies researcher, Lightfoot (Agilent application note, 2002). Quantification of 100ng of RNA spiked with different volumes of phenol (0.5- 5%) was carried out using the Agilent 2100 bioanalyser, a UV spectrophotometer or a ribogreen kit. The results showed that even in the presence of 0.5% phenol contamination, the UV reading for the RNA sample was overestimated.

A posting on a Yahoo microarray group also indicated that other researchers have found that extraction of RNA using Trizol (essentially the same as Tri Reagent), where low amounts of RNA are expected, can result in overestimation of the RNA content if quantified using a Nanodrop (see Appendix 7.1).

To assess if there was any phenol present in the CM samples that may induce an overestimation of the quantity of RNA in the sample, a simple experiment was designed. Firstly Tri Reagent was diluted 1:100 with PBS or UHP and a sample was quantified using the Nanodrop. Fig 3.6.1(A) shows that the diluted Tri Reagent absorbs light at 260 and 270nm as indicated by the peaks on the graph at these wavelengths.

When 250µl of UHP was added to Tri Reagent and the RNA extraction for CM samples

followed (see Section 2.8.2.2), quantification of the resulting solution demonstrated that the Nanodrop falsely calculated that there was 70-108 ng of RNA in the samples (see Fig. 3.6.1 (B) & (C)). Utilisation of an overnight precipitation step slightly increased the false reading (Fig. 3.6.1 (B)). The combination of a Tri Reagent RNA extraction protocol with an RNeasy clean-up protocol reduced the false reading from 68.9/ 108.2ng down to 12.1 ng (Fig 3.6.1 (D)). The volume of water required to elute the sample from the RNeasy column however, was twice the volume used to resuspend the RNA pellet using the Tri Reagent protocol. This would have aided in the reduction of the interference by diluting the sample.

To confirm that the interference was due to the Tri Reagent, an RNeasy extraction was carried out using UHP/ PBS. The RNeasy RNA extraction protocol is a binding-column based procedure that can bind RNA and does not rely on the use of phenol. The quantification of the eluted solution gave a reading of 1.7ng, which is at background levels, and the corresponding absorbance curve in Fig. 3.6.2 shows clearly the difference in absorbance at 260 and 270nm obtained using the two different protocols.

Cell derived RNA extracted using Tri Reagent however does not show evidence of any contaminating peaks at 270nm (see Fig. 3.6.3 (B)) which leads us to the conclusion that phenol carryover only affects/ interferes with the quantification of RNA when the RNA solution cannot be diluted before quantification (see Fig. 3.6.3 (A) for example of CM sample).

Alternative methods for RNA quantification were sought at this time. A Quant-it RNA assay kit (Molecular Probes Inc) was tested as it claimed to be able to quantify as little as 5ng of RNA independent of any DNA that may be in the sample. Whilst it was able to accurately quantify diluted cell RNA diluted down to 50ng, it was unable to accurately quantify below this concentration and was also unable to quantify the CM RNA samples.

4.7 Microarray Analysis of DLKP CM

4.7.1 RNA quantification and QA analysis

Whilst the Nanodrop UV readings of the DLKP CM RNA before amplification (Fig. 3.7.1.1(a)) show two peaks at 260 and 270nm, after two rounds of amplification and with all the inherent cleaning steps, the amplified RNA does not show evidence of phenol contamination as evident by the single peak at 260nm in the amplified sample (Fig. 3.7.1.1(b)).

The three DLKP CM samples (DLKP 1, 2 and 3), showed similar cRNA yields after the 2nd cycle of the two-cycle amplification protocol, despite slightly different yields after the 1st round IVT (Table 3.7.1.1). Bioanalyser traces in Fig. 3.7.1.2 show that although the DLKP 3 CM sample did not show the presence of any RNA fragments before amplification, Fig. 3.7.1.2 (b) clearly shows a large amount of amplified cRNA present in the sample following 2 rounds of amplification. Fig. 3.7.1.2 (c) shows the efficiency of the fragmentation step as the labelled cRNA is uniformly fragmented to nucleotides of approximately 200bps long.

The samples passed the quality controls outlined by Affymetrix when hybridised to the chips with little/no difference between the three chips. The % Present calls were 23.2, 22.6 and 23.5 respectively (Table 3.7.2.1), although this may seem low compared to the average % Present calls of 35- 50% obtained with cell RNA, Affymetrix stipulate that results may vary greatly depending on the sample being analysed, and as long as repeat/ duplicate samples have similar QA results, then a low present call is acceptable.

4.7.2 Scatter plot and Venn diagram Analysis of DLKP CM

Although the % gene transcripts called Present between each CM sample seemed to suggest the samples are quite similar, when these data lists were overlapped using a simple Venn diagram to check the concordance (as indicated by Fig 3.7.3.1), there is quite a bit of variation between the replicate samples. However, this variation seemed to correlate with the Affymetrix probes that had expression values under 100 (Fig 3.7.3.2) as indicated by the grey dots in the scatter plot, the probes sets with higher expression values correlated with the probes that were commonly expressed in all three replicates.

The CM samples were taken from DLKP cells 4 passages apart (DLKP 1= passage 54, DLKP 2, DLKP 3 = passage 58). Fig 3.7.3.2 shows that the difference in cell passage number can have an effect on the expression levels of gene transcripts in the different samples. CM RNA from passage 54 (DLKP 1) cells had more Affymetrix probes expressed at levels over 100 not commonly expressed at these levels with the CM samples taken from cells at passage 58 when based on gene expression over 100 (Fig. 3.7.3.2), however there did not seem to be a huge effect on the number of probe sets called P, M or A. This indicates the importance of taking the CM from cells at the same passage number in order to reduce any variation in the system.

It has been shown that RNA that undergoes a two- cycle amplification protocol has a slightly different profile than RNA that undergoes a one- cycle protocol (Wilson *et al* 2004). The different profiles relate to the fold-changes observed. Certain groups of gene transcripts show higher/lower fold change compared to the unamplified samples. This is why when the CM data is compared to the cell line data; an expression cut-off of 100 is not applied to the DLKP cell data as it would only bias the data.

Fig 3.7.4.1 shows that when DLKP RNA cell data is plotted on a 3-D scatter plot the amount of "noise" is greatly reduced compared to the CM samples, even at levels under

100 the correlation of the replicate data is much better.

When the number of probe sets called P or M in all DLKP cell RNA samples is compared to the number of probe sets called P or M in the DLKP CM samples, Fig 3.7.5.1 shows the number of gene transcripts called P or M common to both DLKP cells and CM is 7645. When only gene transcripts with an expression value of 100 or more in the CM data is included, Fig 3.7.6.1 shows that the number drops to 1492 indicating that the majority of gene transcripts commonly called P or M (between the cell and CM RNA), for DLKP CM has expression levels under 100. Out of the 1,492 probe sets called P or M in both the cell and CM DLKP samples, despite the majority of the CM probes sets having an expression <100, over 65% (1,432) of these probe sets have a P or M call assigned to them.

4.7.3 Ontology of gene expression in DLKP CM using Onto express software

Onto express software is part of an Onto-Tools suite of software which aids in the translation of lists of gene transcripts into a functional profile that offers insights into the cellular mechanisms. Profiles can be constructed for different gene ontology (GO) categories. The data generated comes from GenBank, UniGene, LocusLink and Pubmed.

A gene list was created of the 1,492 gene transcripts that were common to the DLKP CM and cell RNA samples (see Fig. 3.7.7.1) to assess the biological annotation for these transcripts. A large number of gene transcripts belong to a biological category containing less than 20 gene transcripts from the input list ("other" segment of pie chart Fig. 3.7.7.1). There are also a high number of gene transcripts on the list for which no biological annotation is known ("unknownP" segment of pie chart). The most highly represented biological process category is protein biosynthesis.

The molecular function of the same 1,492 gene transcripts belong to molecular function categories containing less than 20 gene transcripts from the input list (Fig. 3.7.7.2). The molecular function category represented on the input list is RNA binding, in fact 14 other molecular function categories represented on the input list refer to binding functions.

Whilst the Onto-Express software aids in assigning biological and molecular function to the list of probe sets, caution must be taken against placing too much value on the results obtained for two reasons. The first reason is that gene transcripts often have more than one molecular function assigned to it and also can be listed as involved in more than one biological process, for this reason it is possible that one gene transcript is represented on the pie chart more than once. This is confirmed when the number of genes in each category is added up and the total number is higher than the number of gene transcripts on the input list. The second reason caution must be taken when interpreting the data is that whilst the

gene transcripts are either assigned as Present or expressed at significantly different levels by microarray analysis, this does not automatically mean that the gene is expressed at levels high enough to be biologically functional. In addition to this, because the Affymetrix probe sets only target the very 3' end of gene transcripts, it is possible that some transcripts are degraded at the 5' end and so are not translatable into protein.

Despite this however, the software is quite amenable to handling large lists of probe set numbers and converting them into pie charts that aid in understanding the underlying biology.

4.8 Analysis of mRNA expressed by various cell line CM

4.8.1 cRNA yield throughout amplification procedure

The cRNA yield of the CM RNA after 2 rounds of amplification was 55.53 ± 26.5μg when the RPMI 2 CM sample was excluded (because it did not amplify the minimum required amount of cRNA). This average is reduced to 53.3 ± 27.8μg if the RPMI 2 CM sample is included. The reason why the RPMI 2 CM sample did not yield enough cRNA for hybridisation (despite two attempts) is not clear. It is possible that the RNA in this sample was degraded as quantification by the Nanodrop would not have detected this, although RT-PCR for the 3' region of β-actin was positive.

Although the Nanodrop may have overestimated the amount of RNA in the CM samples, the amount of cRNA generated after 2 rounds of linear amplification were in the range of quantities of cRNA generated by Schindler *et* al (2005) using cell line RNA that was diluted to 50ng. In the study by Schindler *et al* (2005) they were able to generate between 40.9 and 45.8μg of cRNA from 50ng of input RNA and for 14/24 of the CM samples in this study the amount of cRNA generated was higher than 46μg. An average of 42.2μg ± 3.94 of cRNA was also obtained by Li *et al* (2004) from an average of 60.5ng ± 13.12 ng of input RNA from cell free saliva samples of 10 healthy donors. This indicates that whilst the quantity of CM samples may have been over estimated, for over half of the samples, the yield after 2 rounds of linear amplification, was higher than achieved with 50ng of cell RNA/ 60ng of cell free saliva RNA which shows that the level of inaccuracy of the Nanodrop, for these samples, most likely was less than 40ng.

4.8.2 Microarray QA results for CM samples

The background and noise levels for the CM samples in Section 3.8 were both within acceptable parameters with the exception of MDA-F-TAX-SI. The background in this sample was 113.2, as this fell out of the 3 fold range of most of the other samples it was excluded from further analysis (see Table 3.8.2.1). A high background implies that impurities like cell debris and salts are binding to the array in a non-specific manner.

The scaling factor for all of the CM samples in Section 3.8 were approximately twice the scaling factors of the DLKP CM samples assessed in Section 3.7 (Tables 3.8.2.1 and 3.7.2.1 respectively). The scaling factor is the multiplication factor applied to each signal value on an array and a scaling factor of 1.0 indicates that the average intensity is equal to the target intensity. The only guideline for assessing samples is that they must not differ by 3 fold within an experiment.

Two other QA parameters differ from the results obtained with the DLKP samples; % Presence call and GAPDH ratios. The % Presence call in the DLKP samples were on average 23.1% (see Table 3.7.2.1), however in all of the other CM samples (with the exception of MDA-F-ADR-SI 1) the % Present call was 9.6 ± 2%. In addition to this the 3'/5' and 3'/M GAPDH ratios were only attainable for some of the CM samples do to a large number of the middle target probes being assigned as Absent. The most likely cause of the lower % Present calls and missing GAPDH ratios was due to the time delay in processing the samples. The DLKP CM samples were processed (i.e. filtered and added to Tri Reagent) within 1 minute of removal from cells, whereas the time delay from filtering the CM samples in Section 3.8 until addition of Tri Reagent was up to twenty minutes. This may have lead to the degradation of the more unstable transcripts and thus result in a lower %Present call for these samples. It is possible also that the scaling factor is inversely

correlated to the % Present call. For cell RNA samples hybridised to the Affymetrix Plus 2.0 chips in our laboratory, typical % Present calls were between 35-50% with scaling factors close to 1, indicating that the higher the % Present call of the samples, the lower the scaling factor. As there are no published data on microarray analysis of CM RNA to date, there is nothing to compare the quality assessment measurement results to. Studies by Li *et al* (2004 (a)) detailing results obtained from microarray analysis of RNA extracted from cell-free saliva RNA reveal nothing about the scaling factors but do show % Present calls in the samples of 14.11± 2.98%.

In a study by Schindler *et al* (2005), a 50ng RNA sample that was subjected to two rounds of amplification, when compared to the standard single round of linear amplification of 10μg of the same RNA, showed a reduction in % Present call from approximately 52% to 43.7%. This reduction of % Present call was also correlated to increase with GAPDH and β-actin 3'/5' ratios. These observations were also found by Wilson *et al* (2004).

Schoor *et al* 2003 studied the effects of degradation on mRNA samples subjected to different amplification protocols; the standard protocol and the two round protocol (a precursor protocol to the 2 cycle protocol employed in this study). Both tumour and normal tissue samples were assessed and degradation of the RNA was carried out by incubating aliquots of powdered tissue in PBS at 22°C for different lengths of time. To stop degradation, Trizol was added. The results of the study were that moderate degradation of mRNA did not change the microarray gene expression profile results obtained. In addition to this, a higher scaling factor (2-10) was noted for the two most degraded samples and did result in a reduction of % Present call (although only 11μg of cRNA was hybridised to the array chips for these samples).

The study carried out by Wilson *et al* (2004) compared the 2 cycle amplification protocol using cell RNA diluted down to 100, 10 and 1ngs respectively, to the standard 1 cycle

protocol. They too found a slight increase in the scaling factors of the data (up to 1.96) but not to the same extent as the CM samples presented in this study (see Table 3.8.2.1). This indicates that the high scaling factors obtained for the CM samples in this study reflect the origin of the RNA, and the magnitude of difference between the DLKP CM samples and the other CM samples reflects degradation of RNA within these samples due to the time delay in sample processing.

Whilst an obvious choice for repeating this work would include prompt processing of the CM samples with Tri Reagent to prevent degradation of the CM RNA, possibly more relevant would be a time- course study to look at what mRNA transcripts are degraded over time but more importantly, which ones remain stable.

4.8.3 Expression of Two-Cycle Hybridisation and Labelling Controls

There are two internal controls supplied with the Affymetrix RNA amplification kits. The Poly-A control kit is designed to assess the efficiency of the amplification process and the hybridisation kit assesses the hybridisation part of the process.

The Poly-A controls are *B. subtilis* genes that have been modified by the addition of poly-A tails. These modified genes are then amplified by being cloned into vectors that contain T3 promoters. Amplification of the Poly-A controls with T3 RNA polymerase yields sense RNAs that are spiked into the RNA sample of interest before amplification. The Poly-A controls are pre-mixed at different concentrations. Affymetrix have probe sets that target the 5', middle and 3' region of the Poly-A controls. The results shown in Fig. 3.8.3.2 show that for the CM samples, expression of the 3' region of the Poly-A controls was higher than expression of the middle or 5' region in all samples assessed. Comparison of these results to results obtained from our laboratory with cell RNA that did not undergo the 2 cycle protocol indicate a distinct 3' bias in samples generated using the 2 cycle protocol. The profiles obtained for cell RNA that did not require the 2 cycle protocol showed 3'/M or 3'/5' ratios much closer to 1:1 in comparison to the CM Poly-A controls where the ratios were as high as 8:1 with the *thr* Poly-A control.

This 3' bias effect has also been noted by other users of the 2 cycle amplification protocol (Schoor *et al* 2003, Wilson *et al*, 2004 and Schindler *et al* 2005). Since Affymetrix target their probes to the most 3' 600 bases of the mRNA transcripts, however, the 3' bias introduced by the 2 cycle protocol should not affect the microarray data obtained. This was confirmed by Schindler *et al* (2003), whilst use of the 2 cycle protocol affected the 3'/5' ratios for β-actin and GAPDH, they found no reduction in the % Present call between the samples generated with either the standard labelling protocol or the 2 cycle protocol.

The result shown in Fig. 3.8.3.3 indicates that the ratio of expression of the Poly-A controls did not reflect the ratio of the Poly-A RNA in the original stock solution. However, analysis of microarray data generated in our laboratory using cell RNA yielded similar results. This indicates that the most likely cause for discrepancies in the expression ratios detected was different hybridisation efficiencies for the different probe sets i.e. some probe sets have a greater affinity for their target than others and in these cases the expression values generated are higher. This also should not affect downstream analysis of the microarray data as the probe set, although displaying a higher affinity for the target sequence, will display this higher affinity across all samples tested.

The second controls included with the Affymetrix amplification kit are the hybridisation controls. These are gene transcripts from the biotin synthesis pathway of *E.Coli* and from the recombinase gene of bacteriophage P1. These control sequences are added to the amplified cRNA samples before hybridisation and are used to assess the efficiency of the hybridisation and washing steps of the procedure. The results in Fig. 3.8.3.1 show that these transcripts were detected in the right order and at ratios of approximately 1:1 indicating that the hybridisation and washing steps were without incidence.

4.8.4 Scatter plots and Venn diagrams of CM samples

The scatter plots and Venn diagrams in Section 3.8.4 serve to function as visual aids to interpret the large amount of microarray data generated more easily. The colour of the dots on the scatter plots correlate to the grouping of the probe sets on the adjoining Venn diagram e.g. probe sets that are common to all samples tested regardless of criteria are coloured white on the scatter plot and correspond to the probe sets listed in the centre of the three ringed Venn diagram (in cases where 2 replicate samples are available they are coloured yellow).

For all eight of the cell line CM samples tested, the overlap of gene transcripts common to the replicate samples is significantly higher when only gene transcripts with an expression value over 100 are included. In general, the percentage of gene transcripts commonly called Present or Marginal to the replicate samples was 16% or less (were there were only two replicate samples in the cases of RPMI and MDA-F-TAX-SI this percentage was higher at 20-25%). However this percentage correlation of replicate samples rose to 60-70% for the CM samples when only probe sets with expression values over 100 were included with 2 exceptions. In the MDA-F and MDA-F-ADR-SI CM sample groups this percentage correlation was 55 and 45% respectively. This drop in correlation for these samples was a direct result of one replicate in each group displaying either a higher % Present call or a high background QA parameter. This result shows that for the CM samples, there is a high level of "noisy" gene transcripts with expression levels under 100 i.e. the expression data at that level is highly variable. Schindler *et al* (2005) noticed a two-fold increase in discordance of calls assigned to probe sets when a 2 cycle amplification protocol was used in comparison to a standard one cycle protocol resulting in non-detection of low-abundant RNA with the 2 cycle protocol. This may indicate that some of the discordance in the calls

assigned to the CM samples is attributable to the 2 cycle protocol.

4.8.5 Condition tree and hierarchical clustering of CM samples

The "Condition Tree" option in Genespring software enables similar samples/ conditions to be grouped together based on similar expression profiles. Fig. 3.8.6.1 shows the results obtained for the condition tree of the CM samples using Genespring. Three CM sample groups, H1299 TAX, RPMI ML and RPMI TAX are grouped beside each other. For two other sample groups, H1299 and MDA-F, at least two of the three replicate samples are grouped beside each other. A Pearson correlation was used to assess how similar the samples are to each other.

In dChip, a Pearson correlation can also be selected, however, some steps can be taken to filter out non-essential data before performing hierarchical clustering. To do this any gene transcripts that were not changing across the samples were removed. A second step was included that was not performed on the data before the condition tree analysis was carried out in Genespring. As gene expression data can be quite varied i.e. expression can go from 0- 60,000 expression units, a row-wise standardisation step was carried out on the data. The standardisation step involves linearly scaling the data for a gene transcript across all samples to have a mean 0 and a standard deviation of 1. It is these standardised values that are used to calculate correlations between samples. This enables all data to be viewed so that the range of expression is pre-definable, e.g. the default range is -3 to +3.

The dChip software was able to group the samples better (Fig3.8.6.2). Only one sample, MDA-F 1 was really out of place. A look at the Affymetrix quality control measures assessed for the CM samples (and listed in Table 3.8.2.1) indicates that the reason for this may be due to the fact that the MDA-F 1 sample had the second highest background levels

resulting in a lower % Present call for that sample. This may have influenced how the samples were grouped. For this analysis dChip was found to be better at clustering the samples. This result indicated that based on gene expression profiles alone, different CM samples were distinguishable from each other.

4.8.6 Validation of Microarray data (1)

The first parameter that we decided to attempt to validate was the Presence/ Absence call assigned to different probe sets. This was carried out for three transcripts. The first, β-actin, has probe sets on the microarray that detect the 5'/ middle and 3' regions of the mRNA transcript. Generally, with the exception of two samples, the 5' and middle target probes were assigned as Absent in all of the CM samples. The 3' region was assigned as Present in all of the samples tested. The results in Figs. 3.8.7.2.1-3, show that whilst detection of β-actin gene transcripts corresponding to the 5' UTR region of the β-actin mRNA was only possible in 6-14/ 23 CM samples, PCR products were amplifiable in all original unamplified CM RNA samples targeted by 5' coding- and 3' region specific primers.

The results obtained with the Poly-A control RNA (results Section 3.8.3 and discussion 4.9.3) indicated the possibility that the 3' bias of the 2 cycle protocol was the reason that only 3' regions of the β-actin gene transcript were detected by microarray. To investigate this further, two other gene transcripts, stearoyl-CoA desaturase 5 and APAF1 interacting protein were chosen for validation of the call assigned in the microarray analysis by RT-PCR. Stearoyl-CoA desaturase 5 was assigned as Present in all CM samples and APAF1 mRNA was assigned as Absent in all samples by microarray.

The results shown in Figs. 3.8.7.3.1/2 indicate that for both samples, after 45 cycles of PCR, mRNA transcripts were detected in all CM samples. So whilst the discordance of RT-

PCR and microarray results for β-actin mRNA could be explained by the 3' bias of the 2 cycle protocol, this bias would not have affected the detection of APAF1 interacting protein mRNA. It is generally thought that RT-PCR is a more sensitive technique than microarray. Presence of only a few mRNA transcripts is sometimes enough for a positive RT-PCR result, whereas microarrays depend on the hybridisation of these targets at a single temperature and with stringent washing steps to avoid non-specific binding for all targets. This may result in loss of signal for low abundant gene transcripts.

4.8.7 Gene transcripts significantly up/ down-regulated between parental and drug selected variant CM

Treatment of cancer using chemotherapeutic drugs is frequently hindered by the development of resistance of tumour cells to the anticancer agents, often associated with the emergence of multidrug resistance (MDR) (Liang *et al* 2001). For this reason, the gene expression profiles of the parental CM samples compared to the drug selected variants were evaluated for candidate gene transcripts that may be indicative of drug resistance.

In Section 3.8.8, lists were generated using dChip of gene transcripts that were found to be significantly changed in the parental CM samples in comparison to the drug selected variant CM samples. A gene was determined to be significantly changed if it had a p-value of significance less than 0.05, a fold change greater than 1.2 and a difference in expression greater than 50 fluorescence units.

Of the five parental vs. drug selected CM comparisons, the three parental CM samples, when compared to their taxol selected variants showed more down-regulated gene transcripts in the taxol selected CM samples than for example, when RPMI was compared to its Melphalan selected variant. This was most notable for the MDA-F CM groups. When MDA-F was compared to MDA-F-TAX-SI, 80% of the significantly changed gene transcripts were down-regulated in the MDA-F-TAX-SI CM samples. Comparison of MDA-F-CM samples to MDA-F-ADR-SI CM samples indicated that 90% of the gene transcripts found to be significantly changed were up-regulated in the MDA-F-ADR-SI CM.

Attempts were made to find a gene that may be indicative of taxol resistance detectable in the CM of all three samples. Table 3.8.8.6 shows that when the three parental CM samples are grouped together and compared to the grouping of the taxol-selected CM samples, 34 gene transcripts were determined to be significantly changed (76% of which were down-

regulated). However when the same 34 gene transcripts were analysed in the individual parental vs. taxol selected variant CM samples (Table 3.8.8.7) none of the 34 gene transcripts were found to be significantly changed in all three of the individual groupings. With the exception of 6 targets however, (five of them in the MDA-F vs. MDA-F-TAX-SI group) all gene transcripts displayed the same trend in change of expression. It is possible also that a combination of transcripts, whilst not individually significant, together may yield a biomarker group that would correlate to emergence of drug resistance in these cells. The results in Section 3.8.8 confirm the results in Section 3.4 that different gene transcripts are found in the CM of cancer cells. If this process was a general "dumping" mechanism/ or a mechanism unrelated to the cells themselves, the types of gene transcripts detected would be similar and the number of gene transcripts significantly differentially expressed would be quite small.

4.8.8 Validation of Microarray data (2)

In order to choose gene transcripts for validation of differential expression of gene transcripts in parental vs. drug selected variant CM samples by RT-PCR, strict criteria were followed. The first was obviously that the gene transcripts were assigned as significantly changed (as determined by a p-value <0.05) between the two groups and the second was that a difference in expression of 100 fluorescent units was required. This was to reduce selection of gene transcripts that were expressed at levels under 100 in both samples (as these were shown previously to be quite variable in replicate samples (see discussion in Section 4.9.4)). The third criterion was that the gene transcripts needed to show a fold change between the groups greater than 1.2.

88 gene transcripts differentially expressed in some of the parental and drug selected variant CM samples passed this criterion. Of the 88 gene transcripts, only 41 of the probe sets corresponding to these gene transcripts were "_at" probe sets meaning they uniquely identified one gene transcript. From this list only transcripts with a grade A annotation (see Section 2.13.5) were included. All remaining sequences were examined using BLAT (see Section 2.11.1) to ensure exon-exon boundaries were identifiable so that mRNA specific primers could be designed.

From this shortened list, 10 gene transcripts were chosen for validation by RT-PCR. The gene transcripts chosen displayed a range of fold changes between sample groups of 1.37-5.6. Presence or Absence of the gene transcripts in the different samples did not influence the choice of gene as the results discussed in Section 4.9.7 indicated that these were not reliable. Interestingly, one of the gene transcripts chosen for validation by RT-PCR was APAF1, which was assessed in the study discussed in Section 4.9.7. Despite being assigned as Absent in all CM samples, it was determined to be significantly changed in both RPMI

ML and RPMI TAX CM samples compared to the parental RPMI sample.

The results shown in Figs. 3.8.11.1-20 with corresponding expression data in Tables in 3.8.11.1-10 show that validation of the microarray results was possible for 50% of the gene transcripts tested.

This correlation may seem slightly disappointing at first glance, however the pool of candidate gene transcripts was quite small and they were not chosen for any biological reason i.e. not based on background literature searches or pathway analysis software. In fact only 4 of the 10 gene transcripts chosen for the validation study were annotated for gene function. A closer look at the exact differences between the gene transcripts that were validated and those that were not reveals a distinct pattern.

Table 3.8.11.11 summarises the different attributes for the gene transcripts assessed by RT-PCR for confirmation of the microarray results. For 4 of the 5 transcripts that confirmed results obtained by microarray, the mRNA region targeted were close to or the same as that targeted by the Affymetrix probe. In comparison only 2/5 of the invalidated gene transcripts utilised primers that targeted the Affymetrix targeted region.

For 3 of the 4 gene transcripts, (in which primers that amplified the same region targeted by Affymetrix were used and RT-PCR correlated with microarray data), gene function was annotated. In comparison only 1 of the 5 transcripts not validated by RT-PCR were annotated for gene function. And finally, although the results discussed in Section 4.9.7 indicated that Presence or Absence call assigned to a gene could not be validated by RT-PCR, the results summarised in Table 3.8.11.11 indicate a correlation of Presence call of transcripts studied to validation of the microarray result. 3/5 of the gene transcripts validated by RT-PCR displayed a majority Present call amongst the samples tested in comparison to 0/5 of the transcripts that we were unable to validate.

All of these results together indicate that the criteria employed to select gene transcripts for

confirmation by RT-PCR was flawed. Despite the gene transcripts being statistically the best candidates, more weight should have been given to the Presence/ Absence call of the probe sets for the gene transcripts and also only gene transcripts with well described annotation should have been chosen. Although, as neither microarray nor confirmatory RT-PCR studies have been previously carried out on RNA extracted from CM samples, it is quite important that these criteria have been noted at this early stage. Whilst "blind" selection of gene transcripts for confirmatory RT-PCR resulted in less than desirable validation results, the information obtained, which will aid in selection of candidate gene transcripts in future, is invaluable.

Analysis of the limited published data on microarray studies of cell-free RNA (of which there are three, originating from the same lab) also show discrepancies with validation of the microarray results by RT-PCR methods. Two of the three publications refer to validation of changes found in RNA extracted from the saliva/ serum of oral cancer patients in comparison to healthy volunteers. In the first with saliva RNA (Li *et al* 2004 (b)), 7/9 of the gene transcripts selected were validated by quantitative RT-PCR and in the second more recent study concerning serum RNA only 5/10 gene transcripts selected for validation based on their reported cancer association were confirmed by quantitative RT-PCR to be significantly elevated in the oral cancer specimens (Li *et al* 2006). It is interesting to note that these discrepancies were found despite strict criteria employed for choosing gene transcripts to validate. These criteria were that all samples must have a Present call for the gene of interest and in the first study (Li *et al* 2004(b)) a fold change greater than 3 with a p-value less than 0.01 and in the second study a fold change greater than 2 with a p-value of significance less than 0.05.

Discrepancies relating to validation of microarray results however, are not confined to the

limited studies on extracellular RNA or, indeed, to microarray data generated using RNA amplified with the 2 cycle protocol. Dallas *et al* (2005) tested expression levels for 48 gene transcripts found by microarray to be differentially expressed between their tissues of interest e.g. brain tumour vs. normal. The authors calculated the degree of correlation between the microarray expression scores and expression levels measured for the same gene transcripts using qPCR. The microarray data generated was normalised by MAS 5.0 and also separately by RMA. MAS 5.0 algorithm uses a scalar normalisation technique taking into account perfect match (PM) and mismatch (MM) probe pairs to correct for non-specific hybridisation and RMA is based on a quantile normalisation approach that ignores the MM values.

They noted that for gene transcripts in which the PCR primers don't recognise the same subset of transcripts identified by the Affymetrix probe sets correlation of results was only 41% (7/17 gene transcripts) with both sets of normalised data. Of the remaining 31 gene transcripts for which the microarray probe sets were deemed to recognise the exact same transcript/ subset of transcripts as the qPCR probes 26/31 were validated for MAS 5.0 normalised data and 27/31 were validated for RMA normalised data. This gave an overall non-concordance rate of 13-16%.

A review paper by Draghici *et al* (2005) indicated that results from the Microarray Quality Control Project, a project initiated and led by FDA researchers, shows that above their sensitivity threshold (of between one and ten copies of mRNA per cell), microarray measurements accurately reflect the existence and direction of expression changes in ~70-90% of the gene transcripts. The magnitude of the changes measured with technologies such as RT-PCR however tend to be different to the microarray data.

Etienne *et al* (2004) found that RT-PCR data correlated well with microarray data for gene transcripts with moderate levels of expression that have PCR primers located to the

microarray target region. The authors analysed 26 gene transcripts that had previously been found in their laboratory by RT-PCR to be of interest to their particular study. They concluded that gene transcripts with very high or low levels of expression, or those with larger separation between the location of the PCR primers and microarray target region, often displayed reduced agreement between the two methods. Etienne *et al* (2004) also noted that the region targeted by the Affymetrix probes were mainly in the 3'UTR region of the gene transcripts (only 3/26 were outside of this region).

Whilst carrying out RT-PCR using primers that target the same region as the Affymetrix probes may increase the chances that the results from the two techniques will correlate, consideration must be placed on the question being asked. By only targeting the region of the mRNA transcript targeted by Affymetrix probes (i.e. within last 600 bases of 3'UTR region), the question of whether the mRNA is full length and in a position to be translated, remains unanswered. If presence of the 3' region of the transcript alone however, can be correlated to prognostic status, e.g. presence is correlated with overall survival, then information about the ability of the mRNA to be translated into protein is not necessary.

4.8.8.1 No Reverse Transcription Controls

The results in Section 3.8.11.1 highlight the importance of including proper and full controls for every experiment. Despite careful design, the RT-PCR primers chosen for the microarray validation work were able to amplify a PCR product when no reverse-transcription step was carried out. It is possible that the low annealing temperature (54°C) and position of exon-exon junction in primer sequence (i.e. with enough of the mRNA-specific primer annealing to DNA to enable amplification of PCR product in the presence of DNA) contributed to the amplification of a PCR product in the absence of a reverse transcription step. Another likely reason for amplification of a PCR product is the presence of pseudogenes. A lot of the transcripts chosen for validation were poorly annotated so the presence /sequence of psuedogenes are unlikely to be known.

There are a few options for solving this problem, the first would be to treat all RNA solutions with RNase-free DNase. However this has proven problematic in our laboratory on previous occasions, degradation of RNA has occurred. A second solution is to design primers that anneal to both RNA and DNA but have an intron between them. This would mean that DNA would effectively be distinguishable from RNA as the RNA would not contain the intron sequence and so would be much shorter. This solution would also help to identify whether intron-less pseudogenes are present. Any amplification of a PCR product the same size as the RNA product in a sample that is either; reverse transcribed without reverse transcription enzyme or does not undergo a reverse transcription step would identify the presence of a pseudogene.

4.8.9 Gene Ontology of probe sets called Present in all CM samples or significantly differentially expressed between parent and drug selected variants

The gene transcripts found to be called Present in all CM samples or significantly differentially expressed in parent CM samples compared to their drug selected variants were submitted to Netaffx for assessment of their gene ontology (GO). Netaffx was used in this analysis instead of Onto-Express (that was used for GO analysis of the DLKP CM samples see Section 3.7.7) because there a fixed number of parent categories e.g. binding, catalytic activity *etc.* This meant that a quick visual comparison could be made of the overall gene ontology of the gene transcripts in each list. The annotations, as with Onto-Express are derived from public databases such as UniGene, LocusLink and Homologene.

Fig 3.8.5.1 indicates that the majority of gene transcripts called Present in all CM samples are un-annotated, with binding as the largest represented molecular function group. Comparison of the pie chart in Fig. 3.8.5.1 with the pie chart in Fig. 3.8.9.1 shows that binding is also the largest represented molecular function group assigned to the gene transcripts that were significantly differentially expressed between the parent and drug selected variant groups.

The gene ontology categories of the transcripts significantly differentially expressed between parent and drug selected variant groups however; do not follow the same pattern as the gene transcripts called Present in all CM samples. A smaller percentage of the gene transcripts differentially regulated between the parent and drug selected variant groups relate to structural molecule activity and larger percentage relate to transcription regulator activity and signal transducer activity.

As mentioned before elsewhere, looking at the molecular function of gene transcripts detected by microarray relies on the assumption that the transcripts are full length and

available for translation, however this analysis does enable us to see the differences in the two lists as determined by the biological function of the gene transcripts on those lists.

4.8.10 Analysis of CM mRNA expression profiles against cell mRNA expression profiles

The gene expression profiles of the CM RNA samples were compared to the cell RNA profiles to assess the amount of overlap between the two samples. The aim was to see if (1) gene transcripts called P/ A/ M in the CM samples were equally called P/A/M in the cell RNA samples and (2) to assess how many (if any) gene transcripts found to be significantly up-/ down-regulated in the CM parental vs. drug selected variants were similarly up-/ down-regulated in the cell RNA samples. It is important to note that the RNA used in the microarray study of the cell RNA was generated using the standard one cycle protocol compared to the CM RNA which was processed using the two cycle protocol.

4.8.10.1 Probe sets called P or M in both CM and cell mRNA profiles

The Results shown in Section 3.8.12 show the overlap of gene transcripts called P or M in all three CM samples compared to the gene transcripts called P or M in all three corresponding cell RNA samples. Table 3.8.12.1 summarises all the results for Section 3.8.12. Looking at the results listed in Table 3.8.12.1, it is clear that based on Present/ Absent or Marginal call alone, there is a poor correlation between the gene expression profiles for the CM and cell samples. The results would seem to suggest that as many as 35-51% of gene transcripts present in the CM samples were not present in the corresponding cell RNA samples.

Schindler *et al* (2005) found that when the same cell RNA was processed using the standard one cycle protocol or diluted and processed using a two-cycle protocol that the concordance of P/A calls assigned to probe sets between the two methods was reduced. Up to 10% of probe sets that were called Present in the data obtained with RNA processed with the one cycle were called Absent in the data generated from the same RNA that had been

diluted and subjected to the 2 cycle amplification protocol. A further 1-2% of probe sets went from Absent to Present in the data generated from RNA using the 2 cycle amplification protocol.

The overlap between the CM data and corresponding cell data drops further (40-51%) if the probe sets with an expression value over 100 in the 3 replicate CM samples are overlapped with the probes sets called Present in the 3 replicated cell RNA samples are included. This was carried out because the results shown in Section 3.8.4 indicated that for all the CM samples, gene transcripts that were expressed at levels under 100 were quite variable between the replicate samples. If the replicate samples were overlapped based on the probe sets expressed at values over 100 the correlation between the replicate CM samples increased from an average of 16% (based on Present or Marginal call) to an average of 65% when only the probe sets with an expression value over 100 were included.

The results in Section 3.8.4 indicated that a number of the probe sets with an expression value over 100 in the three replicate samples were also called Absent. For an example see Fig 3.8.4.2, white dots on scatter plot represent probe sets expressed at levels over 100, and compare to Fig. 3.8.4.1, where a number of the probe sets that were coloured white in Fig 3.8.4.2 indicating a high expression value, are coloured grey in Fig. 3.8.4.1, indicating that despite having a high expression value an Absent call was assigned to that probe set. For this reason, the third test was to compare the overlap of probe sets called Present in all three cell RNA samples to the probe sets called Present and with an expression value over 100 in all three CM samples. This was to filter out any noisy gene transcripts (i.e. those with an expression value under 100) and also to filter out any gene transcripts not called Present. The results are also summarised in Table 3.8.12.1. The percentage of gene transcripts overlapping in the cell and CM, when this filter is applied increases from 40-51% up to 62-84%.

A final comparison of the probe sets called P with an expression value over 100 in all replicate samples for a CM group were overlapped with the probe sets called A in all replicate samples for the corresponding cell RNA group. This reduced the discordant number of probe sets to an average of 8% for all of the CM and cell groups.

4.8.10.2 Probe sets significantly changed in both CM groups and cell groups

The gene transcripts deemed to be significantly changed in both CM and cell were compared to see if the trend of up-/ down-regulation in the parental group compared to the drug selected variant group was the same regardless of the origin of the RNA i.e. CM or cell derived.

Schindler *et al* (2005) found that in the comparison of microarray gene expression profiles of RNA that was processed with two different amplification protocols (one cycle and two cycle), that not only was there variation in the call assigned to the probe sets but also in the signal intensities and fold changes. The difference in expression of probe sets between the two protocols was 17.3% of probe sets increased (5.8% of which increased by 2 fold) and 20.9% decreased (of which 8.5% were decreased by 2 fold). For this reason a p-value <0.05 was the only criteria used for selection of probes sets significantly changed.

The results shown in Section 3.8.13 indicate that for all the parental and drug selected variant groups tested, (with the exception of MDA-F vs. MDA-F-TAX) there was a mixture of gene transcripts up-regulated and down-regulated in both CM and cell RNA along with a number of gene transcripts that were differentially regulated in both.

The gene transcripts that were significantly differentially expressed between the parental and drug selected variant CM samples that had a p-value <0.05 with a fold change> 1.2 and a difference in expression over 50 are highlighted in bold in Tables 3.8.13.1- 3.8.13.5.

Analysis of these gene transcripts (which were filtered to exclude some of the gene transcripts that may lie in the "noisy" range of expression in the CM samples- i.e. expression under 100), the correlation of change in expression between the parental and drug selected variants does not increase. This may indicate that gene transcripts are differentially expressed in the two samples as part of a cell specific process. For example, if as a result of selection with a chemotherapeutic drug a cell secretes more of a particular transcript than it retains for translation in the cell, there would be an increase in the amount of that transcript detected in the CM RNA and a decrease in the amount detected in the cell RNA. This is however only a theory and confirmation of such could only be obtained if the RNA samples were both processed with the one amplification protocol i.e. one or two cycle.

4.8.10.3 MICROARRRAY ANALYSIS HMEC CM AND CELLS

One of the end goals of the work presented in this thesis is to identify gene transcripts for further analysis as potential cancer-specific biomarkers. In Section 3.8 gene transcripts that were differentially expressed between different cancer cell line CM samples were identified. Unfortunately no single transcripts were detected in the CM samples that would identify with resistance of that cell line to a chemotherapeutic drug (see Section 3.8.8, Tables 3.8.8.6 and 3.8.8.7).

To assess the possibility that we may be able to detect a transcript in CM that is a cancer-specific marker, we set up a pilot study to culture normal cells. We used HMEC 1001-10 breast mammary cells (HMEC). Unfortunately biological replicates were not available for this sample. As only one $25cm^2$ flask was available and the cells are only viable for approximately 15 days, there was not enough time/ cells to split the flask and passage 3 times to obtain biological replicate samples.

As the CM from normal cells had not been tested previously by us to ensure that mRNA transcripts were detectable from the CM, RT-PCR was firstly carried out on a few transcripts to see if it was possible. Fig. 3.9.1.1 indicates that β-actin primers that target the 5'UTR region of the gene transcript were able to amplify gene transcripts (although the band is quite faint). The PCR results in Figs. 3.9.1.2 and 3.9.1.3 show that other gene transcripts were also detectable.

We then proceeded to amplify 100ng of HMEC CM and cell RNA as quantified by the Nanodrop using the 2 cycle amplification protocol (see Section 2.12). Table 3.9.2.1 shows that there was more cRNA amplified in the cell RNA sample compared to the CM sample (97.65μg compared to 21.12μg).

The details of the QA results of the chips indicate that the two samples cannot be

normalised together as the scaling factors are so different. The scaling factor for the HMEC CM sample, although high, is comparable to the scaling factors obtained for the cancer cell line CM samples (see Table 3.8.2.1). The CM sample used for this pilot study was processed promptly and added to Tri Reagent post filtration within 1 minute. This means that the low % Present call for the HMEC CM sample cannot be attributed to any time delay in sample processing. The most likely explanation is that there are more gene transcripts assigned as Present in the cancer cell CM than in the HMEC CM.

The three regions of β-actin targeted by Affymetrix probe sets were assigned as present in both samples. This is also a reflection of the promtness by which the samples were processed. As with the DLKP CM samples, also processed promptly, β-actin was called P at the 3' target regions. With the CM samples analysed in Section 3.8, only the 3' target region was called P. A possible reason is that β-actin is not stable in CM and most likely is constantly secreted into the CM and mostly degraded (within 15 minutes). With the exception of the GAPDH middle region in the CM sample, the three target regions were also assigned as Present in both samples.

As both samples were amplified using the same protocol it was hoped that the concordance of probe sets called Present in the CM sample would overlap better with the cell RNA than previously for the CM and cell RNA samples (see Table 3.8.12.1 for summary of results for cancer cell line CM and cell RNA samples). The average percentage overlap for the probe sets called Present or Marginal in the cancer cell line RNA with its corresponding CM RNA was 16% (although the probe sets had to be called P or M in all of the replicate samples for each). The percentage overlap of probe sets called P or M in the HMEC CM and cell RNA samples was 67% (see Fig. 3.9.4.1). If an additional cut-off of expression over 100 is applied to the CM sample, the percentage of gene transcripts common to both the CM and cell RNA samples rises to 80%. If the same additional cut-off is applied to the

HMEC cell RNA (i.e. expression over 100) and compared to the HMEC CM RNA the percentage probe sets overlapping between the cell and CM RNA samples decreases to 24%. A criterion of a Present call and expression over 100 in both samples reduces the number of probe sets common to both down to just 3%.

This indicates that whilst comparison of samples using the two different amplification protocols itself introduces discrepancies between the probe sets called Present, Marginal or Absent (as in Schindler *et al* 2005), in the samples studied during this work, a high proportion of the non-concordance is also attributable to the different origins of the RNA i.e. CM vs cell.

The comparison of probe sets that were called Absent in the HMEC compared to the probe sets called present in all of the CM samples studied in this thesis (Section 3.7 and 3.8) yielded a list of 23 gene transcripts called Present in all 26 CM samples and Absent in the HMEC sample. Table 3.9.5.1 shows that most of the gene transcripts are un-annotated.

A second comparison was then generated comparing the gene transcripts expressed in the 26 CM samples compared to the HMEC CM sample. As there were no replicate HMEC samples available statistical significance could not be assigned to the results so a stricter criteria of fold change and difference in expression was employed; fold change >2 and difference in expression >100. The 143 gene transcripts that resulted from this comparison are shown in Table 3.9.5.2. The majority of the transcripts were found to be up-regulated in the cancer cell line CM samples compared to the normal cell CM sample (88 up-regulated and 55 down-regulated in cancer cell line CM compare to the HMEC CM.

4.8.11 Microarray analysis of Breast cancer serum specimens

4.8.11.1 Analysis of QA Parameters, Expression of 2-Cycle Controls and Housekeeping Gene transcripts for Breast Cancer Specimens

The QA results listed in Table 3.10.1.1 show large differences in the serum and tissue samples in regard to the scaling factor, % Present call and GAPDH ratios. The choice of using 2µl of RNA extracted from 1ml of serum for all serum specimens resulted in a decrease of the standard deviation for amplification of the samples. An average of 86 ± 17.17µg of cRNA was generated using this volume of RNA solution. The respective yields of prostate and lung cancer serum specimens generated in our laboratory using 100ng of RNA as determined by the Nanodrop, were 59.26± 30.7µg and 41.7± 24µg. This indicates that the choice to amplify volume amounts over quantified amounts helped decrease the standard deviation of the cRNA generated.

The GAPDH ratios for all tissues except the normal tissue from patient 14 were acceptable. With the tissue from patient 14, the larger GAPDH ratios correlated with a decrease in % Presence call and an increase in the scaling factor compared to the other samples (see Table 3.10.1.1).

Results shown in Section 3.10.2.1 show that the hybridisation controls indicated that the hybridisation and washing steps were without incidence. In Section 3.10.1.2 however there seems to be some discrepancies for the amplification of the Poly-A controls, this may have been due to the quality and quantity of the endogenous RNA content of the samples. For example, the two tissues with the lowest % Present call and highest scaling factors belong to the tissue samples corresponding to patient 14. The % Present calls were 43.5 and 38.7 % and the scaling factors 23. and 2.74 in the normal and tumour tissues respectively. These two tissue samples, as shown in Fig. 3.10.1.2, show higher levels of expression of the 3'

region of the *dap* spike in control. This may indicate that the lower quality of RNA in these samples resulted in higher amplification of the Poly- A controls. Another possible explanation is that it is a patient specific phenomena, the expression levels of the *dap* 3' target region in all samples corresponding to this patient are all expressed at high levels. Lower expression of this Poly-A control target region in the serum specimens 11Pre, 7Post and FN2 did not correlate with a reduction in % Present call for all of the samples. Reasons for the non correlation of the detected expression of the Poly-A controls compared to their original concentration before amplification have been discussed already in Section 4.9.3.

The results showing the expression of the housekeeping gene transcripts β-actin and GAPDH are shown in Figs. 3.10.2.1 and 3.10.2.2 and Tables 3.10.2.1 and 3.10.2.2. There was low expression detected for the three target regions of both transcripts in all of the serum specimens. Whilst expression of GAPDH and β-actin were much higher in the tissue samples there was a distinct difference in the expression of the 3' region of β-actin to the 5' and middle regions (ratio of approximately 5:1).

4.8.11.2 Condition Tree and Hierarchical Clustering of Breast Clinical Samples

The results of the condition tree and hierarchical clustering of the breast clinical samples shown in section 3.10.4 indicates that the results generated using dChip were better at distinguishing the samples from each other (as was the case for the CM samples, see Section 4.8.5 for discussion). Both types of software distinguished that the tissue samples were different to the serum samples (see Figs. 3.10.4.1 and 3.10.4.2). This was most likely as a result of the large differences in % Present call between the tissue and serum samples (~45% in the tissue specimens compared to ~8.3% in the serum specimens) rather than the result of any specific gene expression profile patterns.

When the serum samples were analysed on their own, so that the gene expression data relating to the tissue samples didn't influence the clustering, a better separation of the serum samples was achieved. Fig 3.10.4.3 shows the results of this clustering. The serum specimens are divided into two groups, in one there are 3 Pre serum specimen samples and 2 Normal specimen samples and the other group contains all 4 Post serum specimens along with 4 Normal serum specimens and 1 Pre serum specimen. This indicates the possibility, that with a larger sample number, a distinct gene expression profile for Pre-op/ Post-op/ normal serum specimens may be possible.

4.8.11.3 Probe sets called P or M in different serum groups

Analysis of the overlap of probe sets called Present or Marginal in the pre- operative (Pre) serum specimens compared to the probe sets called Absent in the post- operative (Post) serum specimens shown in Fig. 3.10.5.1 indicate that there are 51 probe sets called P or M in the Pre serum sample and A in the Post serum sample. However, closer inspection of the expression values for these probe sets (see Table 3.10.5.1) reveal that the expression difference between the Pre and the Post samples is very small (if indeed there is any difference at all). This is also reflected in the 6 probe sets that are called P or M in the Pre serum and A in both the Post and normal serum specimens (see Figs. 3.10.5.2 and 3.10.5.3).

This may indicate that these 51 probe sets are borderline calls i.e. just over the threshold to be called P or just under the threshold and so are called M or A. In microarray comparison studies carried out on RNA that was amplified with either 1 or 2 cycles of linear amplification, Schindler *et al* (2005) noted that with the 2 cycle protocol, changes in expression calls assigned to the samples were found but only for probe sets with expression levels under 100. Approximately 10% of the probe sets assigned as P with samples amplified using the standard protocol were switched to A when the 2 cycle protocol was used. Analysis of Table 3.10.5.1 reveals that 49/50 transcripts displayed levels of expression under 100.

This indicates that no gene transcripts were reliably Present in the Pre serum and Absent in the Post serum, and as expected, the changes in gene expression are more subtle making the search for a diagnostic biomarker much harder.

4.8.11.4 Probe sets significantly changed in Pre vs. Post serum specimens

Because of the results discussed in the previous section, it was decided to assess the difference in the Pre and Post serum specimens based on their differential expression. The criteria chosen for gene transcripts to be called significant were that there had to be a difference in expression of 20 fluorescence units, a fold change greater than 1.2 and a significant paired p-value of less than 0.05. The lower difference in expression criteria was employed simply to increase the chances of picking up gene transcripts that may display subtle but significant changes as a direct result of the lumpectomy surgery in the individual patients. The use of the paired p-value was employed to increase the power of the statistical results obtained, i.e. the probe set not only had to show differential expression in the groupwise Pre vs. Post specimen comparisons but also in the individual patient Pre vs. Post specimen comparisons.

Only 4% (5/114) of probe sets were down-regulated in the Post serum specimens, which seems to be an illogical result. We would have expected that removal of the tumour would have corresponded to a decrease in mRNA transcripts detectable in the serum; however this was not the case. To rule out the possibility that using a volume RNA amount rather than a quantified RNA amount in the amplification process may have resulted in this discrepancy, the corresponding tissue samples were then compared. The tissue samples were compared to see if the differentially expressed gene transcripts reflected the results found in the serum specimens. It was found that 80% of the gene transcripts differentially expressed between the tumour and normal tissues were also down-regulated in the tumour tissue compared to the matched normal, indicating that despite being illogical, the results for the gene transcripts differentially expressed in the serum specimens, reflected the overall trend gene expression of the tissue samples.

A third analysis was then performed to rule out whether the differential gene expression profiles were somehow due to patient specific reasons, the expression of the gene transcripts in the pre serum were compared to the expression of the gene transcripts in the normal serum. The result of that analysis showed that 83% of the gene transcripts differentially expressed in the two samples were down-regulated in the Pre serum specimens compared to the normal serum specimens. As the trend of change was similar i.e. the majority of gene transcripts were up-regulated in the normal serum specimens similar to the Post serum specimens any patient- specific reasons for the unusual gene expression pattern were ruled out.

In the microarray study of extracellular saliva mRNA from oral cancer patients and healthy donors, of the 1,679 transcripts deemed to be significantly different (criteria; fold change >2, p-value <0.05), 836 were up-regulated and 843 were down-regulated in the saliva of cancer patients (Li *et al* 2004(b)). A stricter criteria of fold change >3 was used in selection of gene transcripts to validate and 17 gene transcripts passed this criterion, all of which were down-regulated in the serum of healthy controls.

The microarray study of serum mRNA carried out by Li *et al* (2006) identified 335 gene transcripts significantly different in the comparison of oral cancer patients and healthy controls (criteria for significance; fold change >2, p-value <0.05). Of the 335 gene transcripts significantly differentially expressed, 223 were up-regulated and 112 were down-regulated in the serum of cancer patients compared to healthy controls.

4.8.11.5 Study of Randomly Changing Gene Transcripts

The false discovery rate (FDR) is a measurement of how meaningful your gene list is. Since thousands of genes are compared, many genes can be false positives. dChip estimates the false discovery rate by permutation. Other methods of multiple-comparison adjusted p-values, significance analysis of microarray (SAM) and re-sampling based multiple test methods are also available.

To determine the false discovery rate, dChip randomly permutes samples in group-wise comparisons to determine how many gene transcripts would be obtained by the same comparison if the samples were permuted randomly e.g. in a comparison of four samples vs. four samples the false discovery rate would be assessed by moving the eight samples around randomly to compare any 4 samples vs. any four samples. The default in dChip is to test 50 permutations of your data to find the false discovery rate. A meaningful list is determined if the false discovery rate is less than 10% i.e. for every 100 gene transcripts on the comparison gene list, 10 or less will be falsely discovered.

To take this comparison a little further with the serum specimens, the false discovery rate of each list was determined for the sample groupings that make sense e.g. Pre serum vs. Post serum specimens, and also for permutations of the samples to assess if the scrambled groupings also have the ability to generate gene lists with a low false discovery rate.

The results of the false discovery rate generated for all of the different sample groupings are listed in Tables 3.10.7.1, 3.10.7.2 and 3.10.7.3.

With two exceptions, the only sample groupings to generate false discovery rates that were determined to be meaningful (with a minimum number of 4 samples in each group) were the groupings of either Pre vs. Post specimens, Pre vs. Normal specimens or Pre vs. Post and Normal specimens. The two exceptions to this were in the grouping of 4 Pre vs. 4 Post

specimens with 3 Normal specimens randomly added to each group.

None of the comparisons of mixed groups vs. mixed groups yielded gene lists determined to be meaningful. For example, Table 3.10.7.2 lists the false discovery rates attained when any 2 Pre, 2 Post and 3 Normal specimens are compared to a group of the remaining 2 Pre, 2 Post and 3 Normal specimens. None of the group wise comparisons were determined to be meaningful i.e. have a false discovery rate <10%.

Whilst this was not surprising as the sample groups i.e. Pre serum, Post serum and Normal serum were mixed up. What is surprising is that the comparison of these intact groups to each other did not always generate gene lists determined to be meaningful. For example in Table 3.10.7.3, the comparison of the 4 Post serum specimens to the 6 Normal specimens generated a gene list with a false discovery rate of 32.4% indicating that nearly 1/3 of the gene transcripts on that list could have come through by chance. The comparison of the 4 Pre serum specimens with the 6 Normal specimens however, generated a gene list with a false discovery rate of 6.3% which is classed as meaningful.

To summarise, only the gene lists generated with the 4 Pre serum specimens compared to either the 4 Post serum specimens or the 6 Normal serum specimens, or the 4 Pre serum specimens compared to the combination of the 4 Post serum and 6 Normal serum specimens yielded gene lists that were determined to be meaningful. This indicates that despite the fact that the sample groups could not be separated based on their Presence or Absence call, there are distinct differences that separates the Pre serum specimens from both the Post serum specimens and the Normal serum specimens. This finding is also exciting because the Pre and Post serum specimens obviously originate from the same person. The fact that the specimens are more similar to the normal specimens indicates a huge potential for the development of this work in a clinical setting to find a cancer-specific signature of gene expression.

4.8.11.6 SEQUENCE ANALYSIS

The results shown in section 3.11 indicate that there may be subtle changes in the presence of regulatory sequences in the transcripts significantly differentially regulated in parental CM vs. the drug selected CM samples compared to the number of regulatory sequences found in transcripts called Present in the cell RNA and Absent in the CM RNA.

Transcripts in each of the three groups ((1) gene transcripts called Present in all CM samples, (2) gene transcripts significantly differentially regulated between parental and drug selected variant CM samples and (3) gene transcripts called Present in all cell RNA samples and Absent in all CM samples) were positive for 15- Lipoxygenase Differentiation Control Element (15-LOX-DICE) sequences. 15-LOX-DICE sequences are tandem repeat sequences found in Erythroid 15-Lopoxygnease (15-LOX) capable of binding regulatory proteins HnRNP K and HnRNP E1. Binding of these proteins in 15-LOX specifically inhibits 15-LOX mRNA translation.

Whilst the gene transcripts in group (1) (called Present in all CM samples) contained a mixture of multiple 15-LOX-DICE sequences detected in the transcripts, the other two groups showed a subtle pattern to the amount of multiple 15-LOX-DICE sequences found.

In group (2), 3/6 of the transcripts contained 4 15-LOX-DICE sequences, with the remaining 3 transcripts containing 2, 3 & 6 15-LOX-DICE sequences.

In group (3), 5/8 of the transcripts had 2 15-LOX-DICE sequences, of the remaining 3 transcripts in the group, 2 had 3 15-LOX-DICE sequences and 1 had 5 15-LOX-DICE sequences in their mRNA sequences.

The majority of sequences called Present in the cell RNA and Absent in the CM RNA, that were positive for multiple 15-LOX-DICE sites, contained 2 15-LOX-DICE sequences. The majority of sequences differentially expressed in the parent vs. drug selected variant CM

samples contained 4 15-LOX-DICE sequences.

In addition to this, multiple K-Box regulatory sequences were found only in the transcripts significantly differentially expressed between the parent CM samples and the drug selected variant samples (3/20). K-Box sequences are present in one or more copies of many 3'UTRs and mediate negative post-transcriptional regulation. Regulation by the K-box is spatially and temporally ubiquitous, and likely involves the formation of RNA-RNA duplexes with complementary sequences found at the 5' end of many micro RNAs.

However, whilst these subtle changes are noted, it is quite possible that analysis of larger number of sequences from the three categories would show non- compliance to the results noted here. It is also quite possible that presence of regulatory sequences in these transcripts, whilst interesting to note, do not automatically infer that they are functionally active. The reason this analysis was carried out in the first place was because of the few studies carried out to assess the mechanism by which extracellular RNA remains stable. Most of the studies speculate that binding of the RNA to proteins confers stability of the RNA transcripts. Yet no definite reason for the unexpected stability of the transcripts in CM/ serum/ saliva has been elucidated.

The RNA may possibly be complexed to lipids, proteins, lipoproteins, or phospholipids (Rosi *et al* 1988, Stroun *et al* 1978, Masella *et al* 1989, Wieczorek *et al* 1985); bound with DNA in nucleosomes (Stroun *et al* 1978, Sisco *et al* 2001) or protected within apoptotic bodies (Halicka *et al* 2000) or other vesicular structures.

Wieczorek *et al* (1985) were able to isolate an RNA- proteolipid complex from both patient serum or supernatants from cultured cells that seemed to be associated with a malignant state in humans. The RNA- proteolipid complex required respiration for its appearance in the culture medium as production and/ or secretion was inhibited by cytochalasin B, CN^- and monensin that indicates that the complex is not merely a degradation product of the

cancer cells. They were unable to detect the complex in normal cells even when the homogenates were concentrated. They also found that serum RNase C cleaved the isolated RNA under physiological conditions but not RNA associated with the complex. This suggested that the RNA-proteolipid-complex consisted of RNA within a vesicular lipid complex, and the authors felt that the high lipid content of the structure played a part in the inclusion of the RNA within the complex. Investigation of the RNA extracted, showed that 95-98% of it was poly (A) RNA that had double stranded stretches.

Halicka *et al* (2000) studied the location of cellular DNA and RNA both histochemically and immunocytochemically. Cellular RNA was labelled with BrU for different periods of time and the incorporated BrU was detected using anti- BrU mAb while counterstaining DNA with 7-aminoactinomycin D. Also cellular DNA and RNA were differentially stained with Hoechst 33342 and pyronin Y and visualised using fluorescence microscopy. Apoptosis was induced by camptothecin (CPT- a DNA topoisomerase I inhibitor) in MCF-7 cells. They found that cellular RNA became sequestered and packaged into granules and then into apoptotic bodies, separately from DNA in over 90% of the apoptotic bodies.

Sisco, (2001) exposed ethidium bromide-containing serum to different concentrations of RNaseH to selectively degrade RNA in an RNA-DNA heteroduplex molecule. Fluorescence measurements were then taken using a spectrophotometer. As ethidium bromide fluorescence is dependent on base pairing for the highest quantum yield, any decrease in fluorescence would indicate RNaseH activity and presence of RNA-DNA hybrids. The author found a concentration dependent reduction in fluorescence (10% reduction in fluorescence was obtained after incubating the serum with 120 kU/L of RNase H) leading to the conclusion that at least some of the RNA was present in an RNA-DNA heteroduplex.

Hasselmann *et al* (2001) showed that tumour-derived mRNA associated with apoptotic

bodies remained stable in serum. This was possible by culturing a melanoma cell line and inducing apoptosis in the cells using anti- CD95 monoclonal antibody. The cell supernatant was removed after 48 hours exposure and passed through a 0.45μm filter. They then confirmed that the supernatant contained apoptotic bodies and also that RNA was present in some of the apoptotic bodies. Baseline readings of nucleic acid content of the serum were established, and then the supernatant was then mixed with serum from controls for time points ranging from 1 minute to 60 minutes at 37°C. The RNA was subsequently extracted from this mix and also from a separate sample of serum that had 1μg of free/ "naked" RNA (from the melanoma cell line) added. They were able to successfully amplify tyrosinase mRNA from the serum that was mixed with the cell supernatant for time points of 1 minute to 30 minutes but not from the 60 minute sample or from any of the serum specimens that had been mixed with free RNA. This lead the authors to conclude that the apoptotic body structure protected the RNA from degradation.

Tsui *et al* (2002) also showed the stability of plasma/ serum RNA by comparing yields of RNA from plasma/ serum obtained from healthy volunteers. They too found that free RNA added to plasma was rapidly degraded. In the first five seconds after addition of the commercially available extracted RNA, the concentration decreased from 67,605ng/L to 116ng/L and >99% of the RNA was undetectable after 15 seconds incubation. The study also included filtration studies whereby concentrations of endogenous RNA in plasma/ serum were quantified before and after filtration with different size filters and also over time post filtration. They concluded that that most of the RNA was particle- bound in some way that stabilised it and protected it from degradation.

A more recent study by El-Hefnawy *et al* (2004) sought to determine the nature of this protection. They tested the theory that the RNA is protected because it is part of a DNA-RNA hybrid. After addition of RNase-H to plasma samples (which is the only nuclease that

such a hybrid would not be resistant to), they found that there was no difference in the concentrations of RNA extractable from the samples. They also tested the theory that the RNA is protected by inclusion in lipoprotein vesicles that are either actively secreted or released during programmed cell death. Detergents were added (SDS or Triton-X) in an attempt to disrupt any RNA-protein or RNA-lipid complexes. After the addition of the detergents the authors were no longer able to amplify 18SrRNA, which favoured the theory that the RNA is protected in a lipid-containing vesicle.

Studies by Park *et al* (2006) on saliva RNA also revealed that incubation of saliva with Triton X-100 accelerated degradation of the RNA. In addition to this, they tested the recovery of saliva RNA after filtration through 0.22 or 0.45μm pore filters. This was carried out also on purified total RNA as a control. The results showed that >80% of β-actin in the control RNA was recoverable post filtration however, less than 5% of β-actin mRNA was recoverable after filtration of the saliva through the 0.22 and 0.45μm pore filters. Analysis of serum RNA also revealed a significant loss in β-actin mRNA post-filtration. Approximately 8% was recoverable after filtration through a 0.45μm filter and less than 1% was recoverable after filtration through a 0.22μm filter.

Different elements in the 5' and 3' UTR region, like the regulatory elements found in the CM RNA sequences as shown in Section 3.11, are known to confer either stability or instability to an RNA transcript. Examples of these include the occurrence of upstream AUG, upstream ORFs and IRES in 5'UTRs and Poly-A tail in the 3'UTRs which can increase transcript stability and AU-Rich Elements (AREs) in the 3'UTR of RNA sequence which can decrease stability.

In addition to this, different RNA-binding proteins are involved in the regulation of splicing and processing, export of mRNA to the cytoplasm, maintenance of mRNA in the cytoplasm

for translation, and ultimately decay of the mRNA.

There are RNA-binding proteins that shuttle between the nucleus and the cytoplasm with a few of these staying with an mRNA molecule for its lifespan, whilst other RNA-binding proteins are transiently associated with the mRNA. Some RNA-binding proteins have been shown to play a role in the regulation of gene expression by stabilising or destabilising a particular mRNA. A number of stimuli can alter the affinity of RNA-binding proteins for their target *cis*-elements, including hormones and cytokines, redox changes, UV light, cell cycle and developmental stage (Hollams *et al* 2002). Competition between protection or degradation factors or translation activators/ inhibitors will ultimately determine the overall amount of protein produced by one mRNA.

A number of RNA-binding motifs have been characterised like the RNA recognition motif (RRM), which is 90-100 amino acids (aa) in length and often present in one or more copies in the protein sequence. Other binding motifs include the RGG box, the K homology motif (KH), the double-stranded RNA-binding motif (DSRM), the zinc finger/knuckle domain and the cold shock domain (Hollams *et al* 2002).

To summarise, when we started the work shown in this thesis, no proven methods were known for the reproducible and reliable extraction and detection of extracellular mRNA transcripts. It was also not known whether these extracellular mRNA transcripts were present as intact full-length transcripts or whether the detection of these transcripts resulted from cell apoptosis/ non-selective secretion.

To this end, we have answered all of these questions and developed reliable and reproducible methods for the extraction of these mRNA transcripts and subsequent analysis by RT-PCR. PCR primers that amplified either different regions of the same transcript or 5' regions of different transcripts were designed. Use of these primers in conjunction with an oligo dT reverse transcription primer in RT-PCR studies enabled us to confirm the presence of full-length transcripts. As CM samples were taken from healthy viable cultured cells, any interference by apoptosing cells was minimal.

In addition to this, we successfully applied 2 cycle amplification protocols to extracellular RNA isolated from both serum and CM samples to enable global expression analysis of these samples to be performed. No studies have been published regarding the microarray analysis of extracellular RNA from CM samples so this work is very novel. One serum microarray study has since been published relating to microarray analysis of samples taken from oral cancer patient serum and healthy patient serum, however our study design of taking serum specimens pre and post lumpectomy surgery in addition to serum specimens from normal volunteers adds another novel dynamic range to this type of study. The success of this study indicates the possibility that this work could lead to useful diagnostic/ prognostic tests based on gene expression profiles obtained from analysis of the microarray data.

However this work, although having been brought to this stage is still only at an early stage of development. The results need to be confirmed for interesting biomarker candidates. A

much larger study would also need to be performed and results confirmed by RT-PCR methods before application in a clinical environment could be established.

Whilst analysis of the sequences of mRNAs detectable in CM samples compared to those found in cell RNA samples and Absent in CM samples did not identify any significant leads, this work too is in its infancy. Future work is needed on mRNA transcripts confirmed by RT-PCR to be detected in CM to elucidate if the regulatory elements identified in their sequence are in fact functionally active i.e. are they are actively involved in stabilising these transcripts extracellularly. In addition to this, protein-RNA interaction studies should be carried out to assess what role, if any, proteins may have on the stability of the mRNA transcripts. Studies should also be performed to elucidate how the mRNA gets out of the cell.

To conclude we now know that despite the low RNA yields and challenges involved in quantifying and amplifying extracellular mRNA, we can reproducibly amplify and detect extracellular mRNA transcripts by RT-PCR. We also now know that microarray studies on extracellular mRNA from both CM and serum are possible. This has lead to the exciting and very real possibility that methods developed during the course of this study may aid in the selection of cancer biomarkers that would allow on-going/ sequential monitoring of the course of disease (progression, response to therapy, *etc.*).

4.9 SUMMARY

We have demonstrated that many gene transcripts from a broad range of healthy, proliferating, human cancer cell lines can be reproducibly detected in the CM of these cell lines using the protocols developed. The RT-PCR results suggest that the detection of particular transcripts by a given cell line may be somewhat selective. The mRNA found in the CM of cancer cells are apparently full-length transcripts; some of which may be particle-bound.

The utilisation of linear amplification techniques, allowing the CM and serum mRNA to be amplified to amounts required for hybridisation to array chips, enabled global expression analysis of the mRNA found in these samples. Despite the problems encountered when validating the microarray data using RT-PCR, the fact is distinct patterns of expression were achievable for many of the sample groups tested using the microarray data on its own. This indicates the potential for development of disease-specific/ stage-specific/ response-to-therap- specific expression patterns applicable to a clinical environment.

The function of extracellular mRNA remains unclear along with the mechanism by which it remains stable, and also how it gets out of the cell. However its' potential as a future biomarker does not rely on elucidating these roles and Absence/ Presence of transcripts at a particular expression level may be enough to qualify as a diagnostic/ prognostic or predictive biomarker.

4.10 CONCLUSIONS

1. Extracellular mRNA is readily extractable from biological samples using a modified Tri Reagent extraction protocol.
2. The presence of extracellular mRNA in CM samples may be the result of a selective secretion process, as gene transcripts were not always detectable in all CM samples from different cell lines even if the gene was expressed in the different cell types.
3. The mRNAs are derived from the cells and not from the serum in which they are cultured.
4. Gene transcripts are present as full- length mRNAs.
5. The use of different filter sizes affected the abundance of certain mRNA transcripts detected, suggesting that at least some of the extracellular mRNA is likely to be particle- associated in some form.
6. Bioanalyser RNA chips were unable to detect any intact ribosomal RNA, but were able to identify the presence of RNA fragments up to 1000 nucleotides in length from concentrated CM samples.
7. A reliable method for quantification of RNA extracted using Tri Reagent where the sample cannot be diluted, has yet to be found.
8. Extracellular mRNA from CM and serum specimens can be linearly amplified using two rounds of amplification to yield enough cRNA for hybridisation to Affymetrix Human Genome Plus 2.0 microarray chips.
9. Whilst the microarray mRNA profiles of CM mRNA from a range of cancer cell lines/ cancer types are quite similar, distinct sample- specific profiles were achievable for the majority of the samples tested.
10. Although attempts to validate the microarray data using RT-PCR was only partially

successful (approx 50%), it did not take away from the value of the data as a microarray study on its own, as the techniques are inherently different and work on different principles.

11. Application of this microarray technology to serum specimens obtained from the same patient before and after lumpectomy was possible.
12. Despite the small number of patients tested, meaningful gene lists were generated from the groupings of Pre-op specimens against Post-op specimens and normal specimens, and also from the groupings of Pre-op specimens against the normal specimens as defined by their low false discovery rates.
13. The microarray studies shown in this thesis indicate the potential for this technology to be applied to a clinical environment. The matching of distinct gene expression patterns to different diseases or disease states may ultimately aid in the early detection of disease and prediction of prognosis/ response to therapy by monitoring of expression profiles taken during treatment.
14. The function of extracellular mRNA is still unclear, as is the mechanisms by which it remains stable and how it gets out of the cell.

5 FUTURE WORK

5.1 Validation of Breast Cancer Serum Study Microarray Results using RT-PCR/ qPCR

The results of the Breast Cancer Serum Study (see Section 3.10) did not include any validation of the results by a second method e.g. RT-PCR. Some gene transcripts of interest should be chosen for validation by RT-PCR or qPCR (e.g. transcripts expressed highly in the Pre serum specimens and down-regulated in the Post/ Normal serum specimens). The transcripts chosen for validation should be chosen whilst keeping in mind the information obtained from the validation of the CM microarray work, (i.e. choose transcripts called Present in all specimens that are also functionally annotated).

The study should also be enlarged to include serum specimens from many more patients and matched controls. Ideally the samples would be analysed by microarray also, however if this was not possible (e.g. due to the high cost of such an experiment), the serum samples could be tested for the Presence/ expression of any gene transcripts validated in the smaller study to be of interest (i.e. as possible biomarkers).

In addition to this, serum should be collected from a variety of cancer patients and patients with non-malignant diseases together with age/sex matched controls and subjected to microarray analysis using whole genome microarray chips. This would help to establish whether potential biomarkers not only diagnose cancer from normal/ non-malignant disease serum specimens, or predict response to therapy, but also distinguish different types of cancer. It would be also important to include serum from patients with cancers that are typically diagnosed at a late stage e.g. lung or kidney, as with these cancers in particular, late diagnosis often leads to a poorer prognosis for the patients. If a marker could be identified that would indicate development of cancer at an early stage this would be invaluable.

5.2 Microarray Analysis of Conditioned Media from normal cells

The pilot study performed using CM from one batch of HMEC- 1001 breast cells was successful in so far as the yield of labelled cRNA was sufficient to permit hybridisation to whole genome microarrays, and the QA parameters were all acceptable.

The fact that we were able to amplify enough cRNA for hybridisation to the Affymetrix chips means that we can plan a larger study to culture cells from various tissue sources e.g. breast, lung, nasal, *etc* and build up a profile of the gene expression of mRNA detectable from the CM of these cultured cells, including as many biological replicates as desired. The study would be directly comparable to the CM array work described in this thesis.

It may also be possible to subtract the expression of mRNAs found in the CM of normal cells from those found in the cancer cell lines to help to build a picture of cancer-specific mRNAs detectable in the CM of cancer cell lines. In the pilot study this was carried out but as only one CM sample from the normal cells was available, statistical significance could not be assessed.

In addition to this, CM from larger panels of drug resistant/ non-resistant, invasive/ non-invasive samples could be included. Whilst drug selected variants were studied in the work presented here, the numbers were too small and the samples too varied to read too much into the results. It may be possible, for example, with a larger panel of invasive and non-invasive CM samples that markers could be identified that would be indicative of an invasive phenotype.

5.3 Determination of RNA-protein interactions

One conclusion in this thesis is that at least some of the RNA is likely to be is particle-bound. Future studies could, therefore, take this further to try to elucidate if the RNA is bound to proteins which are protecting it from degradation.

Many elements that aid in the translation and regulation of a gene transcript are regulated by cis-elements found in the 3'UTR. In 2002, Hartley *et al* described a technique that could identify sequence-specific RNA binding proteins (RNA-BPs) by using UV-crosslinking methods. This method involves cloning the mRNA sequence of interest (e.g. in Hartley *et* al the 3'UTR sequence of EG and c-mos *Xenopus laevis* maternal mRNAs) into plasmids that have been modified to enable in vitro transcription of the sequence using a T3/ T7 promoter. For in vitro transcription the plasmids were linearized and ^{32}P-labeled uncapped transcripts transcribed using a Riboprobe transcription kit (Promega). The ^{32}P-labeled transcripts were then incubated with cytoplasmic extracts (prepared from 4 hour Xenopus embryos) before being exposed to UV light. Exposure to UV light causes protein to be covalently linked to RNA. The RNA-protein complex was then digested with RNase A to degrade any single stranded RNA i.e. unbound. The results of the reaction were assessed by resolving the entire reaction by SDS-PAGE on a 10% Laemmli gel and visualising the fixed and dried gel by autoradiography. Only proteins with bound radiolabelled transcripts (i.e. our in vitro transcribed transcript of interest) would be visible on the gel.

To adapt this protocol for our work, the procedure would be carried out as above with one exception. Instead of incubating in a cytoplasmic extract, the ^{32}P-labeled transcripts could be incubated in conditioned media so that only proteins endogenously found in CM would be targeted. As it has also been suggested that cell-free mRNA may also be sequestered within vesicles, methods may have to be developed that would reversibly disrupt any

vesicles in the CM long enough for the radiolabelled RNA to get in. Although this method would not identify the protein (if any) that binds the RNA, further analysis using proteomic techniques may aid in the identification of the protein.

6 BIBLIOGRAPHY

- Alaiya, A., Al-Mohanna, M., Linder, S. Clinical cancer proteomics: promises and pitfalls. *J Proteome Res.* 2005 **4** 1213-22
- Anderson, K.S., LaBaer, J The Sentinel Within: Exploiting the Immune System for Cancer Biomarkers *Journal of Proteome Research* (2005) **4** 1123-1133
- Andriani, F., Conte, D., Mastrangelo, T., Leon, M., Ratcliffe, C., Roz, L., Pelosi, G., Goldstraw, P., Sozzi, G., Pastorino, U. Detecting lung cancer in plasma with the use of multiple genetic markers. *Int J Cancer* 2004 **108** (1) 91-**6**
- Aquino, A., Formica, V., Prete, S.P., Correale, P.P., Massara, M.C., Turriziani, M., De Vecchis, L., Bonmassar, E. Drug-induced increase of carcinoembryonic antigen expression in cancer cells. *Pharmacol Res.* 2004 **49** 383-96
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J. M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A.,. Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S, Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000 **1**:25- 29.
- Attallah, A.M., Sakr, H.A., Ismail, H., Ismail, M.F., Ibrahim, A.S., El-Sharabasy, M.M., El-Dosoky, I. Rapid diagnosis and follow up of bladder cancer patients using urinary high molecular weight cytokeratins. *World J Urol* 2006
- Bast, R.C., Feeney, M. Lazarus, H., Nadler, L.M., Colvin, R.B., Knapp, R.C., Reactivity of a monoclonal antibody with human ovarian carcinoma. *J Clin Invest* 1981 **68** 1331-37
- Beau-Faller, M., Gaub, M.P., Schneider, A., Ducrocq, X., Massard, G., Gasser, B., Chenard, M.P., Kessler, R., Anker, P., Stroun, M., Weitzenblum, E., Pauli, G., Wihlm, J.M., Quoiz, E., Oudet, P. Plasma DNA microsatellite panel as sensitive and tumor-specific marker in lung cancer patients. *Int J Cancer* 2003 **105** (3) 361-70
- Bianchi, D.W., Avent, N.D., Costa, J.M., van der Schoot, C.E. Noninvasive prenatal diagnosis of fetal Rhesus D: ready for Prime(r) Time.*Obstet Gynecol.* 2005 **106** 841-4.
- Bozionellou, V., Mavroudis, D., Perraki, M., Papadopoulos, S., Apostolaki, S., Stathopoulos, E., Stathopoulou, A., Lianidou, E., Georgoulias, V. Trastuzumab Administration Can Effectively Target Chemotherapy-Resistant Cytokeratin-19 Messenger RNA-Positive Tumor Cells in the Peripheral Blood and Bone Marrow of

Patients With Breast Cancer. *Clin Cancer Res* 2004 **10** 8185-8194

- Bruhn, N., Beinert, T., Oehm, C., Jandrig, B., Petersen, I., Chen, X.Q., Possinger, K., Fleischhacker, M. Detection of Microsatellite Alterations in the DNA Isolated from Tumor Cells and from Plasma DNA of Patients with Lung Cancer. *Ann NY Acad Sci* 2000 **906** 72-82
- Butler, S.A., Iles, R.K. Ectopic Human Chorionic Gonadotropin β Secretion by Epithelial Tumors and Human Chorionic Gonadotropin β-induced Apoptosis in Kaposi's Sarcoma: Is There a Connection? *Clin Cancer Res* 2003 **9** 4666-4673
- Capone, R.B., Pai, S.I., Koch, W.M., Gillison, M.L., Danish, H.N., Westra, W.H., Daniel, R., Shah, K.V., Sidransky, D. Detection and quantitation of human papillomavirus (HPV) DNA in the sera of patients with HPV- associated head and neck squamous cell carcinoma. 2000 *Clin Cancer Res* 6 4171- 5
- Chan, K.C.A., Zhang, J., Chan, A.T.C., Lei, K.I.K., Leung, S-F., Chan, L.Y.S., Chow, K.C.K, Lo, Y.M.D. Molecular Characterization of Circulating EBV DNA in the Plasma of Nasopharyngeal Carcinoma and Lymphoma Patients. *Cancer Res* 2003 **63** 2028-2032
- Chang, H-W., Shing, M.L., Goodman, S.N., Singer, G., Cho, S.K.R., Sokoll, L.J., Montz, F.J., Roden, R., Zhang, Z., Chan, D.W., Kurman, R.J., Shih, I-M. Assessment of Plasma DNA levels, Allelic Imbalance, and CA 125 as Diagnostic Tests for Cancer *J Nat Cancer Inst* 2002 **94** (22) 1697-1703
- \Chen XQ, Bonnefoi H, Pelte M-F, Lyautey J, Lederrey C, Movaeekhi S, Schaeffer P, Mulcahy HE, Meyer P, Stroun M and Anker P: Telomerase RNA as a detection marker in the serum of breast cancer patients. Clin Cancer Res 6: 3823-3826, 2000.
- Chen, X.Q., Bonnefoi, H., Diebold-Berger, S., Lyautey, J., Lederrey, C., Faltin-Traub, E., Stroun,M., Anker, P. Detecting Tumor-related Alterations in Plasma or Serum DNA of Patients Diagnosed with Breast Cancer *Clin Cancer Res* 1999 **5** 2297-2303
- Chen, X.Q., Stroun, M., Magnenat, J.L., Nicod, L.P., Kurt, A.M., Lyautey, J., Lederrey, C., Anker, P. Microsatellite alterations in plasma DNA of small cell lung cancer patients. *Nat Med* 1996 **2** (9) 1033-5
- Cheung, K.L., Graves, C.R., Robertson, J.F. Tumour marker measurements in the diagnosis and monitoring of breast cancer. *Cancer Treat Rev.* 2000 **26** 91-102

- Codony-Servat, J., Albanell, J., Lopez-Talavera, J.C., Arribas, J., Baselga, J. Cleavage of the HER2 ectodomain is a pervanadate-activable process that is inhibited by the tissue inhibitor of metalloproteases-1 in breast cancer cells. *Cancer Res* 1999 **59** (6) 1196-201
- Coletta, R.D., Almeida, O.P., Vargas, P.A. Cytokeratins 1, 7 and 14 immunoexpression are helpful in the diagnosis of basaloid squamous carcinoma. *Histopathology* 2006 **48** 773-774
- Combaret, V., Audoynaud, C., Iacono, I., Favrot, M-C., Schell, M., Berferon, C., Puisieux, A. Circulating *MYCN* DNA as a Tumor-specific Marker in Neuroblastoma Patients. *Cancer Res* **63** 3646-3648
- Dallas, P.B., Gottardo, N.G., Firth, M.J., Beesley, A.H., Hoffmann, K., Terry, P.A. *et al* Gene expression levels assessed by oligonucleotide microarray analysis and quantitative real- time RT-PCR- how well do they correlate? *BMC Genomics* 2005 **6** 59-69
- Daniele, B., Bencivenga, A., Megna, A. S., Tinessa, V. α-Fetoprotein and Ultrasonography Screening for Hepatocellular Carcinoma. *Gastroenterology* 2004 **127** S108-112
- dChip User's manual (Oct 2005) *Available from:* http://www.dchip.org/
- Demel, U., Tilz, G.P., Foeldes-Papp, Z., Gutierrez, B., Albert, W.H., Böcher, O. Detection of Tumor Cells in the Peripheral Blood of Patients with Breast Cancer. Development of a New Sensitive and Specific Immunomolecular Assay. *J. Exp. Clin. Cancer Res.* 2004 **23** 465-468
- Ding, X., Yang, L. Y., Huang, G.W., Yang, J.Q., Liu, H.L., Wang, W., Peng, J.X., Yang, J.Q., Tao, Y.M., Chang, Z.G., Ling, X.S. Role of AFP mRNA expression in peripheral blood as a predictor for postsurgical recurrence of hepatocellular carcinoma: A systematic review and meta-analysis 2005 *Aorld J Gastroenterol* **11** 2656-2661
- Draghici, S., Khatri, P., Eklund, A.C., Szallasi, Z. Reliability and reproducibility issues in DNA microarray measurements. *Trends Genet* 2005 **22** 101-109
- Duffy, M.J. Carcinoembryonic antigen as a marker for colorectal cancer: is it clinically useful? *Clin Chem.* 2001 **47** 624-30
- El-Hefnawy T, Raja S, Kelly L, Bigbee WL, Kirkwood JM, Luketich JD and

Godfrey TE: Characterization of amplifiable, circulating RNA in plasma and its potential as a tool for cancer diagnostics. Clin Chem 50: 564-573, 2004.

- Emlet, D.R., Schwartz, R., Brown, K.A., Pollice, A.A., Smith, C.A., Shackney, S.E. HER2 expression as a potential marker for response to therapy targeted to the EGFR. *Br J Cancer* 2006 **94** 1144 -1153
- Esteller, M., Sanchez-Cespedes, M., Rosell, R., Sidransky, D., Baylin, S.B., Herman, J.G. Detection of Aberrant Promoter Hypermethylation of Tumor Suppressor Genes in Serum DNA from Non-Small Cell Lung Cancer Patients. *Cancer Res* 1999 59 67-70
- Etienne, W. Meyer, M.H., Peppers, J. Meyer Jr, R.A., Comparison of mRNA expression by RT-PCR and DNA microarray. *Biotechniques* 2004 **36** 618- 626
- Fleischhacker M, Beinert T, Ermitsch M, Seferi D, Possinger K, Engelmann C and Jandrig B: Detection of amplifiable messenger RNA in the serum of patients with lung cancer. *Ann N Y Acad Sci* 2001 **945**: 179-188
- Fletcher, R.H. Carcinoembryonic antigen. *Ann Int Med* 1986 **104** 66-73
- Foon, K.A., John, W.J., Chakraborty, M., Sherratt, A., Garrison, J., Flett, M., Bhattacharya-Chatterjee, M. Clinical and immune responses in advanced colorectal cancer patients treated with anti-idiotype nonoclonal antibody vaccine that minics the carcinoembryonic antigen. *Clin Cancer Res* 1997 **3** 1267-76
- Gacci, M., Serni, S., Lapini, A., Giubilei, G., Dal Canto, M., Paladini, S., Curotto, A., Gallo, F., Carmignani, G., Carini, M. Pre and postoperative quantitative detection of fragments of cytokeratins 8 and 18 (UBC IRMA) as markers of early recurrence of superficial bladder tumor. *Arch Ital Urol Androl* 2006 **78** 5-10
- Gadducci, A., Cosio, S., Carpi, A., Nicolini, A., Genazzani, A.R. Serum tumor markers in the management of ovarian, endometrial and cervical cancer. *Biomedicine & Pharmacotherapy* 2004 **58** 24-38
- Gal, S., Fidler, C., Lo, Y.M.D., Taylor, M., Han, C., Moore, J., Harris, A.L., Wainscoat, J.S. Quantitation of circulating DNA in the serum of breast cancer patients by real-time PCR. *Br J Cancer* 2004 **90** 1211-1215
- Galizia, G., Lieto, E., Ferraraccio, F., DeVita, F., Castellano, P., Orditura, M., Imperatore, V., LaMura, A., LaManna, G., Pinto, M., Catalano, G., Pignatelli, C., Ciardiello, F. Prognostic Significance of Epidermal Growth Factor Receptor

Expression in Colon Cancer Patients Undergoing Curative Surgery. *Annals of Surgical Oncology* 2006 **13** 823-35

- Ginzinger, D.G. Gene quantification using real- time quantitative PCR: An emerging technology hits the mainstream. *Exp Hematology* 2002 **30** 503- 512
- Glynn SA, Gammell P, Heenan M, O'Connor R, Liang Y, Keenan J and Clynes M: A new superinvasive *in vitro* phenotype induced by selection of human breast carcinoma cells with the chemotherapeutic drugs paclitaxel and doxorubicin. *Br J Cancer* 2004 **91** 1800-1807
- Gocke, C.D., Benko, F.A., Kopreski, M.S., McGarrity, T.J. p53 and *APC* Mutations Are Detectable in the Plasma and Serum of Patients with Colorectal Cancer (CRC) or Adenomas *Ann NY Acad Sci* 2000 **906** 44-50
- Gonzalez, R., Silva, J.M., Sanchez, A., Dominguez, G., Garcia, J.M., Chen, X.Q., Stroun, M., Provencio, M., Espana, P., Anker, P., Bonilla, F. Microsatellite alterations and TP53 mutations in plasma DNA of small-cell lung cancer patients: follow-up study and prognostic significance. *Ann Oncol* 2000 **11** (9) 1097-1104
- Greiner, J.W., Guadagni, F., Noguchi, P., Pestka, S., Colcher, D., Fisher, P.B., Schlom, J.Recombinant interferon enhances monoclonal antibody-targeting of carcinoma lesions in vivo. *Science* 1987 **235** 895-898
- Gross-Groupil, M., Saffroy, R., Azoulay, D., Precetti, S., Emile, J.F., Delvart, V., Tindiliere, F., Laurent, A., Bellin, M.F., Bismuth, H., Debuire, B., Lemoine, A. Real-time quantification of AFP mRNA to assess hematogenous dissemination after transarterial chemoembolization of hepatocellular carcinoma. *Ann Surg* 2003 **238** 241-248
- Guo, J., Ziao, B., Jin, Z., Qin, L., Chen, J., Chen, H., Zhang, X, Liu, Z. Detection of cytoderatin 20 mRNA in the peripheral blood of patients with colorectal cancer by immunomagnetic bead enrichment and real-time reverse transcriptase-polymerase chain reaction. *J Gastroenterol Hepatol* 2005 **20** 1279-84
- Halicka, H.D., Bedner, E., Darzynkiewicz, Z. Segregation of RNA and Separate Packaging of DNA and RNA in Apoptotic Bodies during Apoptosis *Exp Cell Res* 2000 **260** 248- 256
- Hartley, H., Le Meuth-Metzinger, V., Osborne, H.B. Screening for sequence-specific RNA-BPs by comprehensive UV crosslinking. *BMC Molecular Biology*

2002 3 8

- Hasselmann, D.O., Rappl, G., Tilgen, W., Reinhold, U. Extracellular Tyronsinase mRNA within Apoptotic Bodies Is protected from Degradation in Human Serum *Clin Chem* **47** 1488-89
- Hibi, K., Taguchi, M., Nakayama, H., Takase, T., Kasai, Y., Ito, K., Akiyama, S., Nakao, A. Molecular Detection of *p16* Promoter Methylation in the Serum of Patients with Esophageal Squamous Cell Carcinoma. *Clin Cancer Res* 7 3135-3138
- Ho, C.M., Yang, S.S., Chien, T.Y., Huang, S.H., Jeng, C.J., Chang, S.F. Detection and quantitation of human papillomavirus type 16, 18 and 52 DNA in the peripheral blood of cervical cancer patients. *Gynecologic Oncology* 2005 **99** 615-621
- Hollams, E. M., Giles, K.M., Thomson, A.M., Leedman, P. J. mRNA stability and the control of gene expression: Implications for human disease. *Neurochemical Research* 2002 **27** (10) 957-980
- Honda, H., Miharu, N., Ohashi, Y., Samura, O., Kinutani, M., Hara, T. Ohama, K Fetal gender determination in early pregnancy through qualitative and quantitative analysis of fetal DNA in maternal serum *Hum Genet* **110** 75-79
- Hsu, K-F., Huang, S-C., Hsiao, J-R., Cheng, Y-M., Wang, S.P.H., Chou, C-Y. Clinical Significance of Serum Human Papillomavirus DNA in Cervical Carcinoma. *Obstet Gynecol* 2003 **102** 1344-1351
- Ijichi, M., Takayama, T., Matsumura, M., Shiratori, Y., Omata, M., Makuuchi, M alpha-Fetoprotein mRNA in the circulation as a predictor of postsurgical recurrence of hepatocellular carcinoma: a prospective study. *Hepatology* 2002 **35** 853- 860
- Ikeda, S., Fujimori, M., Shibata, S., Okajima, M., Ishizaki, Y., Kurihara, T., Miyata, Y., Iseki, M., Shimizu, Y., Tokumoto, N., Ozaki, S., Asahara, T. Combined immunohistochemistry of beta-catenin, cytokeratin 7, and cytokeratin 20 is useful in discriminating primary lung adenocarcinomas from metastatic colorectal cancer. *BMC Cancer* 2006 **2** 31
- Iles, R.K., Lee, C.L., Oliver, R.T.D., Chard, T. Composition of intact hormone and free subunits in the hCG-like material found in serum and urine of patients with carcinoma of the bladder. *Clin Endocrinol* 1990 **32** 355-364
- Johnson, P.J., Leung, N., Cheng, P., Welby, C., Leung, W.T., Lau, W.Y., Yu, S., Ho, S. 'Hepatoma-specific' alphafetoprotein may permit preclinical diagnosis of

malignant change in patients with chronic liver disease. *Br J Cancer.* 1997 **75** 236-40

- Johnson, P.J., Poon, T.C., Hjelm, N.M., Ho, C.S., Blake, C.S., Ho, S.K. Structures of disease-specific serum alpha-fetoprotein isoforms. *Br J Cancer* 2000 **83** 1330-1337
- Johnson, P.J., Poon, T.C.w., Hjelm, N.M., Ho, C.S., Blake, C., Ho, S.K.W. Structures of disease-specific serum α-fetoprotein. *Br J Cancer* 2000 **83** 1188-95
- Kent, W.J. BLAT- The BLAST- Like Alignment Tool. *Genome Res* 2002 **12** 656-664
- Kopreski MS, Benko FA and Gocke CD: Circulating RNA as a tumour marker. Detection of 5T4 mRNA in breast and lung cancer patient serum. *Ann N Y Acad Sci* 2001 **945** 172-178
- Kopreski, M., Benko, F., Kwak, L., Gocke, C. Detection of Tumor Messenger RNA in the Serum of Patients with Malignant Melanoma. *Clin Cancer Res* 1999 **5** 1961-1965,
- Kopreski, M., Benko, F., Kwee, C., Leitzel, K., Eskander, E., Gocke, C. Detection of mutant K-ras DNA in plasma or serum of patients with colorectal cancer. *Br J Cancer* 1997 **76** 1293-9
- Kopreski, M.S., Benko, F.A., Kwee, C., Leitzel, K.E., Eskander, E., Lipton, A., Gocke, C.D. Detection of mutant K-ras DNA I plasma or serum of patients with colorectal cancer *Br J Cancer* 1997 **76** 1293-1299
- Köstler, W., Schwab, B., Singer, C.F., Neumann, R., Rücklinger, E., Brodowicz, T., Tomek, S., Niedermayr, M., Hejna, M., Steger, G.G., Krainer, M., Wiltschke, C., Zielinski, C.C. Monitoring Of Serum Her-2/*Neu* Predicts Response And Progression-Free Survival To Trastuzumab-Based Treatment In Patients With Metastatic Breast Cancer. *Clin Cancer Res* 2004 **10** 1618- 1624
- Koyanagi, K., Kuo, C., Nakagawa, T., Mori, T., Ueno, H., Lorico, A.R., Wang, H-J., Hseuh, E., O'Day, S.J., Hoon, D.S.B. Multimarker Quantitave Real-Time PCR Detection of Circulating Melanoma Cells in Peripheral Blood: Relation to Disease Stage in Melanoma Patients. *Clin Chem* 2005 **51** 981-988
- Law E, Gilvarry U, Lynch V, Gregory B, Grant G and Clynes M: Cytogenetic comparison of two poorly differentiated human lung squamous cell carcinoma lines.

Cancer Genet Cytogenet 1992 **59**111-118

- Leon, S.A., Shapiro, B., Sklaroff, D.M., Yaros, M.J., Free DNA in the serum of cancer patients and the effect of therapy 1997 *Cancer Res* **37** 646-650
- Li, C., Wong, W.H. Model-based analusis of oligpnucleotied arrays: model validation, design issues and standard error application. *Genome Biology*2001(b) **2** 0032.1
- Li, C., Wong, W.H. Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection 2001 *PNAS* (a) **98** 31-36
- Li, D., Mallory, T., Satomura, S. AFP-L3: a new generation of tumor marker for hepatocellular carcinoma 2001 *Clin Chim Acta* **313** 15-19
- Li, Y., Elashoff, D., Oh, M., Sinha, U., St John, M.A.R., Zhou, X., *et al* Serum Circulating Human mRNA Profiling and Its Utility for Oral Cancer Detection 2006 *J Clin Oncol* **24** *in press.*
- Li, Y., Maie, A.R.St.John, Zhou,Z., Kim, Y., Sinha,U., Jordan, R.C.K. Eisele, D., Abemayor, E., Elashoff, D., Park, N.H., Wong, D.T. Salivary Transcriptome Diagnostics for Oral Cancer Detection. *Clin Can Res* 2004 **10** 8442- 8450
- Li, Y., Zhou, X., St John, M.A.R., Wong, D.T.W. RNA Profiling of Cell-free Saliva Using Microarray Technology. *J Dent Res* 2004 **83** 199-203
- Liang Y, Meleady P, Cleary I, McDonnell S, Connolly L and Clynes M: Selection with melphalan or paclitaxel (Taxol) yields variants with different patterns of multidrug resistance, integrin expression and *in vitro* invasiveness. *Eur J Cancer* 2001 **37** 1041-1052
- Liefers, G.J., Cleton-Janson, A.M., Van de Veklde, C.J., Hermans, J., Van Krieken, J.H., Cornelisse, C. J., Tollenaar, R.A. Micrometastases and survival in stage II colorectal cancer. *N Engl J Med* 1998 **339** 223-8
- Lightfoot, S. Quantitation comparison of total RNA using the Agilent 2100 bioanalyzer, ribogreen analysis and UV spectrometry 2002 publication number 5988-7650EN
- Lin, J.C., Chen, K.Y., Wang, W.Y., Jan, J.S., Liang, W.M., Wei, Y.H. Evaluation of cytokeratin-19 mRNA as a tumor marker in the peripheral blood of nasopharyngeal carcinoma patients receiving concurrent chemoradiotherapy. *Int J Cancer* 2002 **97** 548-553

- Lo, K-W. Lo, Y.M., Leung, S-F., Tsang, Y-S., Chan, L.Y.S., Johnson, P.J., Hjelm, M., Lee, J.C.K., Huang, D.P. Analysis of Cell-free Epstein-Barr Virus-associated RNA in the Plasma of Patients with Nasopharyngeal Carcinoma. *Clin Chem* 1999 **45** 1292-1294
- Lo, Y.M.D., Chan, L.Y.S., Lo, K-W., Leung, S-F., Zhang, Z., Chan, A.T.C., Lee, J.C.K., Hjelm, M., Johnson, P.J., Huang, D.P. Quantitative Analysis of Cell-free Epstein-Barr Virus DNA in Plasma of Patients with Nasopharyngeal Carcinoma. *Cancer Res* 1999 **59** 1188-1191
- Lo, Y.M.D., Corbetta, N., Chamberlain, P.F., Rai, V., Sargent, I.L., Redman, C.W.G., Wainscoat, J.S. Presence of fetal DNA in maternal plasma and serum. 1997 *Lancet* **350** 485-87
- Lokeshwar, V.B., Habuchi, T., Grossman, H.B., Murphy, W.M., Hautmann, S.H., Hemstreet III, G.P., Bono, A.V., Getzenberg, R.H., Goebell, P., Schmitz-Dräger, B.J., Schalken, J.A., Fradet, Y., Marberger, M., Messing, E., Droller, M.J. Bladder Tumor Markers Beyond Cytology: International Consensus Panel On Bladder Tumor Markers. *J Urology* 2005 **66** 35-63
- Lopez- Guerrero, J.A., Gilabert, P.B., Gonzalez, E.B., Sanz Alonso, M.A., Perez, J.P., Talens, A.S. Oraval, E.A., de la Rubia, Comos. J., Boix, S.B. Use of reverse-transcriptase polymerase chain reaction (RT-PCR) for carcinoembryonic antigen, cytokeratin 19, and maspin in the detection of tumor cells in leukapheresis products form patients with breast cancer: comparison with immunocytochemistry. *J Hematother* 1999 **8** 53-61
- Ludwig, J.A., Weinstein, J.N. Biomarkers in cancer staging, prognosis and treatment selection. *Nat Rev Cancer.* 2005 **5** 845-56
- Lüftner, D., Lüke, C., Possinger, K. Serum HER-2/*neu* in the management of breast cancer patients *Clinical Biochem* 2003 **36** 233-240
- Manley, G. T., Smitt, P. S., Dalmau, J., Posner, J. B. Hu antigens: reactivity with Hu antibodies, tumor expression, and major immunogenic sites. *Ann Neurol* 1995 **38** 102-110
- Mehta, R.R., McDermott, J.H., Hieken, T.J., Marler, K.C., Patel, M.K., Wild, L.D., Das Gupta, T.K. Plasma c-*erb*B-2 levels in breast cancer patients: prognostic significance in predicting response to chemotherapy. *J Clin Oncol* 1998 **16** 2409-

2416

- Mercatali, L., Valenti, V., Calistri, D., Calpona, S., Rosti, G., Folli, S., Gaudio, M., Frassineti, G.L., Amadori, D., Flamini, E. RT-PCR determination of maspin and mammaglobin B in peripheral blood of healthy donors and breast cancer patients. *Annals of Oncology* 2005 **17** 424-428
- Minata, M., Nishida, N., Komeda, T., Azechi, H., Katsuma, H., Nishimura, T., Kuno, M., Ito, T., Yamaoto, Y., Ikai, I., Yamaoka, Y., Fukuda, Y., Nakao, K. Postoperative detection of alpha-fetoprotein mRNA in blood as a predictor for metastatic recurrence of hepatocellular carcinoma. *J Gastroenterol Hepatol* 2001 **16** 445-451
- Miura, N., Maeda, Y., Kanbe, T., Yazama, H., Takeda, Y., Sato, R., Tsukamoto, T., Sato, E., Marumoto, A., Harada, T., Sano, A., Kishimoto, Y., Hirooka, Y., Murawaki, Y., Hasegawa, J., Shioto, G. Serum Human Telomerase Reverse Transcriptase Messenger RNA as a Novel Tumor Marker for Hepatocellular Carcinoma. *Clin Cancer Res* 2005 **11** 3205-3209
- Morozkin ES, Laktionov PP, Rykova EY and Vlassov VV: Extracellular nucleic acids in cultures of long-term cultivated eukaryotic cells. *Ann N Y Acad Sci* 2004 **1022** 244-249
- Moutzouris, G., Yannapoulos, D., Barbatis, C., Zaharof, A., Theodorou, C. Is β human chorionic gonadotrophin production by transitional cell carcinoma of the bladder a marker of affressive disease and resistance to radiotherapy? *Br J Urol* 1993 **72** 907-909
- Müller, H.M., Widschwendter, A., Fiegl, H., Ivarsson, L., Goebel, G., Perkmann, E., Marth, C., Widschwendter, M. DNA Methylation in Serum of Breast Cancer Patients: An Independent Prognostic Marker. *Cancer Res* 2003 **63** 7641-7645
- Müller, V., Stahmann, N., Riethdorf, S., Rau, T., Zabel, T., Goetz, A., Jänicke, F., Pantel, K. Circulating Tumor Cells in Breast Cancer: Correlation of Bone Marrow Micrometastases, Heterogenous Response to Systemic Therapy and Low Proliferative Activity. *Clin Cancer Res* 2005 **11** 3678-3685
- Nakata, B., Wgawa, Y., Ishikawa, T., Ikeda, K., Kato, Y., Nishino, H., Hirakawa, K. Serum CYFRA 21-1 is one ot the most reliable tumor markers fro breast carcinoma. *Cancer* 2000 **89** 1285-1290

- Nawroz, H., Koch, W., Anker, P., Stroun, M., Sidransky, D. Microsatellite alterations in serum DNA of head and neck cancer patients. *Nat Med* 1996 **2** (9) 1035-7
- Ng EKO, Tsui NBY, Lam NYL, Chiu RWK, Yu SCH, Wong SCC, Lo ESF, Rainer TH, Johnson PJ and Lo YMD: Presence of filterable and nonfilterable mRNA in the plasma of cancer patients and healthy individuals. *Clin Chem* 2002 **48** 1212-1217,.
- Ngan, R.K.C., Yip, T.T.C., Cheng, W-W., Chan, J.K.C., Cho, W.C.S., Ma, V.W.S, Wan, K-K., Au, S-K., Law, C-K., Lau, W-H. Circulating Epstein-Barr Virus DNA in Serum of Patients with Lymphoepithelioma-like Carcinoma of the Lung: A Potential Surrogate Marker for Monitoring Disease. *Clin Cancer Res* 2002 **8** 986-994
- Nogi, H., Takeyama, H., Uchida, K., Agata, T., Horiguchi-Yamada, J., Yamada, H. Detection of MUC1 and keratin 19 mRNAs in the bone marrow by quantitative RT-PCR predicts the risk of distant metastasis in breast cancer patients. *Breast Cancer* 2003 **10** 74-81
- O'Driscoll, L., Kennedy, S., McDermott, E., Kelehan, P., Clynes, M. Multiple drug resistance-related messenger RNA expression in archival formalin-fixed paraffin-embedded human breast tumour tissue. *Eur J Cancer* 1996 **32A** *128-133*
- Pachmann, K., Camara, O., Kavallaris, A., Schneider, U., Schünemann, S., Höffken, K. Quantification of the response of circulating epithelial cells to neodadjuvant treatment for breast cancer: a new tool for therapy monitoring. *Breast Cancer Research* 2005 7 R975-979
- Park, N.J., Li, Y., Yu, T., Brinkman, B.M.N., Wong, D Characterization of RNA in Saliva. *Clin Chem* 2006 **52** *in press*
- Pegram, M.D., Pauletti, G., Slalom, D. J. Her-2/*neu* as a predictive marker of response to breast cancer therapy. *Breast Cancer Res Treat* 1998 **52** 65-77
- Pesole, G., Liuni, S., Grillo, G., Licciulli, F., Larizza, A., Makalowski, W., Saccone, C. UTRdb and UTRsite: specialized databases of sequences and functional elements of 5' and 3' untranslated regions of eukaryotic mRNAs *Nucleic Acids Res* 2000 **28** 193-102
- Pesole, G., Mignone, F., Gissi, C., Grillo, G., Licciulli, F., Liuni, S. Structural and functional features of eukaryotic mRNA untranslated regions *Gene* 2001 **276** 73-81

- Poon, T.C.W., Mok, T.S.K., Chan, A.T.C., Chan, C.M.L., Leong, V., Tsui, S.H.T., Leung, T.W.T., Wong, H.T.M., Ho, S.K.W., Johnson, P. Quantification and Utility of Monosialylated α-Fetoprotein in the Diagnosis of Hepatocellular Carcinoma with Nondiagnostic Serum Total α-Fetoprotein. *Clin Chem* 2002 **48** 1021-1027
- Pornthanakasem, W., Shotelersuk, K., Termrungruanglert, W., Voravud, N., Niruthisard, S., Mutirangura, A. Human papillomavirus DNA in plasma of patients with cervical cancer. *BMC Cancer* 2001 **1** 2
- Qin, L-X., Tang, Z-Y. Recent progress in predictive biomarkers for metastatic recurrence of human hepatocellular carcinoma: a review of the literature *J Cancer Res Clin Oncol* 2004 **130** 497- 513
- Ramirez, J.L., Sarries, C., Lopz do Castro, P., Roig, B., Queralt, C., Escuin, D., De Aguirre, E., Sanchez, J.M., Manzano, J.L., Margeli, M., Sanchez, J.J., Astudillo, J., Taron, M., Rosell, R. Methylation patterns and K-ras mutations in tumor and paired serum of resected non-small-cell lung cancer patients. *Cancer Lett* 2003 **193** 207-216
- Rappl G, Hasselmann DO, Robler M, Ugurel S, Tilgen W and Reinhold U: Detection of tumor-associated circulating mRNA in patients with disseminated malignant melanoma. Ann N Y Acad Sci 2001 **945** 189-191
- Reddi, K.K., Holland, J.F. Elevated serum ribonuclease in patients with pancreatic cancer. *Proc Natl Acad Sci* 1976 **73** 2308-10
- Roncella, S., Ferro, P., Bacigalupo, P., Tognoni, A., Falco, E., Gianquinto, D., Ansaldo, V., Dessanti, P., Fais, F., Rosai, J., Fedeli, F. Human Mammaglobin mRNA is a Reliable Molecular Marker for Detecting Occult Breast Cancer Cells in Peripheral Blood. *J.Exp.Clin. Cancer Res.* 2005 **24** 265-271
- Ryan, B.M., Lefort, F., McManus, R., Daly, J., Keeling, P.W.N., Weir, D.G., Kelleher, D. A prospective study of circulation mutant *KRAS2* in the serum of patients with colorectal neoplasia: strong prognostic indicator in postoperative follow up. *Mol Path* 2003 **56** 172-9
- Sanchez-Cespedes, M., Esteller, M., Wu, Li., Nawroz-Danish, H., Yoo, G.H., Koch, W.M., Jen, J., Herman, J.G., Sidransky, D. Gene Promoter Hypermethylation in Tumors and Serum of Head and Neck Cancer Patients. *Cancer Res* 2000 **60** 892-5
- Schamhart, D., Swinnen, J., Kurth, K-H., Westerhof, A., Kusters, R., Borchers, H.,

Sternberg, C. Numeric Definition of the Clinical Performance of the Nested Reverse Transcription-PCR for Detection of Hematogenous Epithelial Cells and Correction for Specific mRNA of Non-Target Cell Origin as Evaluated for Prostate Cancer Cells. *Clin Chemistry* 2003 **49** 1458-1466

- Schelessinger, J. Cell signalling by receptor tyrosine kinases. *Cell* 2000 **103** 211-25
- Schiavone, N., Rosini, P., Quattrone, A., Donnini, M., Lapucci, A., Citti, L., Bevilacqua, A., Nicolin, A., Capaccioli, S. A conserved AU-rich element in the 3' untranslated region of bcl-2 mRNA is endowed with a destabilizing function that is involved in bcl-2 down-regulation during apoptosis. *FASEB* 2000 **14** 174-184
- Schlinder, H., Wiese, A., Auer, J., Burtscher, H. cRNA target preparation for microarrays: Comparison of gene expression profiles generated with different amplification procedures *Anal Biochem* 2005 **344** 92-101
- Schoor, O., Weinschenk, T., Hennenlotter, J., Corvin, S., Stenzl, A., Rammensee, H.G., Stevanoić Moderate degradation does not preclude microarray analysis of small amounts of RNA. *Biotechniques* 2003 **35** 1192-1201
- Seregni, E., Coli, A., Mazzucca, N. Circulating tumour markers in breast cancer. *Eur J Nucl Med Mol Imaging* 2004 **31** (S1) S15-22
- Serrano-Olivera, A., Dueñas-González, A., Gallardo-Rincón, D., Candelaria, M., De la Garza- Salazar, J. Prognostic, predictive and therapeutic implications of HER2 in invasive epithelial ovarian cancer. *Cancer treatment reviews* 2006 **32** 180-190
- Shaw, G., Kamen, R. A conserved AU sequence form the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell* 1986 **46** 659-667
- Silva JM, Dominguez G, Silva J, Garcia JM, Sanchez A, Rodriguez O, Provencio M, Espana P and Bonilla F: Detection of epithelial messenger RNA in the plasma of breast cancer patients is associated with poor prognosis tumor characteristics. *Clin Cancer Res* 2001 7 2821-2825
- Silva, J.M., Garcia, J.M., Dominguez, G., Silva, J., Miralles, C., Cantos, B., Coca, S., Provencio, M., España, P., Bonilla, F. Persistance of Tumor DNA in Plasma of Breast Cancer Patients After Mastectomy. *Ann Sur Oncol* 2002 **9** 71-6
- Silva, J.M., Silva, J., Sanchez, A., Garcia, J.M., Dominguez, G., Provencio, M., Sanfrutos, L., Jareño, Colas, A., España, P., Bonilla, F. Tumor DNA in Plasma at Diagnosis of Breast Cancer Patients Is a Valuable Predictor of Disease-free Survival

Clin Cancer Res 2002 **8** 3761-3766

- Sisco KL: Is RNA in serum bound to nucleoprotein complexes? *Clin Chem* 2001 **47** 1744-1745
- Slamon, D.J., Clark, G.M., Wong, S.G., Levin, W.J., Ullrich, A., McGuire, W.L. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/ neu oncogene. *Science* 1987 **235** 177-82
- Soeth,E., Grigoleit, U., Moellmann, B., Röder, C., Schniewind, B., Kremer, B., Kalthoff, H. Vogel, I. Detection of tumor cell dissemination in pancrestic ductal carcinoma patients by CK 20 RT-PCR indicates poor survival. *JCancer Res Clin Oncol* 2005 **131** 669-676
- Sorenson, G.D., Pribish, D.M., Valone, F.H., Memoli, V.A., Bzik, D.J., Yao, S.L. Soluble normal and mutated DNA sequences from single copy genes in human blood. *Cancer Epidemiol Biomarkers Prev* 1994 **3** 67-71
- Sozzi, G., Conte, D., Leon, M., Cirincione, R., Roz, L., Ratcliffe, C., Roz, E., Cirenei, N., Bellomi, M., Pelosi, G., Pierotti, M.A., Pastorino, U. Quantification of Free Circulating DNA As a Diagnostic Marker in Lung Cancer. *J Clin Oncol* 2003 **21** 3902-3908
- Sozzi, G., Conte, D., Mariani, L., Lo Vullo, S., Roz, L., Lombardo, C., Pierotti, M.A., Tavecchio, L. Analysis of Circulation Tumor DNA in Plasma at Diagnosis and during Follow-Up of Lung Cancer Patients. *Cancer Res* 2001 **61** 4675-4678
- Sozzi, G., Musso, K., Ratcliffe, C., Goldstraw, P., Pierotti, M.A., Pastorino, U. Detection of Microsatellite Alterations in Plasma DNA of Non-Small Cell Lung Cancer Patients: A Prospect for Early Diagnosis. *Clin Cancer Res* 1999 **5** 2689-2692
- Stathopoulos, E.N., Sanidas, E., Kafousi, M., Mavroudis, D., Askoxylakis, J., Bozionelou, V., Perraki, M., Tsiftsis, D., Georgoulis, V. Detection of CK-19 mRNA-positive cells in the peripheral blood of breast cancer patients with histologically and immunohistochemically negative axillary lymph nodes. *Annals of Oncology* 2005 **16** 240-246
- Stathopoulou, A., Ntoulia, M., Perraki, M., Apostolaki, S., Mavroudis, D., Malamos, N., Georgoulias, V., Lianidou, E.S. A highly specific real-time RT-PCR method for the quantitative determination of CK-19 mRNA positive cells in

peripheral blood of patients with operable breast cancer. *Int J Cancer* 2006 (Online in advance of print)

- Stroun M, Anker P, Beljanski M, Henri J, Lederrey C, Ojha M and Maurice PA: Presence of RNA in the nucleoprotein complex spontaneously released by human lymphocytes and frog auricles in culture. *Cancer Res* 1978 **38** 3546-3554
- Taback, B., Giuliano, A.E., Hansen, N.M., Hoon, D.S.B. Microsatellite Alterations Detected in the Serum of Early Stage Breast Cancer Patients. *Ann NY Acad Sci* 2001 **945** 22-30
- Tangkijvanich, P., Anukulkarnkusol, N., Suwangool, P., Lertmaharit, S., Hanvivatvong, O., Kullavanijaya, P., Poovorawan, Y. Clinical characteristics and prognosis of hepatocellular carcinoma: analysis based on serum alpha-fetoprotein levels. *J Clin Gastroenterol* 2000 **31** 302-308
- Thompson, D.M.P., Krupey, J., Freedman, S.O., Gold, P. The radio-immunoassay of circulating carcinoembryonic antigen of the human digestive system. *Proc Natl Acad Sci USA* 1969 **64** 161-167
- Trivers, G.E., De Benedetti, V.M., Cawley, H.L., Caron, G., Harrington, A.M., Bennet, W.P., Jett, J.R., Colby, T.V., Tazelaar, H., Pairolero, P., Miller, R.D., Harris, C.C. Anti-p53 antibodies in sera from patients with chronic obstructive pulmonary disease can predate a diagnosis of cancer. *Clin Cancer Res* 1996 **2** 1767-1775
- Tsui, N.B.Y., Ng, E.K.O., Lo, Y.M.D. Stability of Endogenous and Added RNA in Blood Specimens, Serum, and Plasma *Clin Chem* 2002 **48** 1647- 1653
- Uen, Y.H., Lin, S,R., Wu, C.H., Hsieh, J.S., Lu, C.Y., Yu, F.J., Huang, T.J., Wang, J.Y. Clinical significance of MUC1 and c-Met RT-PCR detection of circulating tumor cells in patients with gastric carcinoma. *Clinica Chimica Acta* 2006 **367** 55-61
- Uenishi, T., Kubo, S., Hirohashi, K., Tanaka, H., Shuto, T., Yamamoto, T., Nishiguchi, S. Cytokeratin-19 fragments in serum (CYFRA 21-1) as a marker in primary liver cancer. *Br J Cancer* 2003(a) **88** 1894- 1899
- Uenishi, T., Kubo, S., Yamamoto, T., Shuto, T., Ogawa, M., Tanaka, H., Tanaka, S., Kaneda, K., Hirohashi, K. Cytokeratin 19 expression in hepatocellular carcinoma predicts early postoperative recurrence. *Cancer Sci* 2003 (b) **94** 851-857

- Ulmer, A., Schmidt-Kittler, O., Fischer, J., Ellwanger, U., Rassner, G., Riethmuller, G., Fierlbeck, G., Klein, C.A. Immunomagnetic enrichment, genomic characterization, and prognostic impact of circulating melanoma cells. *Clin Cancer Res* 2004 **10** 531-7
- Usadel, H., Brabender, J., Danenberg, K.D., Jerónimo, C., Harden, S., Engles, J., Danenberg, P.V., Yang, S., Sidransky, D. Quantitative Adenomatous Polyposis Coli Promoter Methylation Analysis in Tumor Tissue, Serum, and Plasma DNA of Patients with Lung Cancer. *Cancer Res* 2002 **62** 371-375
- Vasioukhin, V., Anker, P., Maurice, P., Lyautey, J., Lederrey, C., Stroun, M. Point mutations of the N-*ras* gene in the blood plasma DNA of patients with myelodysplastic syndrome or acute myelogenous leukaemia. *Br. J. Haematol.* 1994 **86** 774-779
- Vogel, I., Kalthoff, H. Disseminated tumour cells. *Virchows Arch* 2001 **439** 109-117
- Walknowska, J., Conte, F.A., Grumbach, M.M. Practical and Theoretical Implications of Fetal/Maternal Lymphocyte Transfer. 1969 *Lancet* **293** 1119-1122
- Wang W, Caldwell MC, Lin S, Furneaux H, Gorospe M. HuR regulates cyclin A and cyclin B1 mRNA stability during cell proliferation. *EMBO J.* 2000 **19** 2340-50
- Wang, J.Y., Wu, C.H., Lu, C.Y., Hsieh, J.S., Wu, D.C., Huang, S.Y., Lin, S.R.Molecular Detection of Circulating Tumor Cells in the Peripheral Blood of Patients with Colorectal Cancer Using RT-PCR: Significance of the Prediction of Postoperative Metastasis. *World J Surg* 2006 **30** 1-7
- Weitz, J. Kienle, P., Magener, A., Koch, M., Schrodel, A., Willeke, F. Autschbach F, Lacroix J, Lehnert T, Herfarth C, von Knebel Doeberitz M. Detection of disseminated colorectal cancer cells in lymph nodes, blood and bone marrow. *Clin Cancer Res* 1999 **5** 1830-6.
- Widschwendter, A., Müller, H., Fiegl, H., Ivarsson, L., Wiedemair, A., Müller-Holzner, E., Goebel, G., Marth, C., Widschwendter, M. DNA Methylation in Serum and Tumors of Cervical Cancer Patients. *Clin Cancer Res* 2004 **10** 565-571
- Wieczorek, A.J., Rhyner, C., Blck, L.H. Isolation and characterisation of an RNA-proteolipid complex associated with the malignant state in humans *Proc Natl Acad Sci* **82** 3455-3459

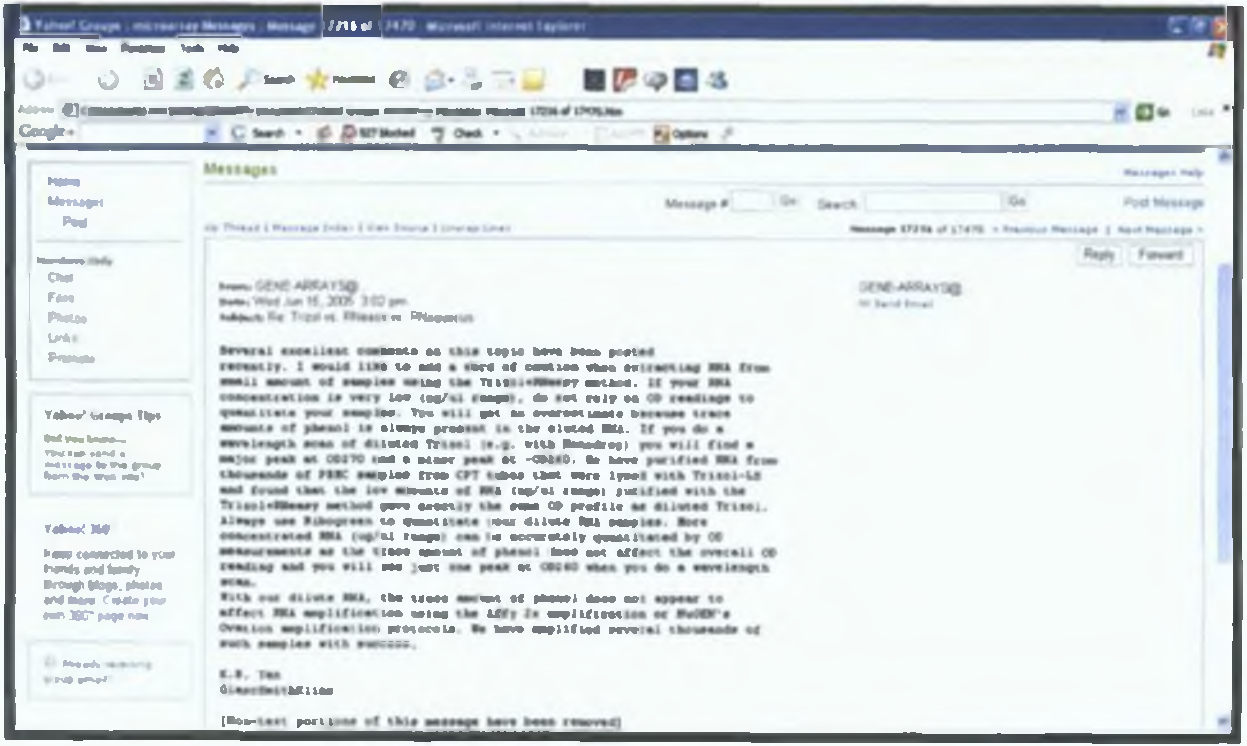
- Wiedswang, G., Borgen, E., Schirmer, C., Käresen, R., Kvalheim, G., Nesland, J.M., Naume, B. Comparison of the clinical significance of occult tumor cells in blood and bone marrow in breast cancer. *Int J Cancer* 2006 **118** 2013-2019
- Wilson, A.P.M., van Dalen, A., Sibley, P.E.C., Dasper, L.A., Durham, A.P., El Shami, A.S. Multicenter tumour marker reference range study. *Anticancer Res* 1999 **19** 2749-52
- Wilson, C., Pepper, S., Hey, Y., Miller, C. Amplification protocols introduce systematic but reproducible errors into gene expression studies. *Biotechniques* 2004 **36** 498- 506
- Wilson, T., Treisman, R. Removal of poly (A) and consequent degradation of *c-fos* mRNA facilitated by 3' AU-rich sequences. Nature 1988 **336** 396-399
- Wong, I.H.N., Lo, Y.M.D., Lai, P.B.S., Johnson, P.J. Relationship of *p16* Methylation Status and Serum α-Fetoprotein Concentration in Hepatocellular Carcinoma Patients. *Clin Chem* 2000 **46** 1420-22,
- Wong, I.H.N., Zhang, Z., Lai, P.B.S., Lau, W.Y., Lo, Y.M.D. Quantitative Analysis of Tumor-derived Methylated *p16INK4a* Sequences in Plasma, Serum, and Blood Cells of Hepatocellular Carcinoma Patients *Clin Cancer Res* 2003a **9** 1047-1052
- Wong, S.C.C., Lo, S.F.E., Cheung, M.T., Ng, K.O.E., Tse, C.W., Lai, B.S.P., Lee K.C., Lo, Y.M.D. Quantification of Plasma β-catenin mRNA in colorectal cancer and adenoma patients. *Clin Cancer Res* 2004 **10** 1613-1617
- Wong, T-S., Kwong, D.L-W., Sham, J.S-T., Wei, W.I., Kwong, Y-L., Yuen, A.P-W. Quantitative Plasma Hypermethylated DNA Markers of undifferentiated Nasopharyngeal Carcinoma. *Clin Cancer Res* 2004 **10** 2401-6
- Wong, T-S., Man, W-L., Lam, A.K-Y., Wei, W.I., Kwong, Y-L., Yuen, A.P-W. The study of *p16* and *p15* gene methylation in head and neck squamous cell carcinoma and their quantitative evaluation in plasma by real-time PCR. *Eur J Cancer* 2003b **39** 1881-1887
- Wu, C.C., Chien, K.Y., Tsang, N.M., Chang, K.P., Hao, S.P., Tsao, C.H., Chang, Y.S., Yu, J.S. Cancer cell-secreted proteomes as a basis for searching potential tumor markers: Nasopharyngeal carcinoma as a model. *Proteomics* 2005 **5** 3173-3182
- Xu, Y., Yao, L., Li, H., Ouyang, T., Li, J., Wang, T., Fan, Z., Lin, B., Lu, Y.,

Larsson, O., Zie, Y. Presence of *erbB2* mRNA in the plasma of breast cancer patients is associated with circulating tumor cells and negative estrogen and progesterone receptor status. *Breast Cancer Research and Treatment* 2006 **97** 49-55

- Yamaguchi, S., Asao, T., Nakamura, J-E., Ide, M., Kuwano, H. High frequency of *DAP-kinase* gene promoter methylation in colorectal cancer specimens and its identification in serum. *Cancer Lett* 2003 **194** 99-105
- Zieglschmid, V., Hollmann, C., Böcher, O. Detection of Disseminated Tumor Cells in Peripheral Blood. *Crit Rev Clin Lab Sci* 2005 **42** 155-196
- Zieglschmid, V., Hollmann, C., Gutierrez, B., Albert, W., Strothoff, D., Gross, E., Böcher Combination of Immunomagnetic Enrichment with Multiplex RT-PCR Analysis for the Detection of Disseminated Tumor Cells. *Anticancer Res* 2005 **25** 1803-1810
- Zigeuner, R.E., Riesenberg, R., Pohla, H., Hofstetter, A., Oberneder, R. Isolation of circulating cancer cells from whole blood by immunomagnetic cell enrichment and unenriched immunocytochemistry in vitro. *J Urol* 2003 **169** 701-5

7 Appendix

7.1 Microarray user group posting



A

7.2 UTRScan sequence results

7.2.1 Gene transcripts called P in all CM samples

Gene	Accession	15-LOX-DICE	[Position] and Sequence	Other	[Position] and Sequence
THRA	NM_199334	5	[95,112] :CGCCGCCTCT GCCGG AGG	---	
			[1730,1744] :CGCTTCCTCC AC ATG		
			[1772,1790] :CCCCCACTCT TCCTCG AGG		
			[2074,2090] :CCCCATCCTCT CAG AAG		
			[2301,2315] :CTCCTCCTCT GG AGG		
RASEF	AK058178	4 +1 matching repeat	[240,254) :CCCCGTCCTTC C AGG	---	
			¶1383,1397] :GCCTGCCTCT GT AGG		
			¶1398,1416] :CTCCACCTCT GGGGGC AGG		
			[1559,1572] :CCCTGACCCC C AAG		
ZNF548	NM_152909	3	[519,533] :GCCCACCCTT GA AAG	---	
			[1975,1993] :CCCTACTTCC CCAGAA ATG		
			[2478,2491] :CTCCGCCTCC C AGG		
SYMPK	NM_004819	5	[665,681] :CCCTGTCACC CCGC ATG	---	
			[748,766] :CCCTACATCC AGTACA ACG		
			[938,951] :ACCTGCCCCC G ACG		
			[1610,1627] :CCATGTCCCC CCTGG AGG		
			[1722,1735] :CCCTACCCTG A AGG		
C21orf114	NM_001012707	---		---	
---	AK091784	1	[256,270] :CTCCACCTCC CC AGG	---	
ZMYND11	NM_006624	---		1 K-Box	[1247,1254] :CTGTGATG
RAB13	NM_002870	1	[723,736] :CCCCACCCCG G AAG	---	
LOC388388	AK027091	2	[337,351] :CCCTGTATCT AT ACG	1 GY-Box	[34,40] :GTCTTCC
			[449,468] :CCCTGCTTCT TTCTCTA AAG		
AUTS2	NM_015570	7	[1971,1989] :CCCACCCACC GCCATC ATG	1 Brd-Box	[4551,4557] :AGCTTTA
			[2735,2754] :CCCCTCCGCC CTGGCTG AAG		
			[2895,2914] :ACCTGTCCTCC CGGTGA ATG		
			[3293,3307] :ACCTGCCTCC AG AGG		
			[3374,3387] :CCCTGGCCTC G ATG		
			[3834,3850] :CCCCGCCTCC CTCG ACG		
			[4011,4028] :CCCGGTCTCT CCCAG AAG		
CYP3A5	AK055879	1	[2279,2295] :CCTCTCCTCC TCCA AAG	---	
SCD5	NM_024906	---		---	
KIAA1641	NM_020970	1	[1396,1409] :GCCCATCTCT C ATG	---	
THSD4	AK130734	2	[40,55] :CCTCCCCTCC CCC ATG	1 K-Box	[416,423] :CTGTGATG
			[948,961] :CTCTGTCCTT A AAG		
FAM72A	NM_207418	1	[828,846] :CCTCATCCTT CCTGCG ACG	1 ADH_DRE	[476,483] :AAGGCTGA
				1 Brd-Box	[453,459] :AGCTTTA
LOC388397	AK057167	2	[1229,1244] :CCCCACCTCC CGA AAG	1 K-Box	[677,884] :CTGTGATT
			[2254,2268] :TCCTTCCTCT CC AGG		
RP11-343N15.3	BC017972	1	[229,246] :CCCTCCCCTC AGTGC AAG	1 SECIS	[914,973] :GTGG AAAACTACA A TGAT AAAAATTT AAA AATCACCAGAG GGATTTA AGAG TATAT TTGC
TRIM26	AK023334	---		---	
LOC389953	AY010114	3	[63,82] :CTCCGCCTCC CGGGTTC ACG	---	
			[1382,1401] :CCCTGCCTCA GTCTCCC AAG		
			[2853,2871] :CCTACCCTAT CCCTGT ATG		
---	AF086452	1	[387,401] :TCCACCCTCC AA ATG	---	

Table 7.2.1 Sequence of regulatory elements found in gene transcripts called P in all CM samples

7.2.2 Gene transcripts differentially regulated between parental CM samples and their drug selected variants

Gene	Accession	15-LOX-DICE	[Position] and Sequence	Other	[Position] and Sequence
SMAF1	NM_001018082	---		1 GY_Box	[738,744] :GTCTTCC
---	AL832767	4	[51,67] :CCTCGCCCTCT GGC AGG	2 K-Box	[1319,1326] :CTGTGATT
			[165,184] :CCCTGTCCCC AACCTGA AAG		[2929,2936] :CTGTGATT
			[4451,4466] :CTCCTCCTCC TGC AAG		
			4524,4540] :CCTATCCTGT GCTC AGG		
---	BC040325	1	[1270,1285] :CCCTGTCTTC AGG AAG	---	
---	BC039476	---		1 K-Box	[812,819] :CTGTGATT
---	AL109711	1	[102,118] :CCACGTCCCC AGCT AGG	---	
GRIK2	AJ252246	1	[2184,2200] :CCTCACCTCT GATT ATG	1 K-Box	[285,292] :CTGTGATC
ABCB4	BC020818	---		---	
NINJ1	NM_004148	3	[943,959] :CTCTCCCTCT CCCA AAG	---	
			[1192,1207] :CCCAACCCCC TCC ACG		
			[1252,1269] :CCCACGCTCT TTGTC AGG		
PHF2	AF043725	6	[729,747] :GCCCATCCTT GTCCCC AAG		
			[771,790] :CCCTGCCCCA ACATTCT ACG		
			[1529,1546] :CCCCACATCT AGTCC AAG		
			[1769,1786] :CCCCGTCTCC CATTG AGG		
			[3255,3273] :CTCTGCCTCC GCCTCC ACG		
			[3480,3498] :CCGTGCCTCC CAGCCC ATG		
NBR2	NM_005821	1	[919,933] :CCCCAACTCC AC AAG	---	
COMT	NM_007754	---		1 GY_Box	[764,770] :GTCTTCC
ADAM7	NM_003817	1	[1134,1147] :ACCTGTCCTT C AGG	1 K-Box	[1367,1374] :CTGTGATG
HLXB9	NM_005515	1	[1872,1887] :CCCCGCGTCC GAG AAG	---	
MMRP19	NM_015957	---		1 -CPE	1080,1226] :TTTTTTAT ATATACTCTCGAGAGAATTCCTCTTAATTTTTAAAGATGCTGGTGAT AATAAA ATTCATT
					AGAAAATTTCCTCATTGTGGAATGAGCATTCTCTTGTTTTAATGTTGGTGTCAGAAAATAAATATGAAACATTAAGTCC
				1IRES	[1134,1226] :TAATA AAATT CATTAG AA AATTTC CTCATTGT GGAATG AGCAT TCTC TTGTT
					TTAATG TTGG TGTCA GAAAA TAAAT ATGAAAC ATTAA GTCC
				3 K-Box	[86,93] :CTGTGATG
					[302,309] :TTGTGATA
					[439,446] :CTGTGATG
THEG	NM_016585	2	[871,886] :GCCACCCTCT TGG AAG	1IRES	[1232,1337] :TATTCT GAG CTTTTT G GCCTG GAGAAGGG AGGGC GGGCT AGAA GGCCA AAAGA
			[1211,1227] :CCCTAACCCT GTGT ATG		AGGATG AGGAGG ACAATA ATATT TATTTT TGCATCCA AAAAAA AAAAAAAAAA
CCDC45	BX641136	3	[918,937] :CCGTGTCCCT CATCTTA ATG	1 ADH_DRE	[422,429] :AAGGCTGA
			[1404,1422] :CTCCCCCTCC TCAGTG ATG	2 K-Box	[1032,1039] :GTGTGATA
			[3460,3473] :CCCTCCCTAT G AGG		[1517,1524] :CTGTGATA
DDI2	BC006011	1	[71,90] :CCCTGCGCCC CGCGCCC AGG	1 Brd-Box	[1030,1036] :AGCTTTA
FLJ10211	AK091238	4	[771,784] :CTCCACCTCC C AGG	1 IRES	[2608,2690] :GTTTC AATGT GATTC GTTTC CCACATTT GAAAA TAGGG GGGGA
			[1265,1284] :CTCTGCCTCC CGGGCTC AAG		TTTTA CTTTA AAATT CAGAT AT CTGGC TCCTC TTTAG GGG
			[1511,1525] :TCCCACCTCT GG AGG		
			[2334,2350] :CCCTACCTTC AGGG ATG		
C13orf24	AK022413	4	[652,669] :CCTTATCTCT ACTAA AAG	---	
			[1374,1392] :CTCCACCTCT GGCGGC AGG		
			[1527,1540] :CCCTGACCCC C AAG		
			[1909,1922] :CTCCCCCTCC A AAG		
---	BX537871	---		1 Brd-Box	[238,244] :AGCTTTA

Table 7.2.2 Sequence of regulatory elements found in gene transcripts differentially regulated in parental CM samples against their drug selected variants