ENDOTHELIAL CELL GENE EXPRESSION

John Matthew Jeff Herbert
Supervisors: Prof. Roy Bicknell and Dr. Victoria Heath

PhD thesis

University of Birmingham

August 2012

ABSTRACT

Tumour angiogenesis is a vital process in the pathology of tumour development and metastasis. Targeting markers of tumour endothelium provide a means of targeted destruction of a tumours oxygen and nutrient supply via destruction of tumour vasculature, which in turn ultimately leads to beneficial consequences to patients. Although current anti -angiogenic and vascular targeting strategies help patients, more potently in combination with chemo therapy, there is still a need for more tumour endothelial marker discoveries as current treatments have cardiovascular and other side effects. For the first time, the analyses of *in-vivo* biotinylation of an embryonic system is performed to obtain putative vascular targets. Also for the first time, deep sequencing is applied to freshly isolated tumour and normal endothelial cells from lung, colon and bladder tissues for the identification of pan-vascular-targets. Integration of the proteomic, deep sequencing, public cDNA libraries and microarrays, delivers 5,892 putative vascular targets to the science community. These analyses identify Endothelial Specific Molecule 1 as a pan vascular target and lysyl oxidase-like 2 as putative novel vascular target. It is envisioned vascular targets and angiogenesis genes in this data will destroy or inhibit tumour vessel growth without the side effects manifest with current clinical regimens.

Dedicated to Robin and Wendy

ACKNOWLEDGEMENTS

A big thank you to everyone who has helped me along the way. Obviously thanks for the support and advice given from my supervisor Roy Bicknell and Vicky Heath (co-supervisor). Thank you to several people and groups who have given me the data to analyse through their hard work and great experimentation. Particularly Andreas Bikfalvi and his group for providing embryonic proteomic data, Patricia Lalor for liver endothelial cells (and Ben James), Xiaodong Zhuang and Forhad Ahmed for providing endothelial cell isolated RNA for deep sequencing, qPCRs and to Peter Noy and Vicky Heath for other experimental assistance. Thank you to the Plate-forme Protéomique de la Génopole Toulouse Midi-Pyrénées group for proteomics data, Genepool Edinburgh for 454 and Illumina sequencing data and the University of Birmingham technology hub for SOLiD4 sequencing data. Thank you to Miss Teresa Smithers, B.Sc. PGCE MIfL, for professionally proof reading, checking grammar and spelling (www.c2btuition.co.uk/). Thank you to Dov Stekel and Francesco Falciani for early inspiration and opportunity. Thank you for kind support from family, Olga Kolomeitz and Molly Herbert.

Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1; INTRODUCTION

1.1 Introduction

The overall theme of this thesis is to provide putative vascular target candidates from new data sources combined with public data, so that others can validate interesting candidate genes in the laboratory. For the first time, deep sequencing is used to measure mRNA contents of freshly isolated tumour endothelial cells and normal endothelial cells from multiple organs. Also, embryonic *in-vivo* biotinylation proteomic data is used to map the vascular and extracellular proteome of the chicken to find vascular targets. The introduction below gives background to relevant topics related to angiogenesis, vascular targeting and RNA-seq gene expression.

1.2 The endothelial cell and angiogenesis

Vertebrates have a network of blood and lymphatic vessels to transport gases and molecules around the body. Endothelial Cells (ECs) line the inside of the heart and both types of vessel throughout the body, forming the interface between blood and vessels. Although ECs have traditionally been classed as epithelial cells, they differ as the EC cytoskeleton is formed with vimentin filaments whereas epithelial cells have keratin. Also, ECs originate from the mesoderm and not the endoderm or ectoderm as epithelial cells do. ECs are surrounded by an Extra Cellular Matrix (ECM) and smooth muscle cells. Endothelium is not inert, it is a dynamic and specialised tissue involved in many important functions including a selective barrier function, inflammation, vascular tone (blood pressure regulation) and angiogenesis.

1.3 Selective barrier and inflammation

One of the main functions of endothelium is to act as a selectively permeable barrier to aid the movement of different molecules between blood and the surrounding tissue. This can be done in two ways: 1) paracellular transport by using special junction molecules at their cell to cell junctions or 2) transcellular transport by endocytosis through the cell. ECs can be specialised depending on the type of tissue or organ they are situated in. For

instance, at the blood brain barrier, ECs form tight junctions to only let certain molecules and a limited number of amino acids [1] enter the brain. In kidneys, ECs have 300 to 400 perforations that enable the export of water, urea and other substances out of the body [2].

During inflammation, another important EC barrier process is the leucocyte adhesion cascade. In response to cytokines, i.e. during inflammation, ECs contact the leucocytes via selectins (capture) and enable rolling of leucocytes along the EC surface (rolling and slow rolling). Then, cell surface leucocyte genes including integrins, LFA-1 and VLA-4 strengthen the adhesion bond (firm adhesion) and cause the leucocytes to flatten and migrate along the ECs [3]. Finally transmigration takes place as they migrate through the endothelium at EC junctions and attach to the extra cellular matrix components at the site of inflammation.

1.4 Blood vessel formation

Angiogenesis and vasculogenesis are both processes that lead to the generation of new blood vessels. Vasculogenesis takes place in the developing embryo and forms the initial cardiovascular system whereas angiogenesis is the growth of blood vessels from existing blood vessels. Angiogenesis can be a normal physiological process such as tissue growth, in exercise, corpus luteum formation, endometrial development in menstrual cycle and wound healing but can also arise in pathologies such tumour angiogenesis or age-dependent macular degeneration [4].

Angiogenesis is firstly initiated by stimulation and activation of endothelial cells that line the inside of established blood vessels. This stimulation compels endothelial cells to secrete and release proteases (for example matrix metallopeptidases) for the purpose of degrading the extracellular matrix to enable endothelial migration escape towards the stimulatory signal, most likely VEGF. A second step in angiogenesis is the emergence of sprouting and migrating endothelial cells out of established vessels towards the stimulus provided by, for example, cancer cell growth factor release. It is the lead tip endothelial cells that first enter the basement membrane and migrate towards the VEGF/growth

factor stimulus [5], which are closely followed by stalk endothelial cells that subsequently orchestrate lumen and extracellular matrix formation [6] and build tight cell-to-cell junctions to strengthen the sprout [7]. An overview diagram of the angiogenesis process is given in figure 1.4.1 and a representation of the different forms of sprouting and intussusceptive vessel formation in figure 1.4.2.

Figure 1.4.1 Overview of tumour angiogenesis process

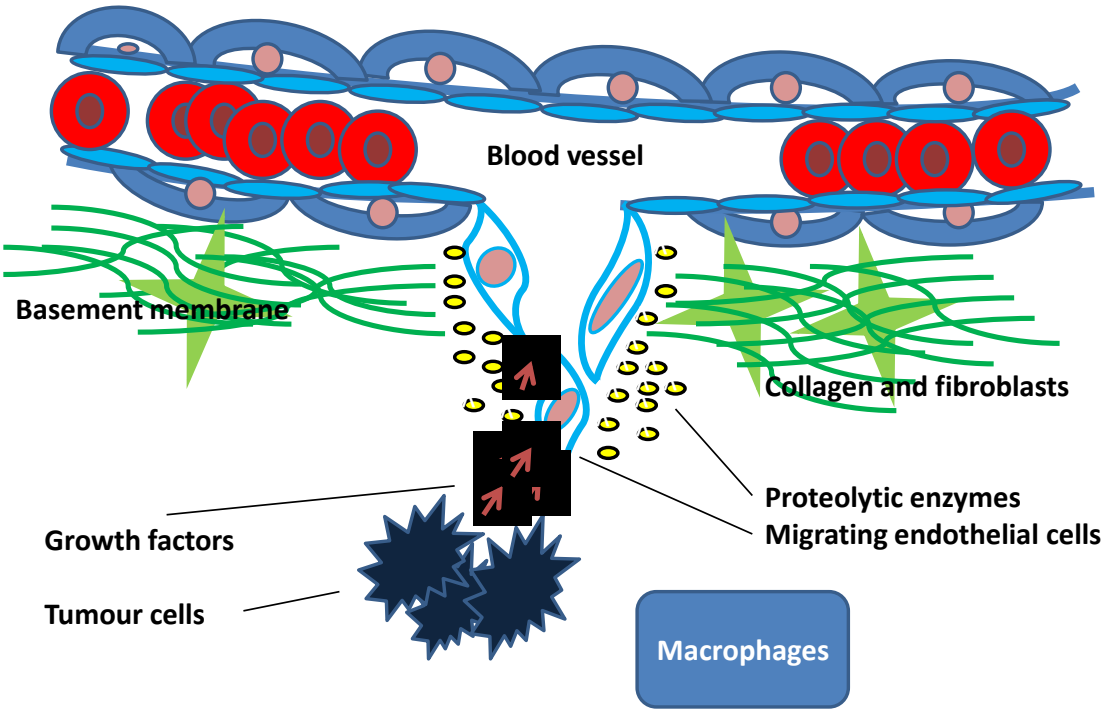


Figure 1.4.1. Angiogenesis is the growth of new blood vessels from existing blood vessels and it is portrayed in this figure. Endothelial cells are shown migrating towards a gradient of pro-angiogenic growth factors, released by nutrient starved tumour cells, whilst they themselves secrete proteolytic enzymes to break down the basement membrane. This leads to the formation of tumour blood vessels and leads to tumour growth and metastasis. This figure was adapted from [8].

Figure 1.4.2 Sprouting and intussusceptive forms of angiogenesis

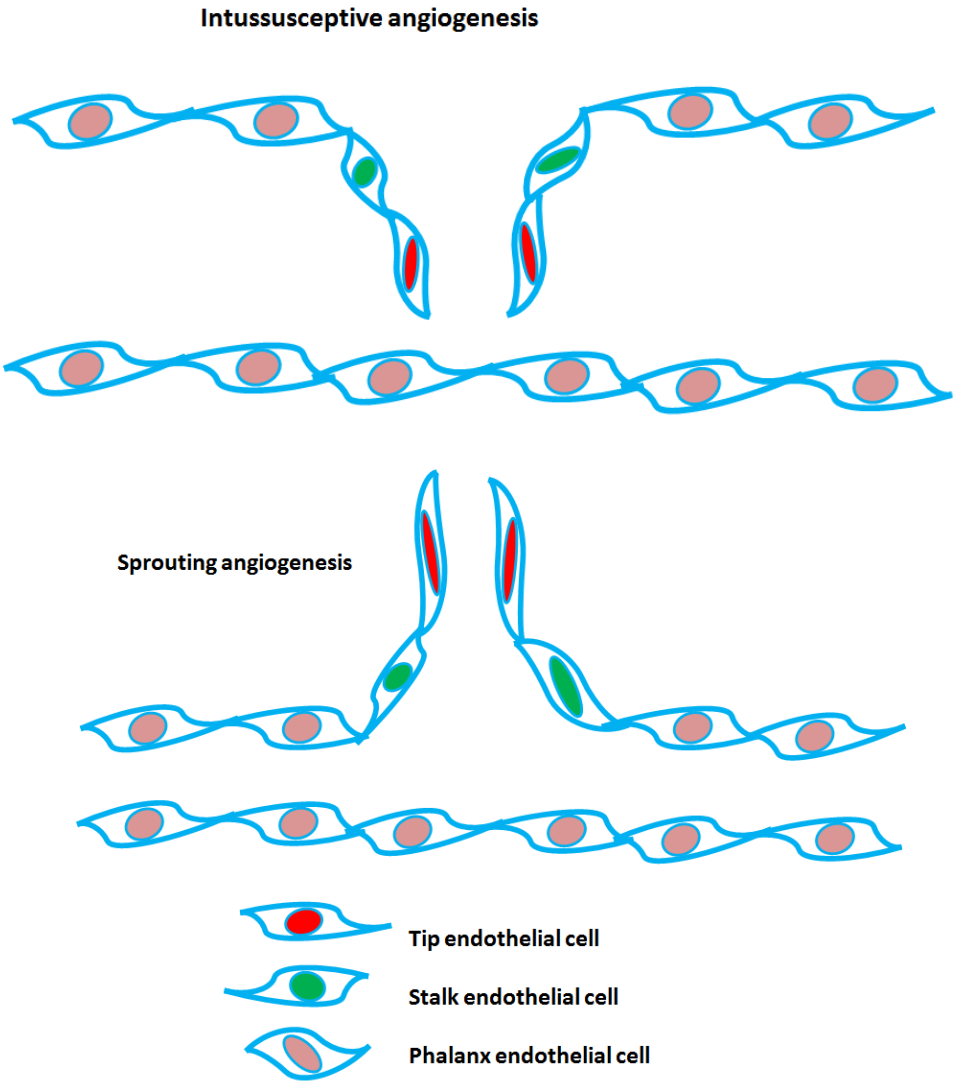


Figure 1.4.2. Intussusceptive is a form of angiogenesis where a single blood vessel divides into two new blood vessels, which is shown on the top half of the illustration. On the bottom panel, the process of sprouting angiogenesis is portrayed where activated endothelial cells release proteolytic enzymes to degrade the basement membrane and then migrate to form tubes. The leading endothelial cells that lead migration are called Tip

cells. Cells directly behind Tip cells are the Stalk cells and the majority of EC of the vessels are termed Phalanx cells [9].

1.5 Tumour angiogenesis

Solid tumours, in their initial growth stage, rely on oxygen and nutrients via diffusion of close by blood vessels and tissues. However, for a tumour to grow beyond a certain size, estimated to be $2mm^3$, and spread, a tumour must acquire its own blood supply. This was proved by an early experiment where they injected the same number of cancer cells onto two different areas of an eye: one injection fixed at the anterior chamber (no blood supply but nutrients) and the other on the iris where blood vessels were able to grow. They showed that tumour cells without a blood supply only grow to $2mm^3$ and tumour growth is unlimited with angiogenesis [10]. The process in which a tumour gets a blood supply is called tumour angiogenesis. As a result of the environment of tumours, being in 1% oxygen (hypoxia) and under oxidative stress, tumour cells release growth factors into the surrounding tissue, which can initiate a response from close by vascular endothelial cells. The most prolific growth factor responsible is VEGF but it can also include bFGF and IL8 growth factors.

1.6 bFGF and VEGF signalling, the most potent angiogenesis initiators and important growth factors for tumour growth

Although there exist a number of molecules that can stimulate angiogenesis (FGF, PGF, TGF, HGF, TNF, angiogenin, IL8 and ANG2) [11], during angiogenesis and vasculogenesis the signalling system that is pivotal to EC migration and proliferation during vessel formation is Vascular Endothelial Growth Factor (VEGF) and its three receptors. VEGFR2 is expressed on the surface of blood vessels and is the main perpetrator of the angiogenic effect of VEGFA. Both *in-vivo* and *in-vitro* studies using VEGFA, which exclusively binds to VEGFR2, showed proliferation, tube formation and chemotaxis of ECs [12]. VEGFR3 is expressed on lymphatic endothelial cells and is the receptor activated by VEGFC or VEGFD, which stimulates lymphoangiogenesis. VEGFR1 is thought to act as a negative regulator of VEGFA, since VEGFA ligand binding only causes weak phosphorylation of the receptors intracellular tyrosine residues

and does not cause proliferation or migration [13]. Additionally, mice devoid of intracellular tyrosine kinase domain of VEGFR1 had no phenotype or any noticeable vascular network defects [14]. A soluble, alternative isoform lacking the tyrosine kinase domains binds VEGFA and is thought to prevent VEGFR2 activation by VEGFA. Once binding of ligands to the VEGF receptors, dimerization of the receptors occurs, which in turn leads to auto-phosphorylation of the intracellular tyrosine kinase domains to activate intracellular molecules that elicit angiogenesis [15].

Basic Fibroblast Growth Factor (bFGF or FGF2) was one of the first angiogenesis growth factors to be discovered and is one of a >20 strong family of growth factors. It is unusual in that bFGF does not contain the typical signal sequence that enables proteins to be exported from cells [16]. On the surface of cells are heparan sulfate proteoglycans, which are proteins that are highly glycosylated, and these bind bFGF and other fibroblast growth factor family members, to give reservoirs of growth factors that can be released by enzymes (heparanases) or by FGF carrier proteins. It has been shown that Fibroblast Growth Factor Binding Protein 1 (FGFBP1) plays a pivotal role in the angiogenic switch in colorectal and pancreatic cancers, acting as an extracellular chaperone for FGF [17]. There are 4 tyrosine kinases FGF receptors before alternative splicing and this can lead to up to 48 different isoforms or paralogs of FGF receptors [18]. It is also known that FGF signalling of cell proliferation and/or angiogenesis will not occur without a heparin sulphate co-receptor [18]. In fact, Heparan sulfate proteoglycans have a massive impact on tumour angiogenesis and this is exemplified if you note that growth, migration, and capillary morphogenesis are stimulated in cell cultures of endothelial cells by the addition of FGF and heparin [19].

1.7 Treatment of cancer and anti-angiogenics

There are several methods of treating solid tumours aside from surgical removal. Two of which are chemotherapy and radio therapy that aim to shut down the tumour cells themselves by taking advantage of the genetic instability characteristic of tumour cells and their hyper-proliferative states. Chemotherapy drugs preferentially target cells that are actively dividing, such as tumour cells and attack those in a certain phase of the cell

cycle, which leads to apoptosis. However, the drawback is that chemotherapy can also damage normal tissues, which is exemplified by the side effects of hair loss and anaemia. In a similar manner, radiotherapy treatment involves bombarding tumours with a stream of high energy waves that damage fragile DNA and cellular components of cancer cells. Whilst it is true that cancer cells are genetically unstable and more susceptible to radio therapy, as with chemotherapy, there can be off-target effects on normal tissues.

Anti-angiogenics offer an alternative or complementary treatment as tumours need a dynamic and expanding vasculature to assist tumour growth and metastasis [20, 21]. Targeting angiogenic pathways is a direct targeted route to treatment with less off-target side effects, though there are still some side effects. Since cancer cells have unstable genomes they are prone to drug resistance due to high rates of mutations. In contrast, the growing endothelial cells of a growing tumour vasculature are essentially normal or un-mutated, so are less likely to become resistant. This method does not kill tumours outright but is rather an approach to containing tumours as even without tumour angiogenesis occurring, established close by vessels can supply a tumour with limited oxygen and nutrients by diffusion. In homeostasis, there is a fine balance between pro and negative angiogenic factors that tightly controls the growth of new blood vessels [22] (corpus luteum formation, wound healing and endometrial development in the menstrual cycle undergo physiological angiogenesis). In contrast, tumour angiogenesis is an uncontrolled excess of pro-angiogenic factors (e.g. hypoxia and chemo-gradient of growth factors) that stimulate angiogenesis. A list of the key genes involved in angiogenesis that could be or are being targeted with anti-angiogenic therapy are listed in table 1.7.1.

One major method of anti-angiogenic therapy involves blocking the activation of endothelial cells by the stimulatory angiogenic factors, released by surrounding stroma and tumour cells. The most prominent factor is VEGF that is highly expressed and released from cancer cells and from stroma cells in close proximity. Elevated VEGF is a critical factor in cancer that has been shown to be a marker of poor outcome in many cancers [22]. Endothelial tip cells express the VEGF receptors [5], which attracts the endothelial cells towards the gradient of VEGF and other stimulatory factors released by

tumours. With VEGF and its receptors being so important in tumour angiogenesis, many anti-angiogenic therapeutics have been developed that target this pathway.

An antibody to VEGF that is able to neutralize VEGF *in-vivo* was approved by the FDA in 2004. It is called Bevacizumab (brand name Avastin) and it had been shown, pre-clinically, to extend the lives of people living with colon cancer, non-small cell lung carcinoma, cancer of the breast or brain tumours [23]. However, following FDA approval and with the exception of renal cell carcinoma, Avastin has only been effective when combined with conventional radiotherapy or chemotherapy. The timing of Avastin administration is critical and it has been determined that the Avastin actually normalizes blood vessels, possibly enabling better conventional therapy delivery to the tumour [24, 25].

There are several hundred anti-angiogenic molecules currently under research, with apparently only 10 of them having made it into clinical use [23]. Besides anti-VEGF pathway inhibitors there are other pathways that can modulate angiogenesis. For example, there are immuno-modulating drugs that target tumour necrosis factor, analogs of thalidomide, that have been FDA approved for treatment of myeloma [26]. Yet another method is to prevent the angiopoietins from binding to the TEK receptor, which has been achieved using a Peptidebody that was shown to inhibit angiogenesis. This peptidebody (AMG_368) has now gone to clinical trials [27].

One interesting anti-tumour and anti-angiogenic therapy approach in recent research is the co-suppression of both Interleukin 6 and VEGF [28]. This abolished glioma U87 tumour cell growth and invasion in the chicken chorioallantoic membrane (CAM) assay and in the brains of xenografted mice [29]. Monoclonal antibodies like Avastin bind to VEGFR ligands but Astra Zeneca are investigating an alternative anti-angiogenesis molecule that binds to all 3 VEGFRs intracellular tyrosine kinase domains, independent of whichever ligands are binding to the receptors. This results in no signal transduction and no angiogenesis. This orally administerd therapy is currently in clinical trials for non-small cell lung carcinoma, colon cancer and recurrent glioblastoma [15]. There are

many other targets and therapeutics under trial, far too many for this introduction, which simply provides a glimpse of anti-angiogenic therapies.

Table 1.7.1 A list of potential anti-angiogenic therapy targets

Extracellular matrix genes	Gene symbols
Fibronectin 1	FN1
Collagens	COL4A2, COL4A1, COL1A1, COL3A1, COL5A1, COL18A1, COL6A2, COL1A2 etc
Thrombospondin	THBS1
Fibrin/Fibrinogen	FGB, FGA and FGG
Endostatin	COL18A1
Tumstatin	COL4A3
cysteine-rich, angiogenic inducer, 61	CYR61
Growth vactors	**Growth factor gene symbols**
vascular endothelial growth factors	VEGFA
fibroblast growth factor 2	FGF2(Bfgf)
Wnt growth factors	WNT2 or WNT9B etc
placental growth factor	PGF (PLGF)
platelet-derived growth factor beta polypeptide	PDGFB (PDGF)
chemokine (C-X-C motif) ligand 12	CXCL12 (SDF1)
Growth factor receptors	**Growth factor receptor gene symbols**
vascular endothelial growth factor receptors 1 to 3	FLT1, KDR and FLT4
fibroblast growth factor receptors	FGFR1, FGFR2, FGFR3 and FGFR4
frizzled family receptors	E.g. FZD4, FZD10
Proteinases	**Proteinase gene symbols**
matrix metallopeptidases	MMP1, MMP11, MMP2, MMP9, MMP14 and MMP3 etc
plasminogen activator, urokinase	PLAU (u-Pa)
plasminogen activator, tissue	PLAT (t-Pa)
Adhesion molecules	**Adhesion gene symbols**
Integrins	ITGA5/ITGB3, ITGA5/ITGB1
N-cadherin	CDH2
VE-cadherin	CDH5
junctional adhesion molecule 3	JAM3 (JAM-C)
Signalling molecules	**Signalling gene symbols**
Protein kinase A (PKA)	PRKACA
protein kinase B (PKB)	AKT1
protein kinase C, alpha	PRKCA
mechanistic target of rapamycin (serine/threonine	MTOR

kinase)	
cytochrome c oxidase subunit II	COX2
ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)	RAC1
Harvey rat sarcoma viral oncogene homolog	HRAS (Ras)
v-raf murine sarcoma viral oncogene homolog B1	BRAF (Raf)
mitogen-activated protein kinase 1	MAPK1 or MAPK8
Metamorphic or guidance	**Metamorphic or guidance genes**
angiopoietins/tyrosine kinase with immunoglobulin-like and EGF-like domains 1	ANGPT1, ANGPT2, ANGPT4 and TIE1, TEK
Ephrins	EPHB4, EPHB3, EPHA2, EPHA7 and EPHB2
Netrins (deleted in colorectal carcinoma)	DCC, UNC5B, UNC5C
Semaphorins	SEMA7A, SEMA4B
Collapsin response mediator protein family	CRMP1, CRMP2, CRMP3, CRMP4 and CRMP5
Roundabout genes	ROBO1, ROBO2, ROBO3 and ROBO4
Slit homolog genes	SLIT1, SLIT2 or SRGAP3
delta-like 4 and notch signalling	DLL4 and NOTCH1
Transcription factors	**Transcription factor gene symbols**
hypoxia inducible factor 1, alpha	HIF1A
inhibitor of DNA binding 1 and 2	ID1, ID2
Kruppel-like Factor	KLF2
forkhead box TFs	FOXO1 and FOXO3
homeobox genes	HOXB3 and HOXD3
NF-kappa B	NFKB1
early growth response 1	EGR1
SRY (sex determining region Y)-box	SOX4, SOX18, SOX7, SOX13

Table 1.7.1; A list of some of the genes that have been implicated in angiogenesis, which could be used as a target in anti-angiogenic therapy. This table is based on previously published literature [23].

1.8 Tumour micro-environment and endothelial cell gene expression repertoires

Tumour vascular targeting is possible because the tumour micro-environment is vastly different to normal conditions, which gives rise to a different gene expression profile and cell surface repertoire of proteins. The environment of a tumour and consequently the endothelium itself is very different to that in normal tissues. The tumour endothelium and adjacent stroma are physiologically and architecturally different to that in corresponding normal environments. Tumour vasculature has the characteristics of being tortuous, random and with a disorganised vascular architecture [30-32]. The walls of vessels are immature and poorly developed with an irregular and discontinuous endothelial cell inner lining and with an incomplete covering of smooth muscle cells, coupled with weak junctions between endothelial cells and accompanying pericytes [30, 33, 34]. Vessels are usually built around a variable and often incomplete basement membrane of aberrant anatomy. The randomness, varying vessel diameters and discontinuous vessel network leads to a high interstitial pressure. The blood flow around tumour vessels is reduced, stopping and starting, which makes tumours susceptible to changes in pressure. In tumours, red blood cells and haemoglobin are deficient in oxygen, which can lead to variable levels of tissue oxygenation, sometimes as low as 1%. Some regions of a tumour are more hypoxic than others [35]. Because of the extensive anaerobic respiration of tumour cells, a high level of lactate is present and this produces oxidative stress [36]. The tumour environment is also starved of glucose. All these physical forces and attributes of the tumour micro-environment lead to endothelial cells being in an active angiogenic state, as against a quiescent state found in normal tissues [37]. This ultimately leads to a different gene expression profile that is specific to tumour endothelium and this can be exploited using vascular disrupting agents (VDA), which specifically target the established tumour blood vessels [32]. Most treatments of solid tumours are unspecific and since tumour vasculature is irregular and under interstitial pressure, chemotherapy agents lack tumour mass penetration, which is not the case for VDAs.

1.9 Treatment of cancer and vascular targeting

As early as 1923, it was proposed by Woglom that damage of tumour capillaries may have the highest efficacy in slowing tumour growth [38]. Subsequently, in 1990, Denekamp observed that cancer therapeutics were affecting tumour vasculature and it should not be ignored [39]. However, even more compelling evidence was demonstrated in 1993 when Burrows et al. destroyed a tumour by targeting the ricin poison to a marker induced to be expressed on the tumour endothelium [40]. Studies have shown that phosphatidylserine becomes exposed on the surfaces of tumour vasculature and can be targeted by an antibody Bavituximab, coupled to a radioactive Arsenic analog. This approach to targeting tumour vasculature has been used in imaging studies and highlights that targeting is possible [41]. It should be noted that there is a differentiation between targeting the existing tumour vasculature and targeting angiogenesis molecules expressed through active angiogenesis. These are both classed differently: anti-angiogenics, or angiogenic inhibitors = AIs and vascular disrupting agents = VDAs. These are different approaches and vascular targeting aims to shut down existing tumour vasculature to starve a tumour of oxygen and nutrients, plus remove lymphatic vessels that take away cellular waste (VDAs). In contrast, anti-angiogenics or angiogenic inhibitors aim to stop the active angiogenesis process in tumours (AIs).

In a review by Loi et al. it is stated there are two classes of VDAs, which are 1) small molecules and 2) ligand based, which are agents like antibodies that carry a therapeutic agent to a specific tumour endothelial target for treatment [32]. The flavone, acetic acid, was demonstrated to retard the growth of tumours grown in mice, which was attributed to a reduction in vessel perfusion [42]. Other studies of the effects of flavonoids on endothelium show that this class of molecules relax vessels, which could be a reason for reduced vessel perfusion in the mouse tumour models [43].

An example of the targeted ligand approach, one which is a promising therapy, involves coupling an antibody to liposomes that contain the therapeutic agent. This way, a characteristic of tumours is exploited, the characteristic that tumours harbour liposomes because of vessel architecture. Liposome delivery of therapeutics therefore give a more

potent hit because of this unique architecture [44]. One such case, using targeting peptides that have an affinity for integrins αvβ3 expressed on tumour vasculature, was to deliver liposomes containing the therapeutic agent Combretastatin A4 to tumours [45].

1.10 Reasons for vascular targeting

Membrane or secreted vascular targets are an attractive option as they are in contact with blood and as such an injected therapeutic can be easily administered to the blood supply. Vascular targets also offer an advantage over direct tumour targets as tumour targets can sometimes become obsolete due to the genomic instability of cancer cells, whereas tumour vasculature is genetically stable. It is also possible that some receptor targets could internalise the therapeutic, thus maximising the potential therapeutic hit. An example of successful homing of ligands to receptors was performed in a mouse prostate disease model, using a peptide found by phage display technology. They coupled a pro-apoptotic protein to a phage protein which demonstrated a delay in the formation of prostate tumours in a genetically engineered mouse, prone to prostate tumours [46].

1.11 A recent history of vascular target discovery

1.12 Silica pellicle

In early vascular targeting studies, silica pellicle coated endothelium was used to isolate membrane proteins [47]. Silica gel was perfused into blood vessels of rat where it formed a thin layer onto the surfaces of endothelial cell membranes. Once set, the coated tissues were extracted and fragments isolated using gradient centrifugation. The extracted regions of coated plaslemma were then processed biochemically and subject to proteomic mass spectrophotometer analyses in order to identify plasma associated proteins. A similar procedure has been patented [48]. An endothelial caveole protein PV-1 was discovered this way in the rat, for which there is the human ortholog, PLVAP [49, 50]. An antibody, called Pathologische Anatomie Leiden-endothelium or PALE, has been used as a marker for vascular endothelium for greater than 20 years without the knowledge of which gene product it was binding to. It has now been discovered that this antibody binds to the gene PLVAP [51]. Using a similar approach to the silica method,

Oh et al. identified Anexin A1 and aminopeptidase-P as endothelial cell membrane markers, using a pipeline of: subcellular fractionation, subtractive mass spectrometry proteomics and bioinformatic analyses [52].

1.13 Chemically coupling membrane and extracellular proteins to biotin esters

In this technique, vascular target genes were identified utilising a chemical reaction that occurs between injected biotin and proteins that are in contact with the blood supply (extracellular or membrane). The reaction occurs between a functional group of biotin and the terminal amine groups of proteins. Following careful homogenization of the biotin perfused tissues using a detergent, biotin coupled proteins were isolated using streptavidin beads, and subject to trypsin digestion and tandem mass spectrometry/chromatography analyses for protein identification. This technique has been applied *in-vivo* to mice with subcutaneous tumours [53-55] and, more recently, to *ex-vivo* human resected organs with tumours [56]. From the mouse experiments, Rybak et al. [53] listed 9 proteins that were specific to F9 and/or RENCA tumours. The human orthologs of these 9 proteins are: ITGAV, COL4A2, PTK7, ACACA, VTN, TKT, TIAM1, TNNI2 and GBA. This *in-vivo* biotinylation method has been customised by removing detergents and applying hot water for the recovery process and, for the first time, is applied in the chicken embryo, granulation tissue and CAM implanted tumour tissue (see chapter 2).

1.14 Serial Analyses of Gene Expression (SAGE)

St. Croix et al. used a comprehensive procedure to isolate endothelial cells from normal colon mucosa and tumour colon. For positive antibody selection, they proposed a better marker of endothelial cells than had been previously used [57, 58], which was MCAM rather than CD31 as this gene is also expressed on macrophages. They included several negative selection purification steps to remove non-endothelial cells, using antibodies to EPCAM (epithelial cells), CD45 (leukocytes), CD64 (macrophages) and CD14 (monocyte cells). They validated purity of endothelial cells in two ways: 1) by comparing semi-quantitative PCR results of bulk tumour with isolated endothelial cells, using markers of endothelium and epithelium, and 2) quantification of SAGE tags for a

selection of tissue markers in 4 pools of SAGE libraries (endothelial cell libraries from positive bead selection pool, hematopoietic cell pool from the negative selection purification procedure, cultured HUVEC pool and the fourth pool, cancer epithelial cell lines).

They identified pan endothelial markers (PEMs) by comparing highly expressed genes in both tumour and normal colon cells, with 1.8 million SAGE tags derived from cancer cell lines. They chose genes as PEMs if their expression was ≥ 20 fold higher in endothelium versus cancer cell lines. *In-situ* hybridization was carried out on frozen human tissue sections to validate PEM3 and PEM6, whilst other PEMs were validated based on SAGE expression patterns in cultured endothelial cells (HUVEC and human dermal microvascular endothelial cells (HMVEC)). They noted that several genes highly expressed *in-vivo* were not expressed in cultured endothelial cells, which was exemplified by the Hevin (SPARCL1) gene, where they showed clear *in-situ* hybridization staining in colon tumour endothelium *in-vivo* but was absent from HUVEC SAGE libraries. Tumour Endothelial Markers (TEMs) were also identified; 46 tumour colon endothelial enriched SAGE tags were found with ≥ 10 fold enrichment in tumour endothelium versus normal endothelium and they focused on the 25 tags most enriched (23 genes), more specifically the 9 novel tags for which only EST data was available. Firstly, they showed a positive endothelial cell control from normal and tumour derived endothelial cells using semi quantitative PCR for both VWF and PEM6 (C-type lectin domain family 14, member A). They were highly expressed in endothelial cells of tumour and normal colon but absent from an epithelial control. TEM1 (endosialin), TEM7 (plexin domain containing 1) and TEM9 (mannose receptor, C type 2) were all absent from the epithelial control and normal colon endothelium but strongly expressed in tumour colon endothelium. Both TEM7 and TEM1 clearly stained tumour endothelial cells from *in-situ* analyses. TEM7 was further shown to stain endothelium on colon metastasis to the liver, and also breast, lung, pancreatic, and brain primary tumour endothelium. Corpus luteum vessels also showed TEM7 staining, demonstrating overlap between tumour and physiological angiogenesis. This article, along with 9 TEM genes, has been cited by many subsequent related articles.

1.15 cDNA or Expressed Sequence Tag (EST) gene discovery

cDNA library sequencing is the high throughput and systematic single pass sequencing of bacterial clone inserts, picked at random, from a cDNA library representing poly-A mRNA from a cell type of interest. It facilitates a snap shot of the gene expression signatures of a tissue. Although the transcriptome coverage is low, because of the prohibitive cost of sequencing and labour involved, many groups have successfully used this method to identify novel genes and genes enriched in target tissues [59-63]. Sequences range from 200 to 600bp in length and libraries often contain less than 10,000 sequences.

Huminiecki and Bicknell [63] explored public expression databases and combined differential gene expression analyses of cDNA and SAGE libraries to produce a list of 37 endothelial markers (4 of them novel). The bioinformatics analyses were done in two stages: 1) The first approach used BLAST to search two pools of cDNA libraries, one pool endothelial and other pool non-endothelial cell lines, versus Unigene sequences. 2) The second approach utilised the xProfiler software [64] where a similar screen was performed with public SAGE libraries. The intersection of results were the 37 endothelial genes reported. Two of the most promising candidate genes, ROBO4 and TXNDC5 were further experimentally validated using *in-situ* hybridisation and were shown to be tumour endothelial enriched [65, 66].

Herbert et al. performed an extended version of the Huminiecki and Bicknell analyses using an EST to gene assignment that negated SAGE library confirmation. In the 8 year time lapse since the Huminiecki and Bicknell publication, many new cDNA and SAGE libraries had been added to public databases. The latest SAGE and cDNA libraries were combined, employing likelihood ratio statistics, to identify 459 endothelial gene candidates including CLEC14A, MMP1, ESM1, RHOJ and ECSCR [62]. RHOJ has subsequently been functionally investigated [67].

1.16 Differentiating physiological and pathological angiogenesis

Other researchers aimed to differentiate physiological and pathological forms of angiogenesis. Van Beijnum et al. performed an elegant study where suppression subtractive hybridization (SSH) experiments were followed by microarray analyses to distinguish genes enriched in tumour angiogenesis, *in-vivo* general angiogenesis and *in-vitro* general angiogenesis [68]. They isolated colon tumour endothelial cells (TEC) and normal colon endothelial cells (NEC) using a modified version of the method reported by St. Croix et al. [69]. This involved combining antibodies for positive selection: the EN4 antibody (reported to be anti CD31 [70]), another anti-CD31 antibody and an anti-CD34 antibody. Bead isolated TEC, NEC and placenta endothelial cells (PLEC), plus tumour conditioned HUVEC (T.HUVEC) and quiescent HUVEC (Q.HUVEC), were subject to suppressive subtractive hybridisation experiments. The resulting 2,746 cDNA clones found by SSH were spotted on to microarray filters, where labelled cDNA (^{33}P-dCTP) from the different endothelial cells types were hybridized to find differentially expressed genes. In the results they present 17 tumour angiogenesis genes (TAGs), 46 *in-vivo* general angiogenesis genes (GAG) and 22 *in-vitro* angiogenesis genes (GAG/B). 15 of the TAGs were confirmed by qPCR both in TEC vs. NEC and TEC vs. PLEC contrasts. The TAGs CD59, HMGB1, IGFBP7 and VIM were then validated at the protein level; Immunohistochemistry results showed that all 4 genes were over-expressed in tumour colon endothelium, with CD59 and IGFBP7 producing the most vascular specific staining pattern. They then tested if these genes had a role in angiogenesis *in-vitro* by applying antibodies to these genes onto a collagen sprouting angiogenesis assay. Three out of the four genes inhibited angiogenesis in this model (HMGB1, IGFBP7 and VIM). The chicken chorioallantoic membrane (CAM) was then used in an *in-vivo* assay to see if the same antibodies could inhibit *in-vivo* angiogenesis. As in the *in-vitro* collagen sprouting assay, the same 3 genes inhibited up to 45% of the angiogenesis in the CAM. This suggests that embryonic chicken undergoes similar angiogenesis to tumours. Finally, they performed an experiment in Athymic mice with implanted colon tumours, where the gene VIM was targeted with an antibody. It was shown that tumour growth was inhibited in a dose dependent way and that the microvessel density of the treated mice was reduced compared to controls.

Likewise, Seaman et al. endeavoured to dissect pathological and physiological angiogenesis using mouse endothelium from a variety of tissues and angiogenic states [71]. To achieve this, they used mouse tissues from normal liver, liver tumour and regenerating liver (also undergoing angiogenesis). In addition, they too distinguished gene signatures of endothelial cells from different organs and identified 15 liver specific and 27 brain specific endothelial markers. To do this, endothelial cells were isolated from 7 normal mouse tissues including: brain, heart, lung, kidney, liver, muscle and spleen. In addition, they also extracted endothelial cells from regenerating liver at 3 time points and tumour endothelial cells from 5 human/mouse tumour types including liver (subcutaneously for all tumour cells implanted into mice, except for the liver metastasis which was performed intrasplenically). The implanted tumour cell sources were:

- Mouse colon carcinoma cell line called CT26 [72]
- Mouse mammary carcinoma cell line called EMT6
- Human colonic carcinoma cell lines KM12SM
- Human colonic carcinoma cell lines HCT116
- Human colonic carcinoma cell lines SW620
- Mouse Lewis lung carcinoma cell line LLC
- Human colon adenocarcinoma cell line LS174T

The endothelial cells were isolated using a magnetic antibody coupled beads protocol that included 2 negative selection stages: 1) antibodies for the CD19 (B-lymphocyte antigen) and 2) CD45 (protein tyrosine phosphatase, receptor type, C or leukocyte common antigen). Just prior to positive selection, they were careful to prevent random and non-specific binding to any FC receptor displaying cells by adding anti CD16/32 antibodies to the cell mixtures. Positive selection was carried out by using 3 combinations of 2 antibodies, depending on tissue source. For heart, liver, kidney, intestine, lung, KM12SM and LS174T samples, the positive endothelial cell antibody selection was done with Endoglin (CD105). For brain, muscle, SW620 and EMT6 samples the positive selection was achieved using a mixture of the antibodies to CDH5 (vascular endothelial cadherin) and Endoglin. For the spleen, LLC and CT26 samples, endothelial cells were isolated

with antibodies to CDH5 alone. Once the endothelial cells were isolated, they were subject to SAGE library sequencing and analyses to find differentially expressed genes enriched in tumour, angiogenic and normal endothelium from normal organ tissues. Quality control of endothelial isolates was performed using qPCR comparing whole tissues with the isolates; they used the endothelial gene ve-cadherin (CDH5) and showed ≥ 200 fold enrichment of this marker in isolates versus whole tissues. QPCR was also used to validate genes that were up-regulated in regenerating liver endothelial cells and/or tumour endothelial cells. 25 genes were validated as tumour endothelial enriched, 13 of which were not expressed in regenerating liver (physiological angiogenesis). Then Seaman et al. also demonstrated 2 brain (Slc2a1 and Slco1a4) and liver (Dnase1l3 and Oit3) specific endothelial markers respectively, using *in-situ* hybridisation. All 4 genes clearly stained endothelium in their respective tissues and staining patterns were matched using a positive control endothelial staining pattern of the VEGF receptor 2. Several tumour endothelial markers were then validated using *in-situ* hybridisation in a serious of different tumours: genes included Apelin, Doppel, Cd109, Cd137, Cd276, Ptprn, Mirp2, Ets4 and Vscp that all showed positive vessel staining. A comparison was made of Cd276 between tumour and normal endothelium. The human ortholog of CD276 was then tested with a mono-clonal antibody on a series of paired tumour and normal patient colon mucosa samples. A recombinant protein of CD276 was used as a positive control and it was evident that CD276 was enriched in the tumour samples from all patients, though it was not absent in normal colon mucosa. A range of other tumours: non-small cell lung carcinoma, oesophageal cancer, bladder and breast cancers were stained with a poly-clonal antibody and these all showed positive staining for CD276 on tumour endothelial cells. Then more proof was provided by immunohistochemistry to CD276 and VWF, which co-localised in the tumour colon endothelium but was absent in normal colon endothelium (only VWF was present). Finally, the manuscript lists 27 brain specific endothelial markers (25 of which have a human ortholog), 15 liver specific endothelial markers (with 13 orthologs in human), 12 angiogenic endothelial markers (12 orthologs in human) and lastly, 13 tumour endothelial markers (all with human orthologs).

Madden et al. isolated brain endothelial cells from tumour and normal brain samples using antibody coupled beads [73]. They removed leukocytes using CD45 antibody coupled beads before a positive antibody-bead selection with either anti-MCAM or an endothelial-specific plant protein Ulex europaeus agglutinin 1 (UEA-I), which binds to alpha-L-fucose on the surface of endothelial cells [74]. This was followed by Serial Analyses of Gene Expression (long SAGE [75, 76]) sequencing to reveal 16 genes that were ≥ 4 fold up-regulated in glioma tumour endothelium. They also combined their brain SAGE libraries with published colon tumour and normal SAGE libraries from St. Croix et al. [69] to identify common tumour induced endothelial genes. Finally, they also presented a table of genes that had a tumour endothelial specific expression pattern, with little or no expression in normal brain and no counts in other cell types (public SAGE libraries). In total, the combined list of tumour endothelial up-regulated and tumour endothelial specific genes was 21. They focused on the plasmalemma vesicle associated protein (PLVAP) using *in-situ* hybridisation, custom microarray and PCR. The *in-situ* staining showed PLVAP to be up-regulated in glioma tumour endothelium. PCR/microarray results showed that PLVAP was up-regulated in a matrigel medium versus endothelial cells grown on plastic plates, thus concluding PLVAP is expressed under active angiogenesis. This is a very interesting result since PLVAP has been previously shown to be the antigen of the PALE antibody that has been used for several years to stain vasculature (see above).

Bhati et al. performed microarray analyses on micro-dissected breast tumour and normal endothelial cells to find vascular targets. They reported a list of 48 known genes that were greater than four fold up-regulated in tumour endothelial cells versus normal endothelial cells. Immunohistochemistry experiments on breast tumours validated several of these genes: FAB, SFRP2, JAK3 and SMPD3 [77].

1.17 Phage display technology and vascular zip codes

Pasqualini et al. coined the term vascular zip codes (post codes) in a review that illuminated research on vascular targeting [78]. Vascular post codes refers to the protein heterogeneity between different vascular beds of various tissues and the vascular surface

protein differences between normal and pathological states. The post codes enable the body to deliver precise and appropriate cells or molecules to specific sites in the body; one example is the homing of haematopoietic stem cells (HSCs) to the bone marrow, which is mediated by a combination of α4β1 integrins, vascular cell adhesion molecule 1, β2-integrins and/or selectin genes [79]. Because of these different vascular post codes, depending on tissue type or pathological state, an opportunity exists to target these vascular tissue specific post codes for therapeutic or imaging purposes. The key to this is to find luminal targets specific to tumour vasculature. Aside from *in-vivo* biotinylation, whole genome arrays and sequencing technologies as discussed previously, there exists a combinatorial approach that aims to find both protein receptor post codes and produce ligands for these receptors, at the same time. This is called phage display screening [78], which is a high-through-put technique to decipher protein-protein interactions (in this case, receptor and ligand interactions). This procedure can be performed *in-vitro* or *in-vivo*.

In the phage display system, fragments of DNA coding for 8 or 9 amino acid ligand peptides are fused to capsid sequences, which code for phage protein coats, so that ligands are present (displayed) on the surface of filamentous phage, hence the phage display name [78, 80]. The 8 to 9 amino acid ligands are specifically designed to maximise biological effect, usually containing a central tyrosine residue, without jeopardising the bacteria phage. Many combinations of ligands can be produced and these can be used in panning experiments on a selected surface or microenvironment of interest. This has also been achieved *in-vivo*, where phage display was used to map ligands to different vascular beds of different organs and tumours to facilitate therapeutic delivery or imaging [81]. One of the peptide ligands was to interleukin 11, which binds to both endothelium and prostate epithelium (also seen by others [82, 83]). They validated the peptide mimic of interleukin 11, using a receptor-ligand binding assay where immobilised interleukin 11 receptor bound the phage display ligand peptide. This was subsequently confirmed with competitive binding with interleukin 11 itself, versus the mimic peptide.

1.18 DNA and sequencing technologies

1.19 Sequencing of DNA/RNA

The polymerase chain reaction and the sequencing of DNA molecules underpins many advances in molecular biology. Without being able to determine the sequences of A, G, C and T embedded within genomes and transcriptomes, vascular targeting would be limited to protein based methods. Although protein methods are good, without the technique of PCR to amplify regions of DNA or whole gene sequences, many experiments we take for granted would not be possible. The technique of sequencing cloned DNA (cDNA) representing poly-A mRNA has facilitated gene discovery and also has indirectly led to the production of whole transcriptome gene expression arrays. As this work relies heavily on DNA and sequencing, a proportion of this introduction is devoted to illuminating a brief history of DNA and sequencing, which goes on to briefly describe the modern techniques of 2nd generation sequencing, which have been heavily utilized in this work.

1.20 Quick history of DNA

It is apparent that a scientist in 1868, Friedrich Miescher, noticed a substance derived from leukocytes that he could not class as protein or fat. Because it originated from the centre of the cell, the nuclei, he termed it nuclein. At one stage, he even proposed that this new substance could contain the heritable information of cells [83] but later rejected it. So it was Friedrich Miescher who first became aware of the existence of DNA. In the year 1878, another scientist, Albrecht Kossel, identified the organic chemical bases of DNA and RNA (Guanine (G), Cytosine (C), Thymine (T), Adenine (A) and Uracil (U)). Later in 1919, an American-Lithuanian scientist identified the sugar/phosphorus building blocks of DNA, which he noted were composed of Deoxyribose, Guanine, Thymine, Adenine, Cytosine and Phosphorus [84].

The first system to show and prove that DNA carried heritable information was provided by Frederick Griffith. He showed that injection of mice with a mixture of heat killed "smooth" bacteria and live harmless "rough" bacteria would result in the transformation of the harmless to the virulent ("rough" to "smooth") strain. It was Avery, Macleod and

McCartney who later proposed that this transformation was caused by the transfer of DNA rather than transfer of proteins between bacterium [85, 86]. Some years later, in 1937, the first evidence that DNA was a regular structure was provided from the first x-Ray diffraction pattern of DNA, which was produced by William Astbury [87]. Research on DNA structure intensified between 1944 and 1953 that led to the recognised structure of DNA.

Between 1949 and 1951, it was Erwin Chargaff who first noticed that the base content or genetic composition differed between different species and simultaneously he observed that, within the same species, nucleotide bases were evident at fixed ratio levels. These findings mean that there were the same number of guanine bases as cytosine bases and likewise for the adenine and thymine bases [88, 89]. This was a big step in the right direction. It was finally confirmed for sure that DNA was the makeup of heritable genetic material. Chase and Hershey showed this by observing the fact that filamentous phage inject mostly DNA into cells rather than proteins, when phage infection of bacteria is taking place [90]. It was Rosalind Franklin and Maurice Wilkins who decided to produce a crystal of DNA to be used in x-Ray crystallography experiments. This extended the earlier work by Astbury, which was not successful. Finally, in the end, Rosalind Franklin produced two X-ray diffraction pattern photographs of high quality, which were taken by Maurice Wilkins and shown to Watson and Crick, who subsequently confirmed the double alpha helix model structure of DNA. It is apparent that Wilkins, Franklin, Watson and Crick all published data on x-Ray crystallography and the structure of DNA in the same Nature journal [91-95].

The basic structure of DNA is 4 bases in a double alpha-helix structure. The coded sequencing DNA or genome is what gives an organism its phenotype, and to better understand an organism as well as to find novel and differentially expressed genes, DNA sequencing is a pivotal process.

1.21 Maxam-Gilbert's early method of sequencing (1st generation)

A first method of DNA sequencing was produced in 1977 by Maxam-Gilbert [96]. They developed a method of sequencing using radio isotope labelling, chemical cleavage and fragment resolution via gel electrophoresis. First each 5' end of the molecule is labelled with a radioactive isotope (usually gamma-^{32}P ATP). Then the DNA is broken up by a series of 4 chemicals that aim to cleave the DNA strand only once and these can cleave at either a specific single base or at two base locations. Guanines are treated with a methylating agent (dimethyl sulphate) that aims to methylate a single guanine per molecule of DNA. This methylated guanine is then treated with piperdine to remove the methylated base and additionally cleave the DNA strand at this position. A second reaction is applied to a different aliquot of DNA, which also uses dimethyl sulphate but this time in the presence of formic acid. Under these conditions both adenine and guanine are both methylated and cleaved following the piperdine treatment. A similar process works for cytosine and thymine together using the chemical hydrazine alone, followed by piperdine. Or for cleaving Cytosine alone, hydrazine in the presence of NaCl followed by piperdine treatment is used. As cleavage occurs at different bases, the DNA fragments are all of differing lengths and the 4 separate chemically treated DNA can be resolved on a Polyacrylamide gel, where a radio-autograph produces an image of the isotope bands. Actual DNA sequences can then be finally deduced. Figure 1.21.1 displays an overview of the process.

Figure 1.21.1 Overview of Maxam-Gilbert's sequencing process

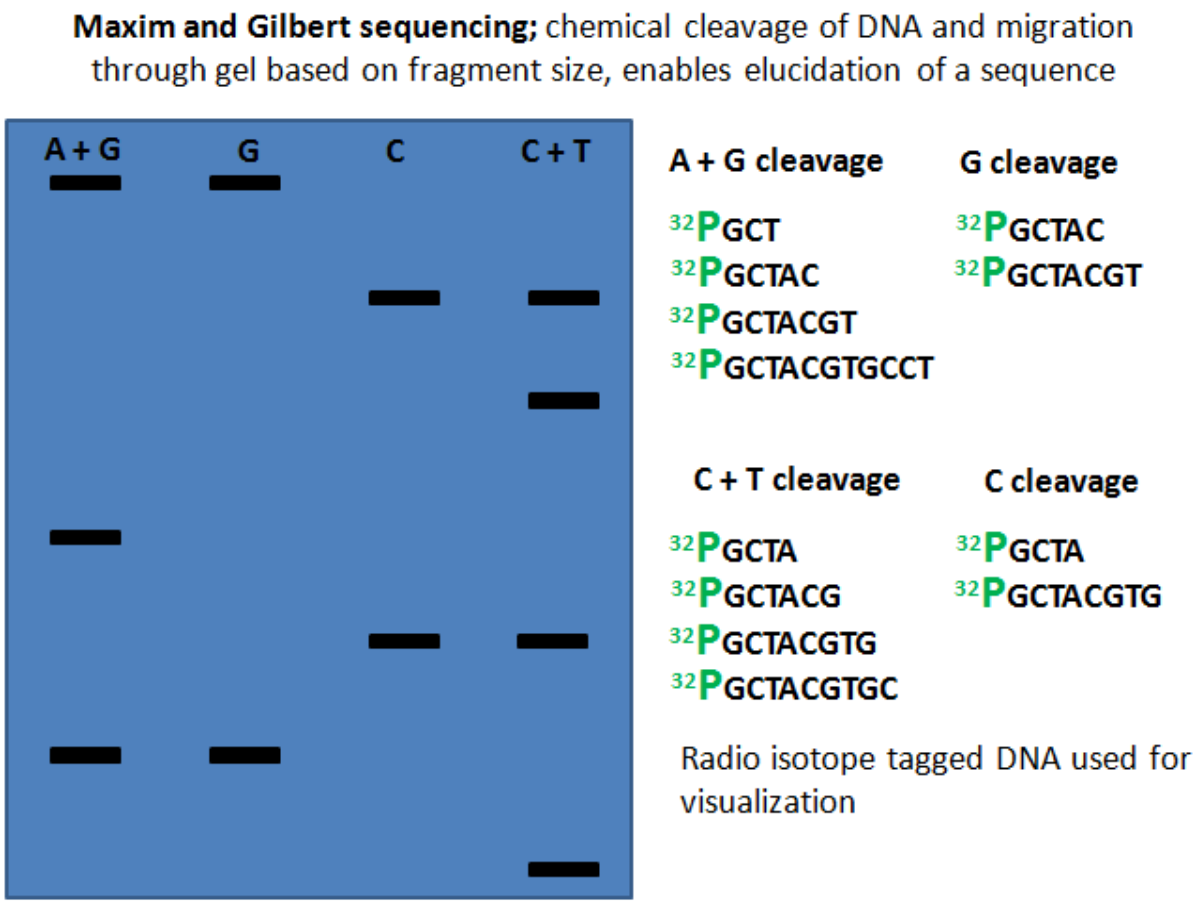


Figure 1.21.1 Gel electrophoresis separates out fragments of chemically treated DNA that have been cut at specific sites in the sequence. Based on fragment lengths and chemical digestions (in different lanes), a sequence can be deduced.

1.22 Sanger method of sequencing (1st generation)

The Sanger method of sequencing was brought about by Frederick Sanger in the year 1977 [97]. The molecular biology that is pivotal to this technique is the DNA chain terminators or ddNTPs (dideoxynucleotide triphosphates) and DNA polymerase that can synthesise DNA. ddNTPs terminate the sequence at the point of insertion during the polymerase phase of sequencing and these can be labelled with a radioactive isotope or, as is more recent, with a fluorescent label. The basic method involves denaturing the double stranded template DNA into single strands and then adding a short primer sequence that binds to this DNA. A second strand of DNA is then synthesised using the DNA polymerase, in the presence of both normal nucleotides and ddNTPs. The growing strand is randomly terminated when a ddNTP is incorporated into the synthesised strand. This produces many strands of differing lengths that cover the full sequence. Fragments are then resolved on a Polyacrylamide gel, initially having the 4 different bases run separately in different lanes, with radioactive nucleotide labels. Please see figure 1.22.1 for clarity. These are read on an autoradiograph. More recently, fluorescent labels replaced radio activity and all 4 bases could be combined into a single reaction. The human genome project used this method of sequencing, with the majority of machinery coming from Applied Biosystems E.g. the ABI 377 slab gel and/or the ABI 3700 capillary sequencer machines.

Figure 1.22.1 Sanger sequencing overview

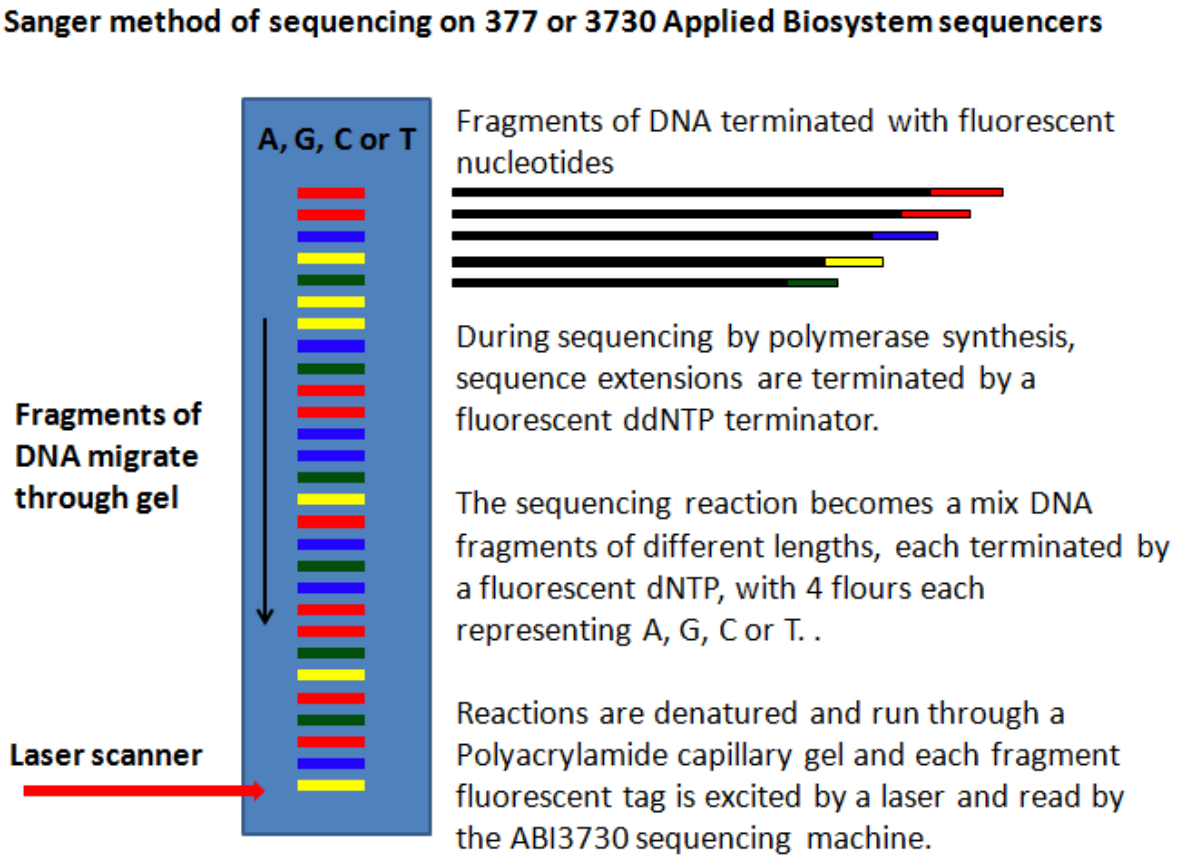


Figure 1.22.1 Polyacrylamide gel electrophoresis on a slab or capillary gel separates out different lengths of fluorescently labelled single stranded DNA. As fragments migrate through gel, they pass through a laser, which excites the attached dyes for base calls.

1.23 First generation to second generation

One problem with 1st generation sequencing, such as the Sanger big dye terminator sequencing, is that a lot of space is needed. You have to have a gel that can separate fragments of DNA of different lengths, each with a coloured ddNTP on the end. This will end up giving you a 500bp sequence but the problem is, you can only run 96 sequences per ABI 377 sequencer. Another inconvenience is that you had to pour Polyacrylamide slab gels several hours before a sequencing run. Even more fiddly was the loading of 96 samples onto the gel after carefully removing a comb. A welcome upgrade to the 377 was an Applied Biosystems ABI 3700, where you can run up to 96 different sequencing reactions using capillary gels instead of slap gels. This saved a lot of preparation time. However, even with these improvements, a major problem remained, which was that a maximum of 96 samples could be sequenced per run.

1.24 Illumina (Solexa)

2nd generation sequencing is completely different. If you consider the Solexa/Illumina technology, this immobilises millions of fragments of DNA onto a glass slide or flow cell. These separated single stranded DNA molecules are then bridge amplified so that you gain clonal clusters that enable enough fluorescent signal strength at each round of sequencing (Figure 1.24.1). This is then subject to sequencing by synthesis using a polymerase and reversible dye terminators of all 4 bases (dNTPs). The cycle of events in the reaction is: 1) A nucleotide is incorporated, 2) fluorescent signal is imaged, 3) the blocking chemical group is removed and 4) the next base is incorporated (Figure 1.24.2). Solexa was the original name of the technology that was launched in 2006 with the Solexa GI analyser that could sequence several million 32bp length sequences per run. In 2007 Solexa was bought out by another company, Illumina, who replaced the name Solexa. The latest machine, an Illumina HiSeq 2500 is able to sequence 600Gb of DNA/RNA per 11 day run (approximately equivalent to sequencing 66 human genomes to 30 times coverage). This can be paired end or single end sequencing of 100bp length reads. The technology that really advanced Illumina was the patent of the reversible dye terminator chemistry that was developed by Turcatti et al. [98].

Figure 1.24.1 Illumina sequencing, bridge amplification

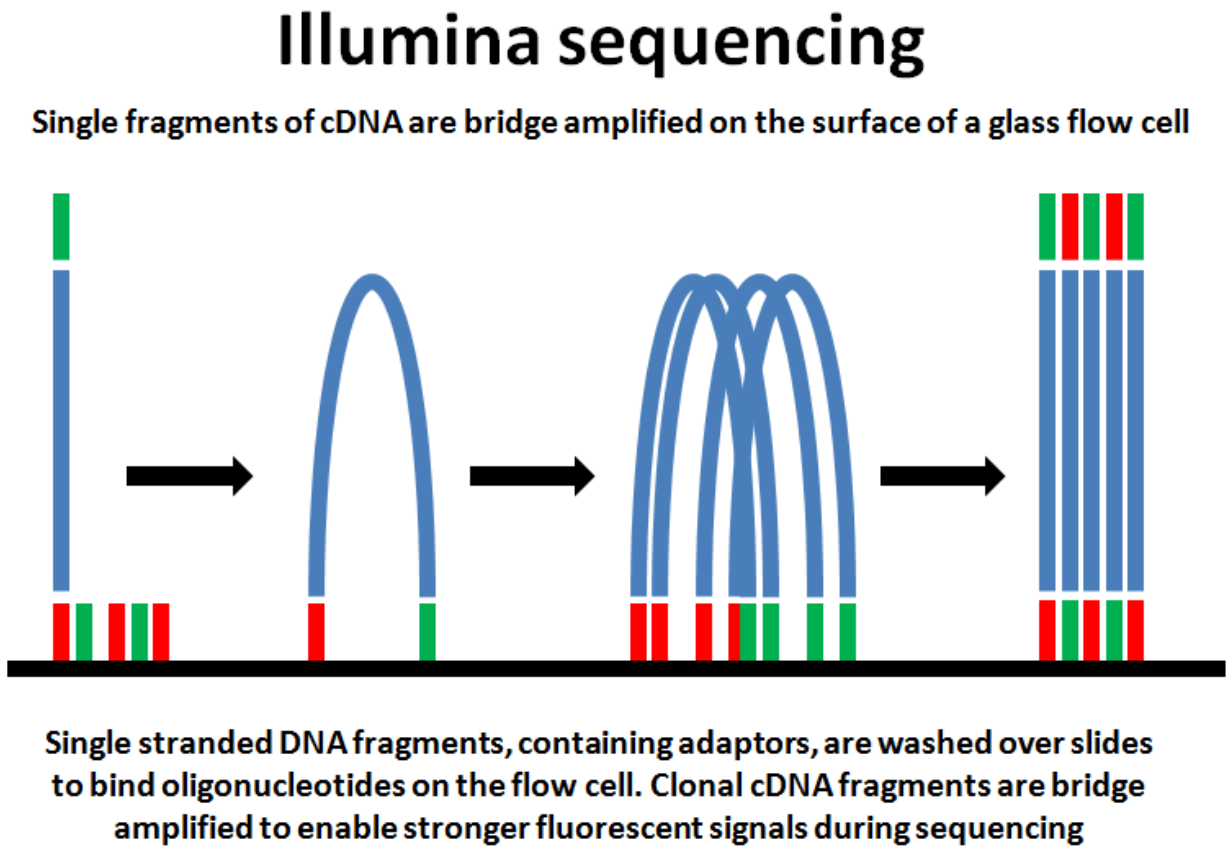


Figure 1.24.1 Single stranded and size selected cDNA with attached adaptors are washed through a lane of a flow cell. The adaptors complement slide bound oligonucleotides. Clonal clusters are amplified through a bridge amplification process to increase fluorescence signal.

Figure 1.24.2 Illumina, sequencing by synthesis using dye terminators

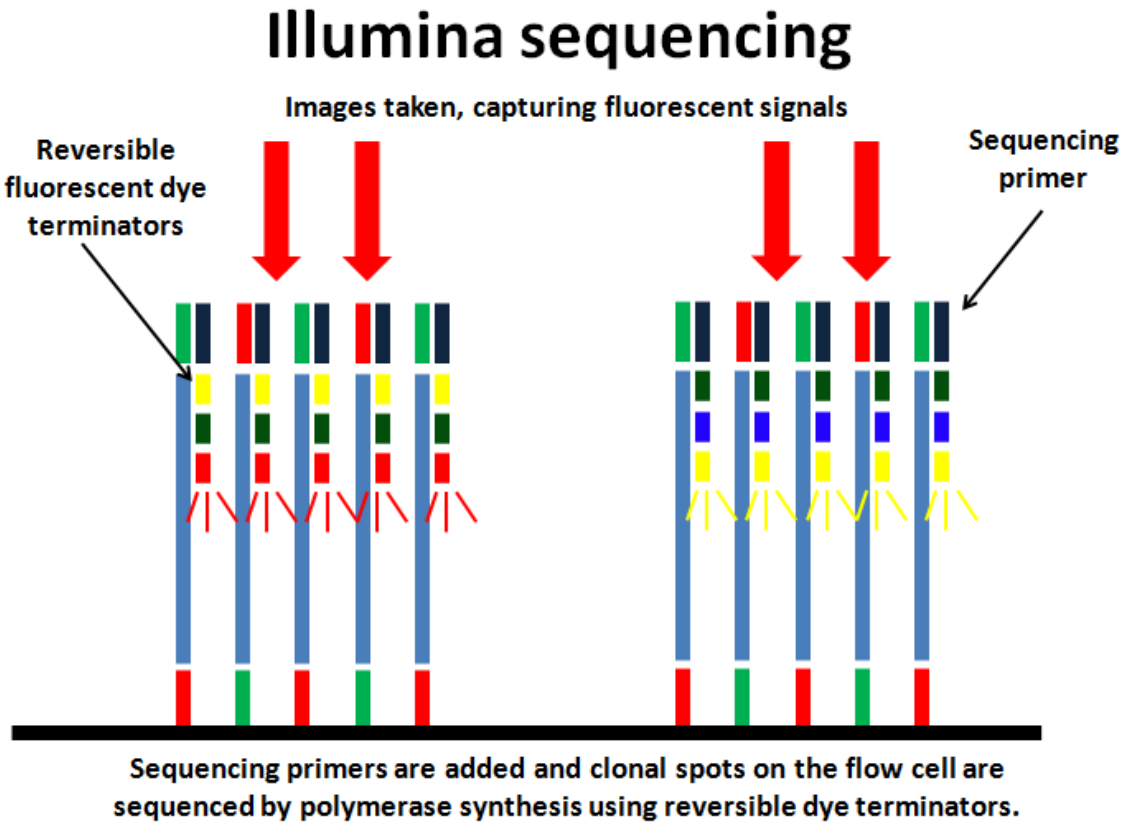


Figure 1.24.2 Illumina sequencing takes advantage of reversible dye terminators to sequence by synthesis fragments of DNA using a polymerase. As of 2012, typical sequencing runs produced ~80 million x 100bp paired end reads per lane of a flow cell. After each base incorporation, an image is taken of the flow cell to record base calls.

1.25 Roche 454 titanium

In 2005 Life Sciences launched the first 2nd generation sequencer called a 454. During RNA/DNA preparation for this technology, single stranded DNA molecules are bound to beads that are subject to clonal amplification using a technique called emulsion PCR. Here the reagents of PCR are in aqueous phase droplets within an oil solution. The aim is to have a single bead, with a single template of DNA, which is clonally amplified during the emulsion PCR process. The method of sequencing used is Pyrosequencing, that is a sequencing by synthesis method that adds each base (A, G, C or T) separately and then detects the release of pyrophosphate when a nucleotide or nucleotides are incorporated [99]. It utilizes a DNA polymerase and 4 different enzymes; the polymerase incorporates a base or bases depending on template and this produces a proportional amount of light depending on the number of nucleotides incorporated. When pyrophosphate is released, an enzyme called ATP sulfurylase converts pyrophosphate to ATP in conjunction with Adenosine 5' phosphosulfate (APS). In turn, the produced ATP compels the luciferase driven conversion of Luciferin to Oxyluciferin. This generates an amount of light that is proportional to the amount of ATP, which is consequentially proportional to the number of incorporated bases. The light peaks, after base additions, are analysed to produce a sequence, which is prone to poly-A tracts. The main difference if compared to Illumina is that the sequence read lengths are variable, from 200 to 500bp in length. However, the longer read lengths are offset by the fact that Roche 454 only produces at most 1 million sequences per run whereas Illumina can produce 6 billion x 100bp, paired end reads.

1.26 Applied Biosystems SOLiD sequencing and RNA-seq

A third 2nd generation sequencing platform is the SOLiD system. The scientific principles and methodology behind this technology were first reported by Shendure et al. [100] in 2005. This technology also uses the emulsion PCR to clonally amplify beads containing templates of DNA. This is followed by an enrichment of clonal beads that are then covalently immobilised onto a glass slide. The sequencing chemistry used with the SOLiD system is different to both Roche 454 and Illumina. Using a preparation method

of RNA-seq as an example, the SOLID system of sequencing will be described (Figures 1.26.1 to 1.26.5).

Poly-A mRNA is selected from total RNA using a poly-T cellulose medium, from a tissue of interest (e.g. tumour endothelial cells). This is then fragmented using an RNAse III enzyme digest and the resulting fragmented RNA is hybridised/ligated to adaptors. cDNA generation is achieved by reverse transcription and second strand synthesis. The purified cDNA is then size selected, amplified and finally subject to bead template preparation and emulsion PCR [101]. On the glass slide, the covalently bound beads are combined with a mixture of ligase, a universal sequencing primer and a large pool of di-base probes. These di base probes are labelled with four fluorescent dyes. Each coloured dye represents 4 out of the 16 possible di-nucleotide combinations. The complementary probe hybridises to the template DNA strand using the di-nucleotide sequence to interrogate the template. This probe is then ligated with the ligase enzyme and the fluorescent dye imaged. Following this, the dye is cleaved off, and 7 more cycles of ligation, imaging and cleaving occurs. The sequencing is performed in colour-space and each base in the template sequence is interrogated twice. The colour space sequences are finally converted to nucleotide sequences by using knowledge of the first base in the universal sequencing primer. A cartoon of the process is displayed in figure 1.26.4. A guide for the different di-nucleotide codes/colours and how a colour space sequence is decoded to base-space, is in figure 1.26.5.

Figure 1.26.1 Overview of RNAseq protocol (A); poly-A mRNA isolation

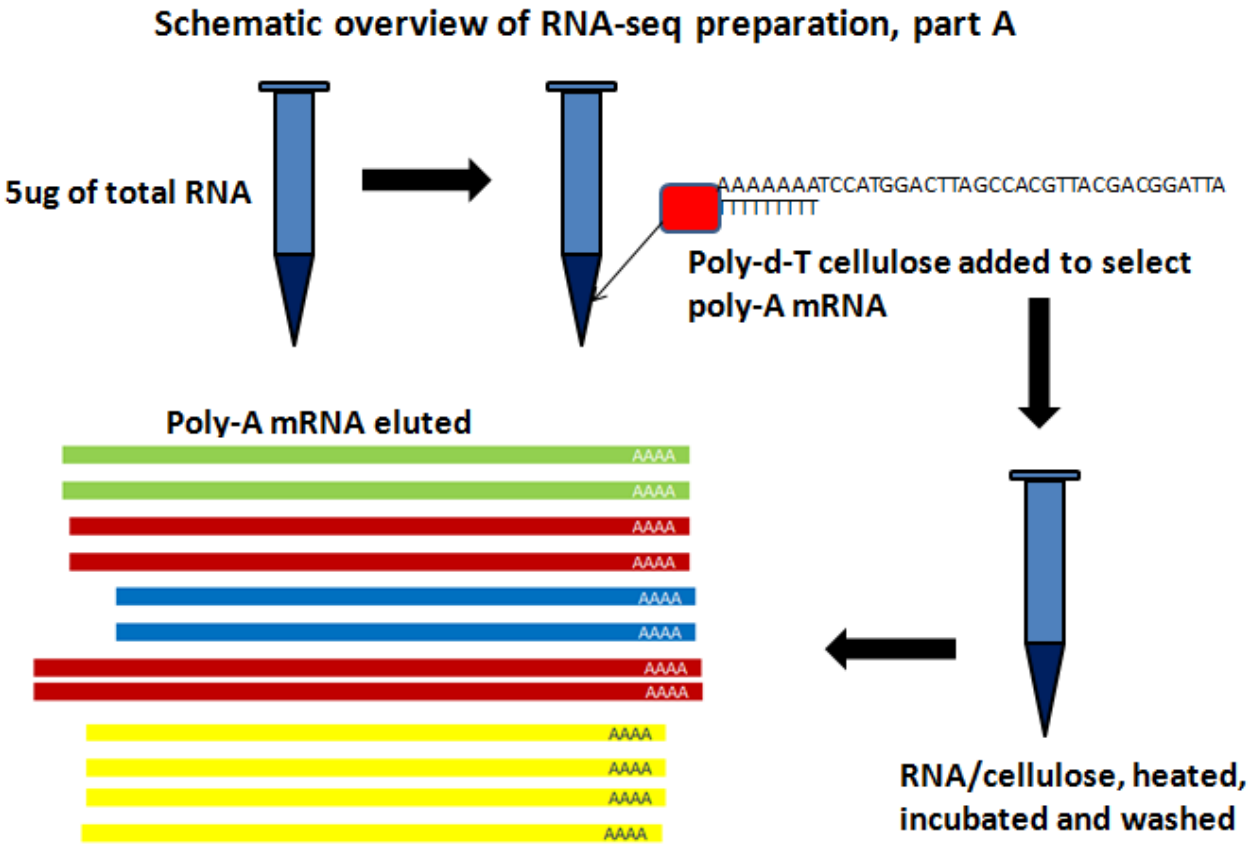


Figure 1.26.1 Basic overview of the experimental process of RNAseq (part A). Total RNA is isolated from endothelial cells (~5uG). This is incubated with cellulose containing poly-T sequences to bind poly-A tails of mRNA. Bound mRNA is washed and incubated before a final elusion to produce an enriched population of poly-A mRNA.

Figure 1.26.2 Overview of RNAseq protocol (B); fragmentation, cDNA synthesis

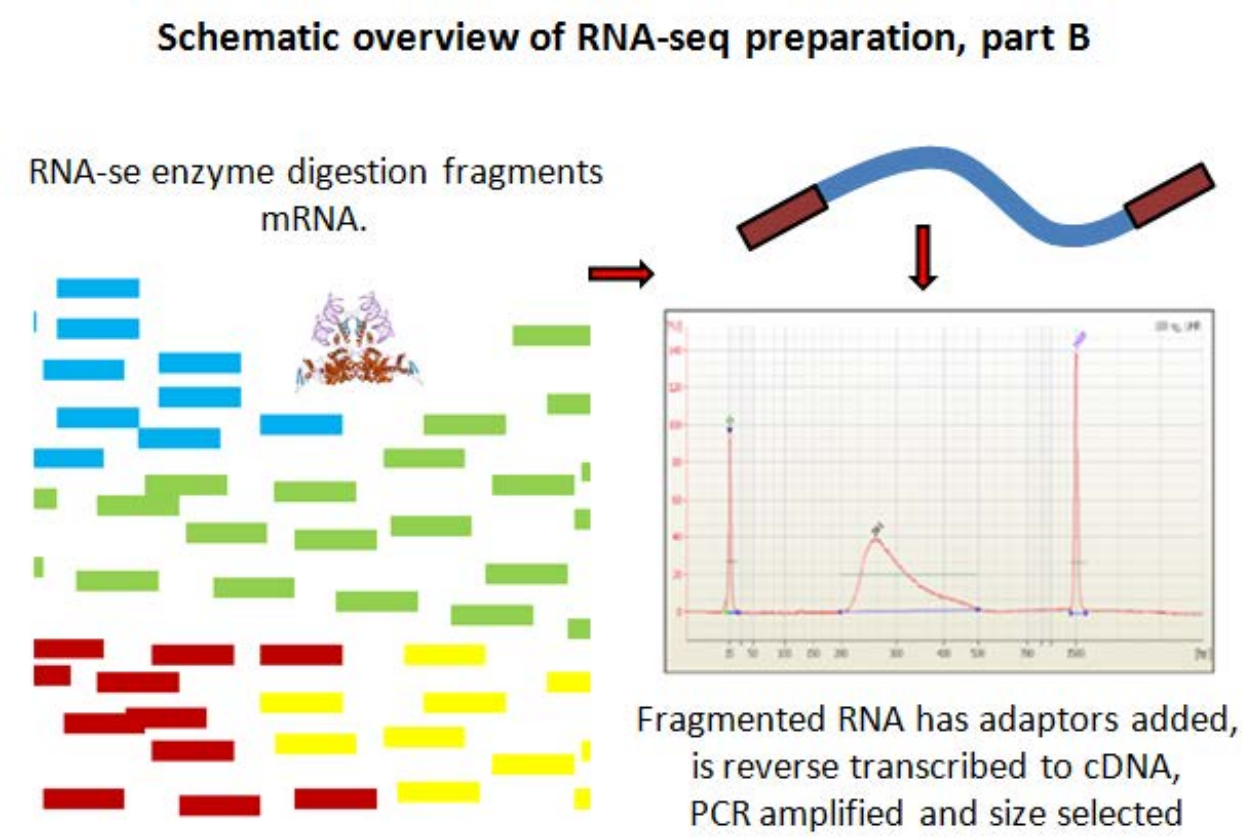


Figure 1.26.2; Basic overview of the experimental process of RNAseq (part B). RNA is hybridised/ligated to adaptors, followed by cDNA generation achieved by reverse transcription and second strand synthesis. The purified cDNA is then size selected, amplified and finally subject to bead template preparation and emulsion PCR (for Roche or SOLiD technologies).

Figure 1.26.3 Overview of RNAseq protocol (C); emulsion PCR and sequencing

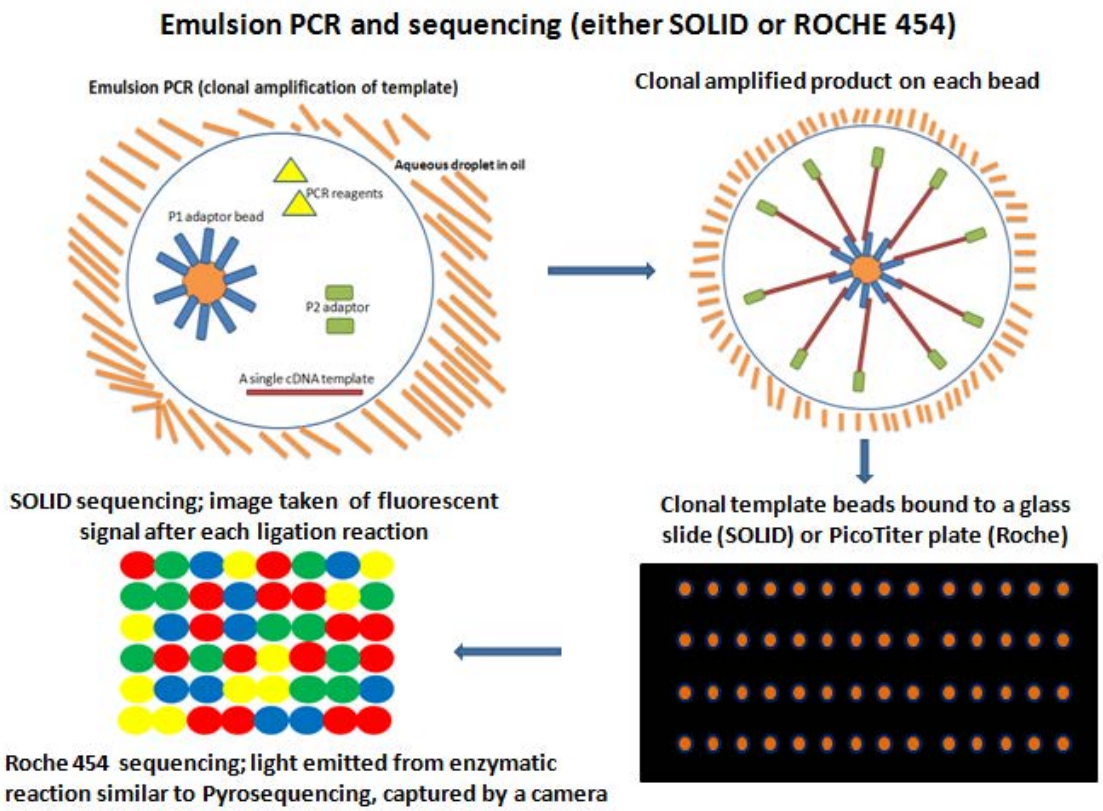


Figure 1.26.3 Basic overview of the experimental process of RNAseq (part C). Sequencing reagents (P1 adaptor beads, P2 adaptors and PCR reagents) are suspended in an oil/aqueous liquid. Single stranded cDNA molecules are mixed so that a single cDNA molecule in each bead droplet of the mixture is attained. Emulsion PCR takes place to amplify a clonal fragment, prior to sequencing.

Figure 1.26.4 Overview of SOLiD colour-space sequencing process

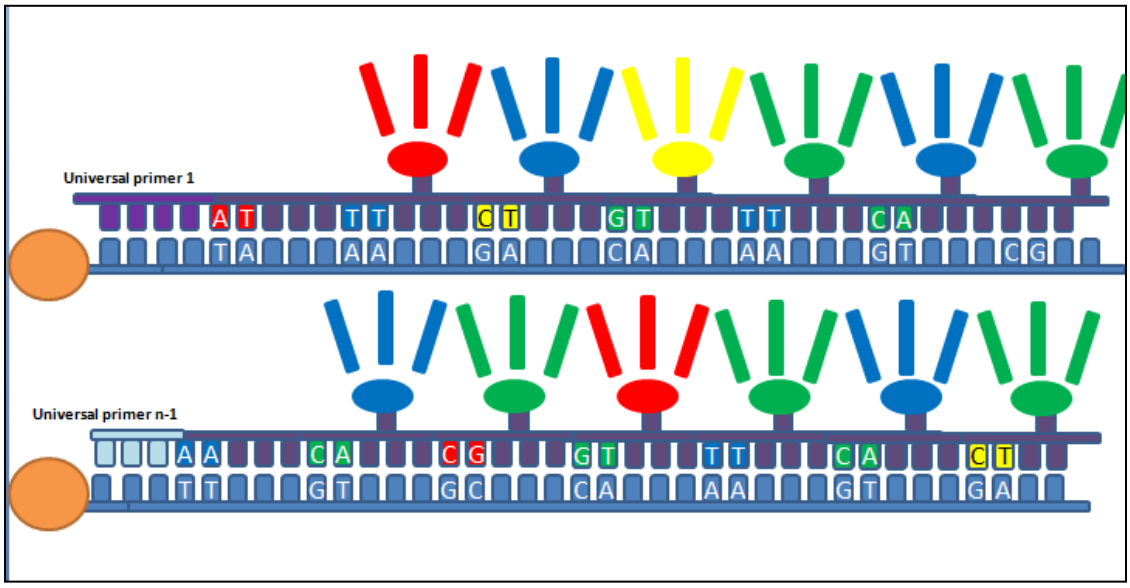


Figure 1.26.4 SOLiD sequencing; First the universal primer 1 is hybridised to the clonal bead template (top). Then one of the 16 di-nucleotide dye-labelled probes hybridises to the template, ligation occurs and an image is taken. This is followed by cleavage of the dye and another round of hybridization, ligation, imaging and cleavage. There are gaps in the sequencing, so all probes are removed and a second universal primer is applied which is offset by one base (bottom). Again the ligation, image, cleave sequencing procedure is repeated and eventually there are 26 consecutive colours representing a 50bp sequence.

Figure 1.26.5 SOLiD colour space decoding

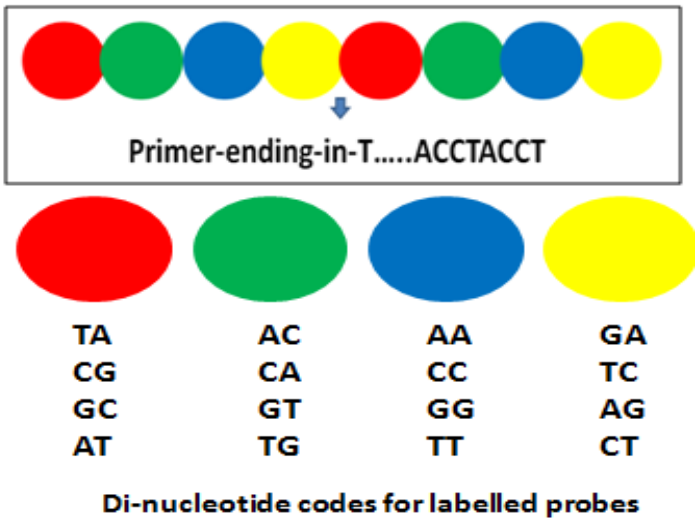


Figure 1.26.5; The image defines the 16 possible di-nucleotide codes. You start to decode the colour space sequence by using the last T base from the universal primer as the first di-nucleotide base. A T base followed by a red colour is an A base. Then an A base followed by a green colour is a C base and so on. A sequence is built up and this is all done with bioinformatics, with a read aligner of choice (possibly bowtie [102]).

1.27 Transcriptome RNA-seq or deep sequencing

Gene expression, although it is not always the case, measures the mRNA level that is loosely proportional to protein level. Therefore certain technologies have been developed that aim to measure gene expression at the mRNA level on a genome scale. These techniques are: cDNA libraries (low coverage), SAGE libraries, microarrays and more recently, 2nd generation RNA-seq sequencing. In addition, there are also shotgun proteomic approaches that can reveal differentially expressed genes in tissues.

RNA-seq is the whole transcriptome sequencing, performed by 2nd generation sequencers (also known as deep sequencing). mRNA is fragmented into 250bp fragments and is subject to amplification and massive parallel sequencing using Illumina or SOLiD sequencers. RNA-seq enables researchers to attain a gene expression snapshot and the transcripts present within in a tissue or disease state of interest. One advantage that RNA-

seq has over microarrays is that it does not need any previous knowledge of sequences and as such it can provide researchers with novel transcripts and new isoforms of known genes. In addition, the dynamic range of gene expression surpasses that of microarray and there are no possible cross-hybridising probes. This technique enables you to measure gene expression, find new transcripts, discover non-coding RNAs, unearth micro-RNAs and find pseudo genes. As an illustration, a recent Illumina RNA-seq data set contained 80 million x 100bp paired end reads (~500x transcriptome coverage). Paired end means 100bp of each end of a 250bp fragment is sequenced. Read alignments and paired end read information can aid transcript assembly programs like Cufflinks to assemble transcripts isoforms, by understanding the distance between paired ends and using intron spanning reads. This information is applied in statistical models to calculate the most likely isoforms and transcript assemblies (see Tophat and Cufflinks [103, 104]).

In this work, Illumina, SOLiD and Roche 454 were all used for different samples of endothelial cells, isolated from different environments. The only 2nd generation sequencing method used was RNA-seq, which is a method for finding differentially expressed genes, isoforms and identifying the presence of novel sequences.

1.28 Need for new targets is the driving force behind work

Figure 1.28.1 displays a Metro newspaper article published December 2010, which gives one good example of the need and motivation to find new vascular targets. To provide possibly cheaper, alternative strategies by using one or more of the new targets found in this work. It is the aim of the author to supply the scientific community with as many potential markers of established tumour vasculature and/or angiogenic targets as possible, which could, in the long run, end up with patient benefit and treatment of solid tumour cancers.

Figure 1.28.1 Newspaper article relating to the expense of some current treatments

Wednesday, December 8, 2010 METRO 13

Cancer drug rejected for use on NHS

By **Fred Attewill**

THOUSANDS of breast cancer victims will be denied access to the so-called 'miracle' drug Avastin.

The £33,000 drug used in many countries to treat different types of advanced cancer has been rejected for use on the NHS.

That comes despite claims it can help thousands of women live at least five months longer.

But Avastin (or bevacizumab) offers only 'limited and uncertain benefit' for breast cancer patients, according to the The National Institute for Health and Clinical Excellence.

Its manufacturer, Roche, said it had presented Nice with evidence showing 'clinically meaningful benefits, particularly in certain breast cancer patients with poor outcomes'.

Cancer sufferers living in France, Germany, Italy, Scandinavia and Canada are all prescribed it for free.

Nice initially banned the drug last year for use in bowel cancer, prompting an outcry from patient groups and campaigners. Since then, Roche has come up with cheaper ways of prescribing it – but the rationing body still believes it is too expensive.

Data on women pre-treated with chemotherapy and given Avastin with the taxane drug showed an average increase in survival of 5.6 months, Roche said.

But it is expensive, the average cost per patient being £33,000.

Avastin works by stopping tumours from forming their own blood supply, which restricts growth and spread.

Nice appraised Avastin in combination with a type of chemotherapy for treating breast cancer which has spread around the body.

Sir Andrew Dillon, chief executive of Nice, said: 'Although the data seemed to show that the drug may slow the growth and spread of the cancer, the size of this effect varied between studies.'

Figure 1.28.1; There is an apparent need for new and cheaper vascular targets and/or angiogenic disrupting agents.

1.29 Chapter 2, Overview; characterisation and identification of vascular targets in an embryonic system.

For the first time, *in-vivo* biotinylation of a chicken embryo is combined with both high-resolution proteomics and bioinformatics to characterise an embryonic organism and to find potential vascular targets. This additionally looks at differences between tumour and wound models, also used in the chicken embryonic system.

1.30 Chapter 3 combining 1st and 2nd generation sequenced cDNA libraries to identify vascular targets

For the first time, public cDNA libraries are combined with long read 2nd generation sequencing of Roche 454 Titanium. The long read lengths of 454 provide some similarity to cDNA libraries but with a vast superior transcriptome coverage. Public and in-house data are combined to provide a list of putative vascular targets.

1.31 Chapter 4; using RNA-seq deep sequencing to find putative pan-vascular targets

In this chapter, the latest 2nd generation sequencing technologies are utilised to illuminate the tumour endothelial transcriptome of freshly isolated tumour endothelial and normal endothelial cells from multiple organs. One overriding aim is to find vascular targets that are present in multiple organ types. A second analyses using tumour conditioned HUVEC and liver endothelial cells, cultured for 24 hours in 1% oxygen, are combined and compared against multiple normal adult human RNA-seq libraries from the public domain (RPKM analyses).

1.32 Chapter 5; identifying vascular targets in lung cancer using microarray data from multiple patient tumour endothelium

Here, 4 biological replicate tumour and normal fresh isolated endothelium were subject to microarray analyses. This was to take into account biological variation and to see if there were perpetual vascular targets in lung tumours.

CHAPTER 2; CHARACTERISATION AND IDENTIFICATION OF VASCULAR TARGETS IN AN EMBRYONIC SYSTEM

2.1 Introduction

A novel *in-vivo* biotinylation procedure has been developed by the Bikfalvi group to facilitate the elucidation of vascular extracellular and membrane proteins of the chicken embryo. This is the first time *in-vivo* biotinylation has been performed in an embryo. It also extends and improves on a previous method presented by Rybak et al. 2005 [53] to capture a more complete proteome. *In-vivo* biotinylation involves a functionally reactive analog of biotin, which is injected into the blood stream of the chicken chorioallantoic membrane (CAM), where it perfuses blood vessels and surrounding tissues. Here the biotin chemical reacts and covalently bonds to the N-terminal amine groups of vessel and stroma genes that are in contact with blood and the vasculature. These are then recovered by homogenisation and elution with hot water but not detergent. To make the technique more sensitive than in previous work, which involved digesting streptavidin bound biotin-coupled-proteins with trypsin (Rybak et al. [53]), this work removed proteins with hot water prior to trypsin digestion to produce a more complete proteome. This procedure was applied to chicken embryonic liver, kidney, intestine, CAM and granulation (wound incision on the CAM) tissues as well as glioma cells implanted onto a CAM. Mass spectrometry proteomic data from each condition was combined with bioinformatic analyses to characterise the data and to identify likely vascular and matrix associated genes that could be used as vascular targets. The high resolution proteomics also enabled differential expression between the different tissues. Two pools of proteomic data were generated: 1) CIKL, which stands for CAM, Intestine, Kidney and Liver pool and 2) CTW that represents the CAM, Tumour and Wound pool. To characterize matrix and vascular proteins of the chicken embryo within these samples, some of which potentially could be suitable vascular targets, the following bioinformatics analyses were carried out:

- biological replicate reproducibility and principle component analyses

- human orthologs identification of chicken proteins
- distribution and differential expression between the different tissues
- identification of gene ontology and molecular pathways
- peptide matching to identify glioma specific genes
- the generation of an "Angioscore", derived from literature keyword scanning
- endothelial cell expression signatures of orthologs
- potential vascular targets

Please note, appendix tables are given in a separate volume, also downloadable from http://sara.molbiol.ox.ac.uk/userweb/jherbert/tissue_diffex/appendix_files.html.

2.2 Principle component analyses

Three biological replicates were generated for all conditions and chicken proteins were identified from mass-spectrums using the Mascot proprietary software [105, 106]. Briefly, the Mascot mass spectrometry analyses were performed using label free quantification to produce Protein Abundance Index (PAI) values, which were computed using mass-spectrum peaks from the 3 most abundant peptides, from the same Uniprot protein.

Principle component analyses were then performed on the PAI data to see if biological replicates of like samples clustered. Prior data transformations were carried out as follows: any missing PAI values were imputed, replicate protein PAI values were averaged and all data was quantile normalized (see methods). The results of the principle component analyses of the normalized data can be seen in figures 2.2.1 to 2.2.3. The results show that, with the possible exception of one kidney replicate, alike samples cluster close together and organ specific samples show more variation along the first principle component than the CAM, tumour and wound samples (Fig. 2.2.1). To attain higher resolution of each individual pool, principle component analyses were carried out independently (CIKL and CTW). As can be seen in figures. 2.2.2 and 2.2.3 there is greater variation between sample groups of both pools. Pearson correlation coefficient

(PCC) values were generated to assess the reproducibility of biological replicates and figure 2.2.4 shows the full list of values, which were all greater than 0.83.

Figure 2.2.1 Principle Component Analyses of normalized protein counts for CIKL and CTW

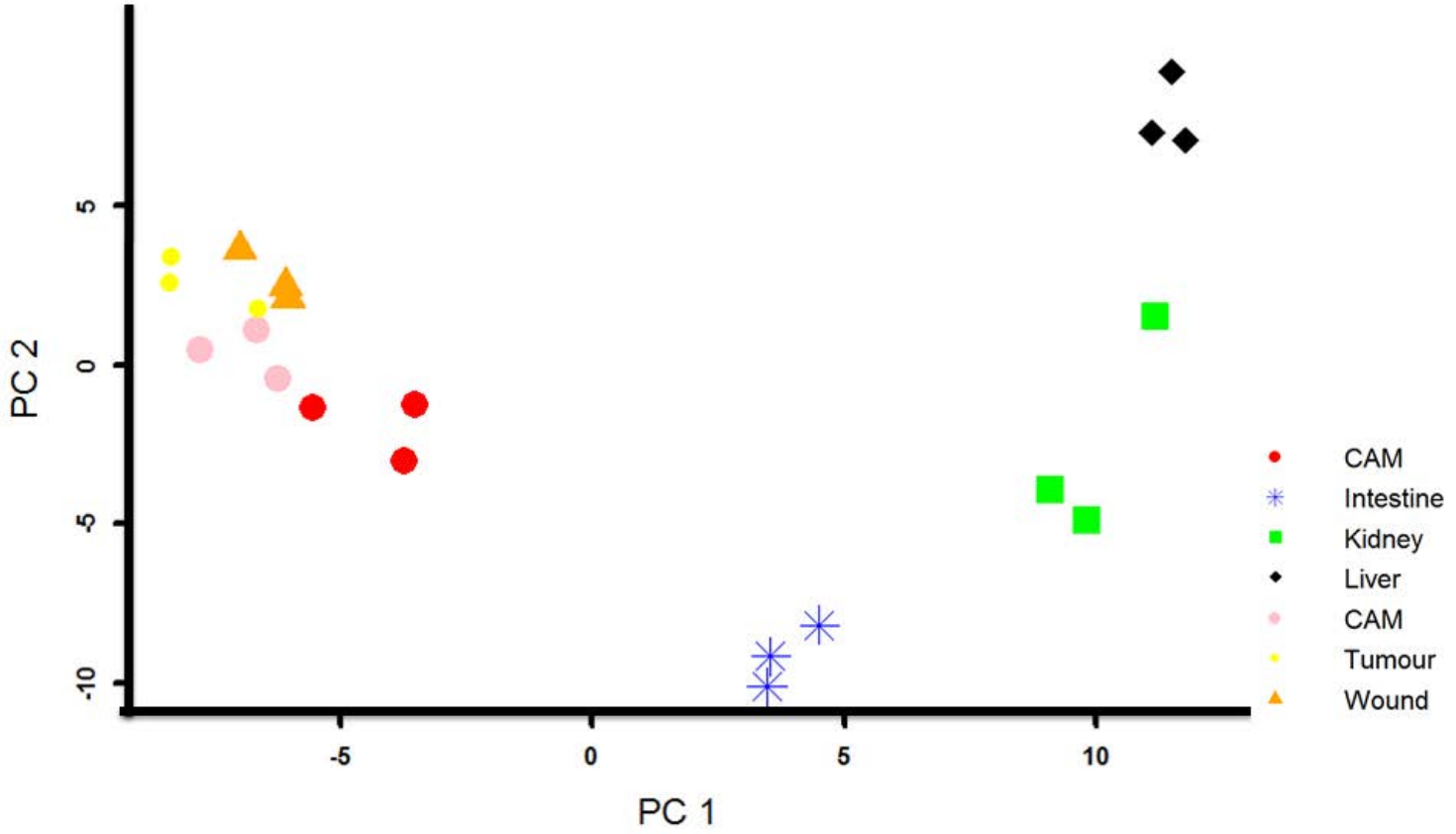


Figure 2.2.1 shows the plot of a principle component analyses of both protein pools combined. This plot shows the two directions of most variance in the data and it can be seen that the same tissue replicates cluster together. The liver, kidney and intestine samples show more variation along the first principle component than the CAM, tumour and wound samples.

Figure 2.2.2 Principle Component Analyses of the CIKL pool

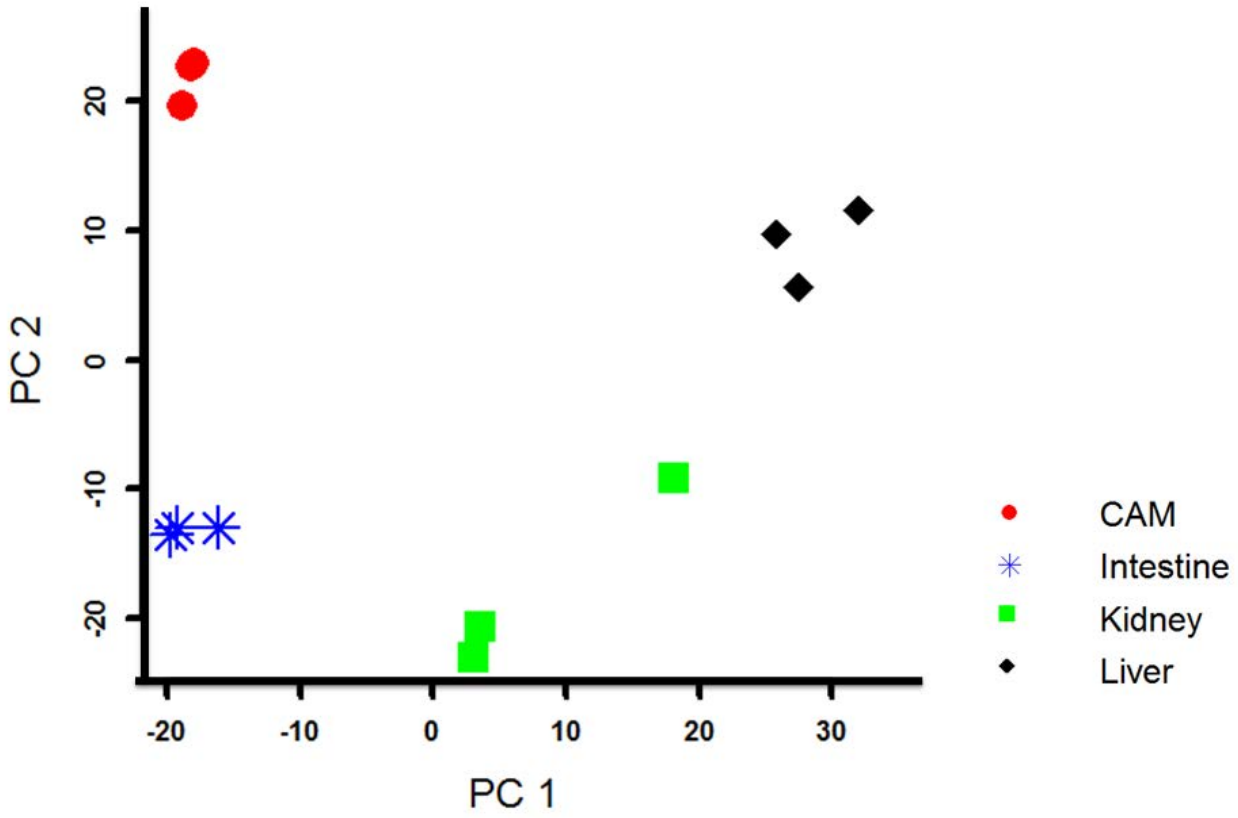


Figure 2.2.2; this figure portrays an independent principle component plot of CIKL pool. The higher resolution reveals a kidney replicate as being different to the other kidney samples.

Figure 2.2.3 Principle Component Analyses of the CTW pool

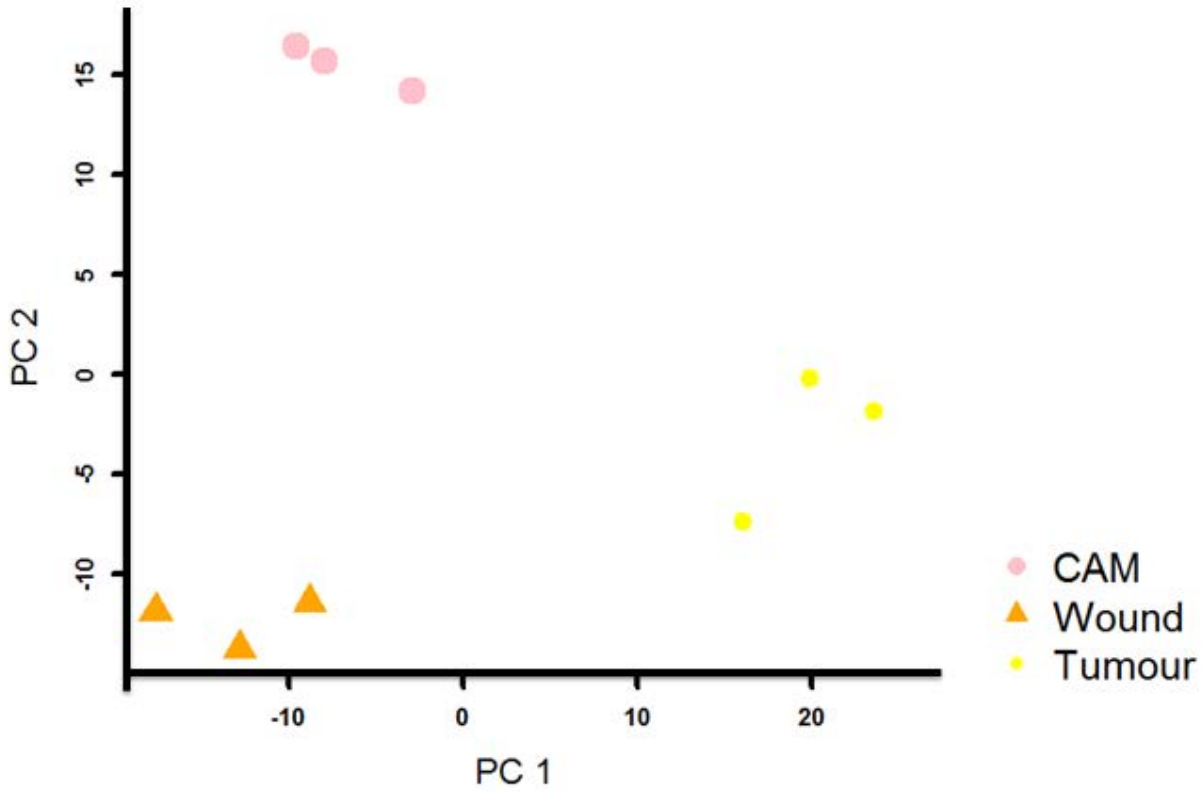


Figure 2.2.3; shows an independent principle component analyses of the CAM, wound and tumour samples. The largest variation in the first principle component shows that the wound and CAM samples show less global gene expression difference than the tumour sample.

Figure 2.2.4 Pearson Correlation Coefficient values of biological replicates

	Cam1	Cam2	Cam3	Tumour1	Tumour2	Tumour3	Wound1	Wound2	Wound3
Cam1	1	0.94	0.94	0.86	0.85	0.83	0.84	0.83	0.84
Cam2	0.94	1	0.96	0.85	0.87	0.84	0.84	0.85	0.84
Cam3	0.94	0.96	1	0.87	0.89	0.86	0.85	0.85	0.85
Tumour1	0.86	0.85	0.87	1	0.95	0.91	0.84	0.80	0.81
Tumour2	0.85	0.87	0.89	0.95	1	0.91	0.84	0.82	0.81
Tumour3	0.83	0.84	0.86	0.91	0.91	1	0.81	0.79	0.83
Wound1	0.84	0.84	0.85	0.84	0.84	0.81	1	0.96	0.94
Wound2	0.83	0.85	0.85	0.80	0.82	0.79	0.96	1	0.93
Wound3	0.84	0.84	0.85	0.81	0.81	0.83	0.94	0.93	1

	Cam1b	Cam2b	Cam3b	Intest1	Intest2	Intest3	Liver1	Liver2	Liver3	Kidney1	Kidney2	Kidney3
Cam1b	1	0.96	0.95	0.70	0.70	0.71	0.52	0.56	0.49	0.56	0.55	0.58
Cam2b	0.96	1	0.97	0.73	0.73	0.74	0.52	0.56	0.48	0.57	0.58	0.60
Cam3b	0.95	0.97	1	0.70	0.70	0.72	0.50	0.54	0.47	0.54	0.54	0.57
Intest1	0.70	0.73	0.70	1	0.95	0.95	0.52	0.51	0.47	0.66	0.69	0.70
Intest2	0.70	0.73	0.70	0.95	1	0.96	0.48	0.48	0.43	0.61	0.67	0.68
Intest3	0.71	0.74	0.72	0.95	0.96	1	0.49	0.49	0.44	0.63	0.68	0.70
Liver1	0.52	0.52	0.50	0.52	0.48	0.49	1	0.92	0.91	0.82	0.67	0.69
Liver2	0.56	0.56	0.54	0.51	0.48	0.49	0.92	1	0.91	0.80	0.66	0.69
Liver3	0.49	0.48	0.47	0.47	0.43	0.44	0.91	0.91	1	0.78	0.61	0.64
Kidney1	0.56	0.57	0.54	0.66	0.61	0.63	0.82	0.80	0.78	1	0.84	0.86
Kidney2	0.55	0.58	0.54	0.69	0.67	0.68	0.67	0.66	0.61	0.84	1	0.93
Kidney3	0.58	0.60	0.57	0.70	0.68	0.70	0.69	0.69	0.64	0.86	0.93	1

Figure 2.2.4; this figure shows a heat-map of the Pearson Correlation Coefficient results of all biological replicates from both pools. For replicates from the same tissue type, a minimum correlation of 0.84 was seen.

2.3 Human ortholog identification

Although chicken is a model organism and there exists a draft genome, the transcriptome of the chicken, unlike human, is not comprehensively annotated. To facilitate literature keyword, gene ontology and endothelial expression analyses, the human ortholog of each chicken protein was sought. The entire set of proteins from both pools totalled 1370 (non-redundant) from the Uniprot database. The orthologs of each protein was found employing a BLAST sequence analyses pipeline similar to Herbert et al. [107]. Appendix table A.2.3.1 shows the full protein and ortholog results for both CIKL and CTW pools. Table 2.3.1 displays the counts for proteins and orthologs; of the total proteins found, 18 proteins were not assigned an ortholog since: 1) 11 proteins were chicken specific (no human orthologs) and 2) 7 proteins did not hit a human gene with a significant enough BLAST hit (1e-9). 1352 proteins were assigned an ortholog and the non-redundant set of human orthologs was 1157.

Table 2.3.1 lists the number of proteins assigned an ortholog

	Number of Chicken proteins	Number of orthologs found
PAI quantification data	1370	1157

Table 2.3.1; This table depicts the number of non-redundant proteins and orthologs found in the PAI data for both the CTW and the CIKL pools combined.

2.4 Differential expression of proteins

Normal distribution; the linear model moderated T-test assumes a normal distribution of data. To test the proteomic PAI data for Gaussian/normal distribution, a histogram for each pool was produced (Fig. 2.4.1). The results show a bell like shape for both pools, characteristic of a normal distribution, which enables valid parametric test statistics to be used on this data.

Figure 2.4.1 PAI value distribution histograms of both proteome pools (CIKL and CTW)

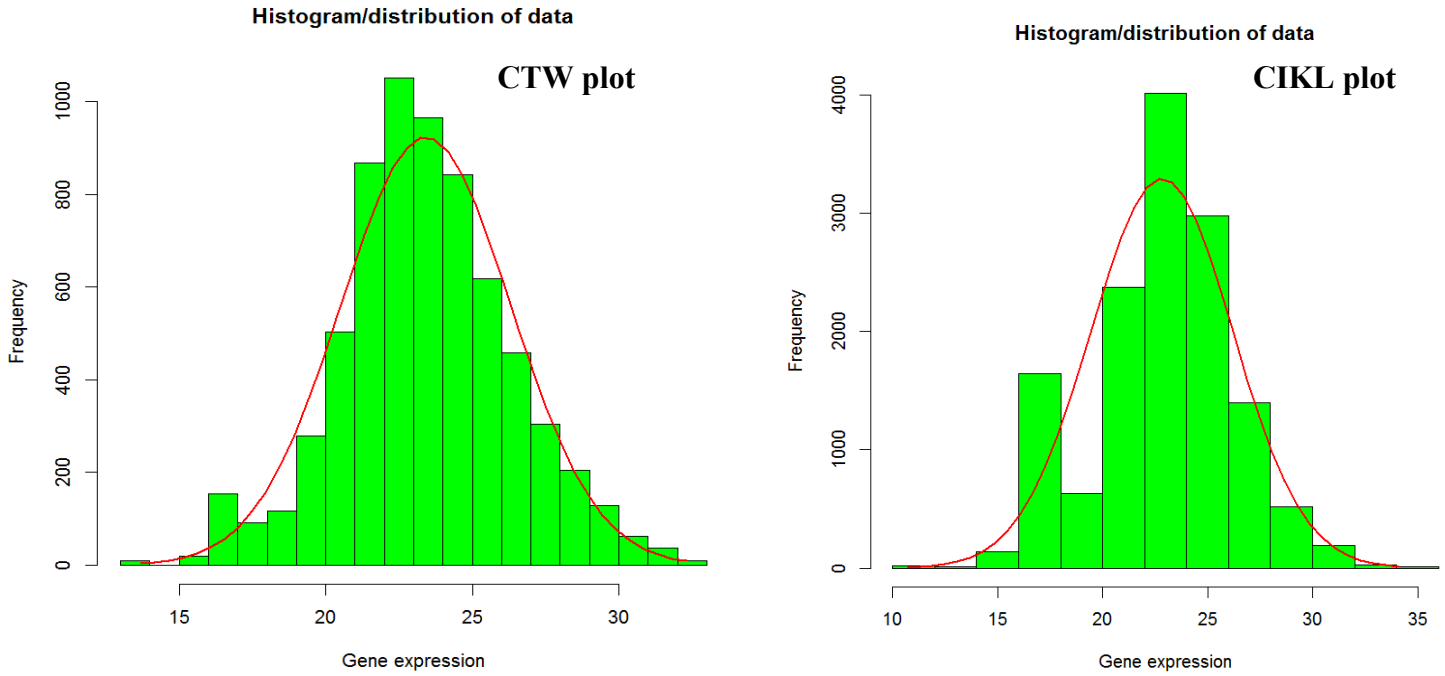


Figure 2.4.1. The histogram plots show that both proteomic PAI pool values show an approximate normal distribution, enabling statistically valid differential gene expression with a parametric moderated T-test statistic.

2.5 Differentially expressed genes and tissue enriched genes

This approach used the PAI data in a moderated T-test differential gene expression analyses, called "decide tests" from the Bioconductor Limma package [108]. This tests for differential expression between multiple tissues and the results of the most specific genes for each tissue from the CIKL pool are listed in table 2.5.1. The column of this table labelled "Significant enrichment" details a tissue for which a gene was enriched. There are three numerical codes for the decide tests; 1 stands for significantly up-regulated, -1 for significantly down regulated and 0 for no differential expression. Statistical significance was reached for a contrast with the following conditions met; a logarithmic to base 2 fold change of ≥ 1 and an adjusted p-value ≤ 0.01. The number of tissue enriched genes for CIKL: liver = 16, kidney = 10, intestine 20 and CAM = 33. Table 2.5.2 lists the genes enriched in the CTW pool and the number of enriched genes were for the CAM = 3, the tumour = 6 and the wound = 10. To confirm tissue specific expression, 2-dimensional clustering was performed using normalized expression PAI values of these genes. Biological replicate samples clustered together as well as differentially expressed genes as can be seen in the cluster heatmaps presented in figures 2.5.1 and 2.5.2. This provides confirmation of the tissue enriched expression.

A download file of tissue clustered genes with AngioScore and cDNA library endothelial enrichment can be downloaded from;
http://sara.molbiol.ox.ac.uk/userweb/jherbert/tissue_diffex/appendix_files.html (file called "Tissue enriched protein clusters with AngioScore and Endothelial cDNA enrichment").

Table 2.5.1 Linear Model decide tests applied to the CIKL pool PAI data set

Gene	intestine - kidney	intestine - liver	Liver - kidney	intestine - cam	kidney - cam	liver - cam	Significant enrichment
SEPT2	-1	0	-1	0	1	0	Kidney
ANO4	-1	0	-1	0	1	0	Kidney
CNNM2	-1	0	-1	0	1	0	Kidney
CRB2	-1	0	-1	0	1	0	Kidney
DBH	-1	0	-1	0	1	0	Kidney
FAAH	-1	0	-1	0	1	0	Kidney
FLT4	-1	0	-1	0	1	0	Kidney
MGST3	-1	0	-1	0	1	0	Kidney
RIMS2	-1	0	-1	0	1	0	Kidney
SLC16A7	-1	0	-1	0	1	0	Kidney
BRP44L	0	-1	1	0	0	1	Liver
CPZ	0	-1	1	0	0	1	Liver
DYNLL1	0	-1	1	0	0	1	Liver
FABP2	0	-1	1	0	0	1	Liver
GOT1	0	-1	1	0	0	1	Liver
GRHPR	0	-1	1	0	0	1	Liver
GSTA1	0	-1	1	0	0	1	Liver
GSTZ1	0	-1	1	0	0	1	Liver
HAPLN1	0	-1	1	0	0	1	Liver
HPD	0	-1	1	0	0	1	Liver
HPGDS	0	-1	1	0	0	1	Liver
LAMC3	0	-1	1	0	0	1	Liver
LDHB	0	-1	1	0	0	1	Liver
NQO1	0	-1	1	0	0	1	Liver
PYGL	0	-1	1	0	0	1	Liver

SEC23A	**0**	**-1**	**1**	**0**	**0**	**1**	**Liver**
ADAMTSL1	**1**	**1**	**0**	**1**	**0**	**0**	**Intestine**
C9orf30-TMEFF1	**1**	**1**	**0**	**1**	**0**	**0**	**Intestine**
CAV1	**1**	**1**	**0**	**1**	**0**	**0**	**Intestine**
CNTNAP2	**1**	**1**	**0**	**1**	**0**	**0**	**Intestine**
EML4	**1**	**1**	**0**	**1**	**0**	**0**	**Intestine**
EPHA1	**1**	**1**	**0**	**1**	**0**	**0**	**Intestine**
FNDC1	**1**	**1**	**0**	**1**	**0**	**0**	**Intestine**
GPA33	**1**	**1**	**0**	**1**	**0**	**0**	**Intestine**
GPLD1	**1**	**1**	**0**	**1**	**0**	**0**	**Intestine**
GRM7	**1**	**1**	**0**	**1**	**0**	**0**	**Intestine**
KDELR2	**1**	**1**	**0**	**1**	**0**	**0**	**Intestine**
KIAA1731	**1**	**1**	**0**	**1**	**0**	**0**	**Intestine**
LIPH	**1**	**1**	**0**	**1**	**0**	**0**	**Intestine**
LRIG1	**1**	**1**	**0**	**1**	**0**	**0**	**Intestine**
MDGA1	**1**	**1**	**0**	**1**	**0**	**0**	**Intestine**
MEGF6	**1**	**1**	**0**	**1**	**0**	**0**	**Intestine**
MYLK	**1**	**1**	**0**	**1**	**0**	**0**	**Intestine**
NTNG1	**1**	**1**	**0**	**1**	**0**	**0**	**Intestine**
TNR	**1**	**1**	**0**	**1**	**0**	**0**	**Intestine**
SYK	**1**	**1**	**0**	**1**	**0**	**0**	**Intestine**
ACAT2	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
ACTA1	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
ACTA2	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
ACTB	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
ALDOC	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
ANXA1	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
BPNT1	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
CNN2	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
DSTN	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**

E1BYN5	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
E1BYN6	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
EPCAM	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
FAT2	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
GSTT1	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
JUP	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
KRT17	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
LCP1	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
LYG2	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
MYL6	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
NEU2	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
NUCB2	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
P02659	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
P02845	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
PAFAH1B2	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
PLS3	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
PPIC	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
PSCA	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
S100A11	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
SERPINB3	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
TMPRSS6	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
TNFRSF6B	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
VWA2	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**
ZP3	**0**	**0**	**0**	**-1**	**-1**	**-1**	**Cam**

Table 2.5.1. This tables shows the Linear Model decide tests differential gene expression contrasts between the liver, kidney, intestine and CAM tissues. The following number of genes were significantly enriched in each tissue with a logarithmic to base 2 fold change of ≥ 1 and an adjusted p-value ≤ 0.01. Tissue enriched genes: kidney = 10, liver = 16, intestine 20 and CAM = 33. A value of 1 is up-regulated, -1 down regulated and 0 is no change.

Table 2.5.2 Linear Model decide tests applied to the CTW pool PAI data set

Gene	tumour - cam	tumour - wound	wound - cam	Significant enrichment
CNTFR	**-1**	**0**	**-1**	**Cam**
F13A1	**-1**	**0**	**-1**	**Cam**
SGCD	**-1**	**0**	**-1**	**Cam**
C1R	**1**	**1**	**0**	**Tumour**
C1S	**1**	**1**	**0**	**Tumour**
CTHRC1	**1**	**1**	**0**	**Tumour**
GFPT1	**1**	**1**	**0**	**Tumour**
OAS3	**1**	**1**	**0**	**Tumour**
VIM	**1**	**1**	**0**	**Tumour**
ANK1	**0**	**-1**	**1**	**Wound**
BPGM	**0**	**-1**	**1**	**Wound**
CD276	**0**	**-1**	**1**	**Wound**
EPB41	**0**	**-1**	**1**	**Wound**
EPB42	**0**	**-1**	**1**	**Wound**
FECH	**0**	**-1**	**1**	**Wound**
PON2	**0**	**-1**	**1**	**Wound**
PTPRB	**0**	**-1**	**1**	**Wound**
SLC4A1	**0**	**-1**	**1**	**Wound**
TSPO	**0**	**-1**	**1**	**Wound**

Table 2.5.2. This tables shows the Linear Models differential gene expression contrasts between tumour, wound and CAM samples. The following number of genes were significantly enriched in each tissue with a logarithmic to base 2 fold change of ≥ 1 and an adjusted p-value ≤ 0.01. Tissue enriched genes were the CAM = 3, tumour = 6 and wound = 10. A value of 1 is up-regulated, -1 down-regulated and 0 is no change.

Figure 2.5.1 Two-dimensional clustering of differentially expressed genes of CIKL samples

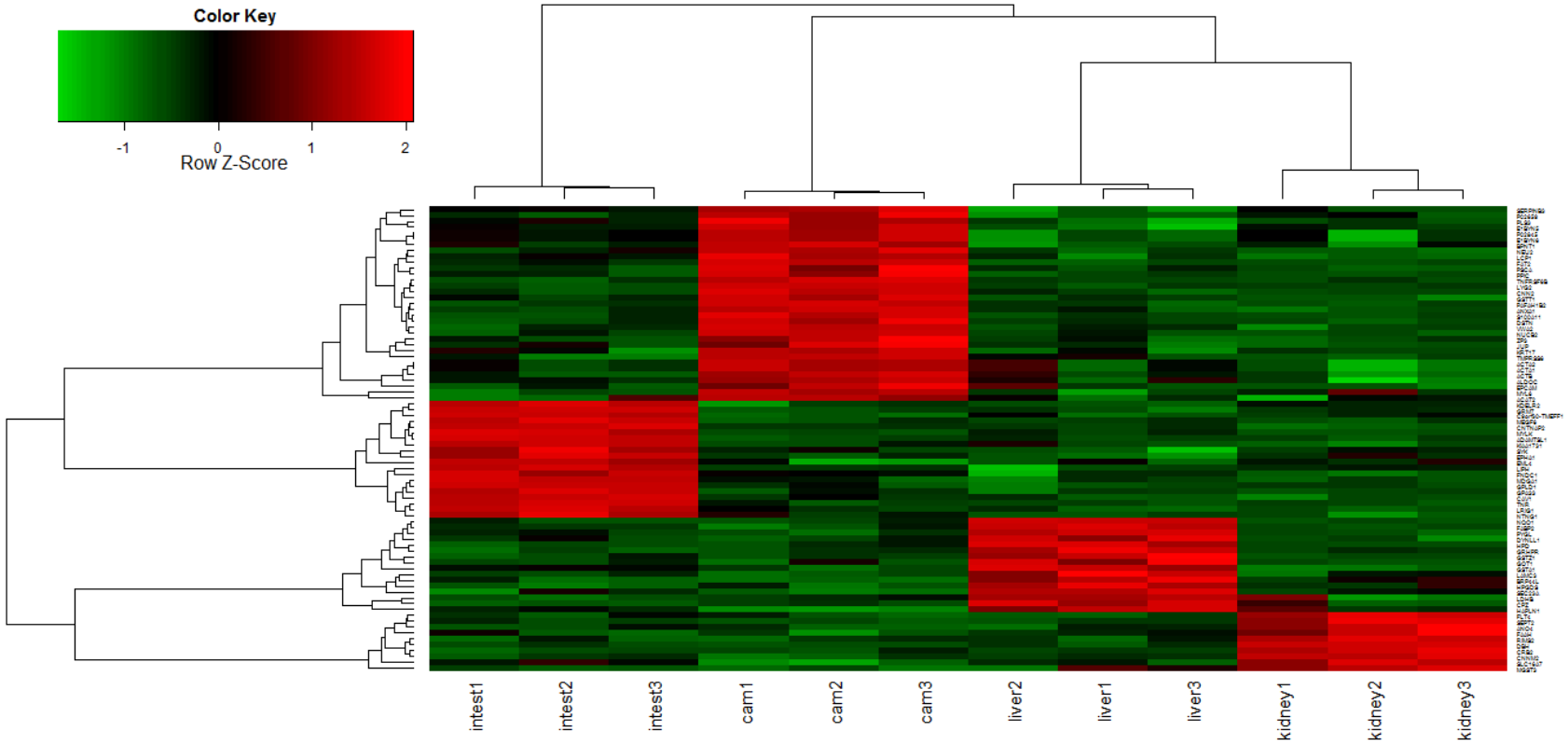


Figure 2.5.1. Hierarchical 2-dimensional clustering of differentially expressed genes found in the CIKL pool by Limma decide tests. Cluster distance calculated by 1-correlation. There is clear separation of replicate tissues by the differentially expressed genes.

Figure 2.5.2 Two-dimensional clustering of differentially expressed genes of CTW samples

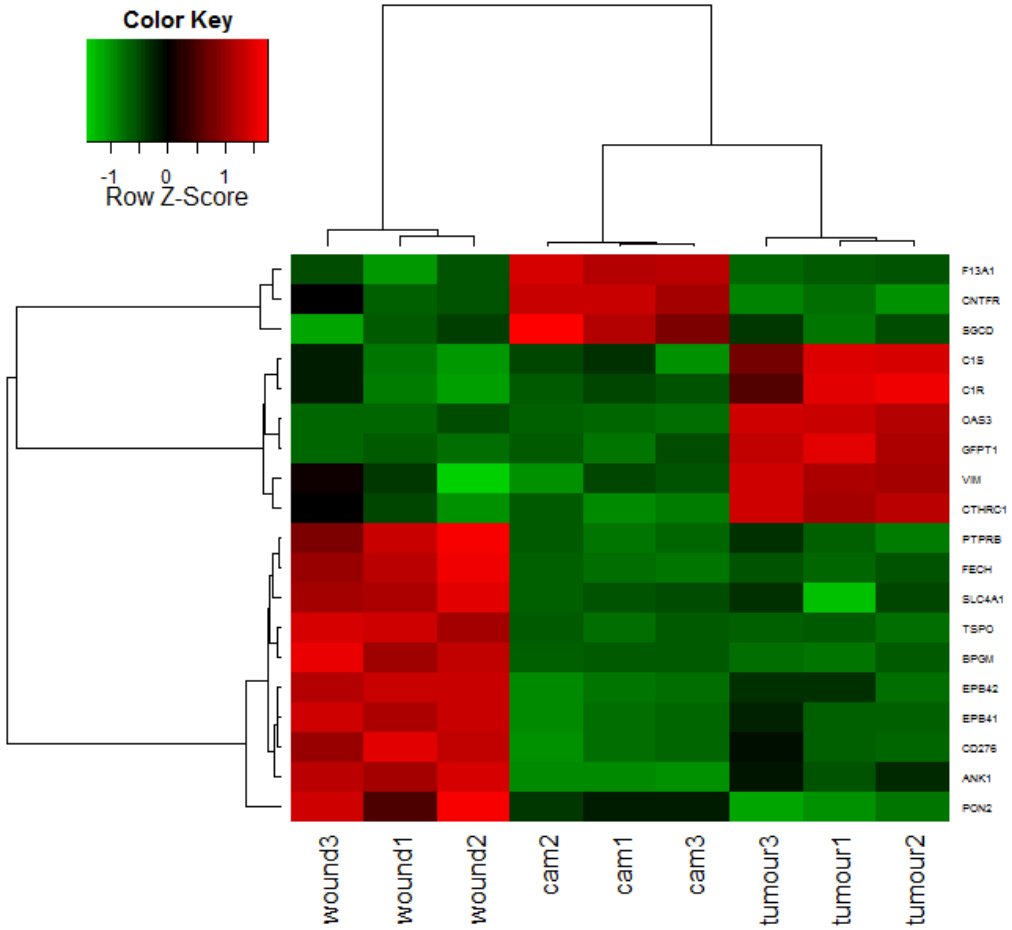


Figure 2.5.2. Hierarchical 2-dimensional clustering of differentially expressed genes from the CTW pool found by Limma decide tests. Cluster distance calculated by 1-correlation.

2.6 Tumour versus wound differential expression

Wounds and tumours both undergo angiogenesis but one is a homeostatic process and the other pathological. To determine the molecular differences between the wound and tumour implanted samples, an additional moderated T-test was performed between tumour and wound samples alone. A list of the most significant differentially expressed genes, with positive B-value > 0 and adjusted p-value ≤ 0.02, are listed in table 2.6.1. A volcano plot displaying the most significant differentially expressed genes can be seen in figure 2.6.1, which are those genes highlighted in blue.

Table 2.6.1 Tumour vs. wound differentially expressed genes

ID	Product	logFC	AveExpr	adj.P.Val	B
NAMPT	nicotinamide phosphoribosyltransferase	6.13	19.08	0.00	6.52
C1R	complement component 1, r subcomponent	5.71	18.76	0.00	4.86
OAS3	2'-5'-oligoadenylate synthetase 3, 100kDa	5.68	18.91	0.00	2.13
RAC2	ras-related C3 botulinum toxin substrate 2 (rho family, small GTP binding protein Rac2)	5.37	18.12	0.00	3.56
POMGNT1	protein O-linked mannose beta1,2-N-acetylglucosaminyltransferase	5.24	18.71	0.00	6.28
P27177		5.15	18.70	0.00	5.93
PSMA7	proteasome (prosome, macropain) subunit, alpha type, 7	4.98	18.65	0.00	4.46
PDIA3	protein disulfide isomerase family A, member 3	4.60	25.53	0.00	6.19
TAGLN	Transgelin	3.78	18.60	0.00	2.89
SPTB	spectrin, beta, erythrocytic	3.64	25.02	0.00	2.94
GFPT1	glutamine--fructose-6-phosphate transaminase 1	3.43	18.36	0.00	3.18
TNFAIP6	tumor necrosis factor, alpha-induced	2.61	21.31	0.00	1.68

	protein 6				
RAB27A	RAB27A, member RAS oncogene family	2.52	18.22	0.01	0.42
SERPINA10	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 10	2.30	18.22	0.01	0.51
S100P	S100 calcium binding protein P	2.06	26.50	0.01	0.00
ITGB5	integrin, beta 5	1.90	23.93	0.01	0.24
CALCRL	calcitonin receptor-like	-6.37	18.91	0.00	3.26
H1F0	H1 histone family, member 0	-5.77	19.51	0.00	2.83
FECH	Ferrochelatase	-5.49	19.30	0.00	3.74
EPB42	erythrocyte membrane protein band 4.2	-5.33	20.18	0.00	6.58
AHNAK2	AHNAK nucleoprotein 2	-4.84	18.43	0.00	1.77
EPB41	erythrocyte membrane protein band 4.1 (elliptocytosis 1, RH-linked)	-4.81	19.85	0.00	4.63
ANK1	ankyrin 1, erythrocytic	-4.69	21.16	0.00	5.65
BPGM	2,3-bisphosphoglycerate mutase	-4.57	19.16	0.00	4.17
CD276	CD276 molecule	-4.51	18.51	0.00	3.97
SLC4A1	solute carrier family 4, anion exchanger, member 1 (erythrocyte membrane protein band 3, Diego blood group)	-4.22	23.95	0.00	1.72
PTPRB	protein tyrosine phosphatase, receptor type, B	-4.03	19.40	0.00	1.77
TSPO	translocator protein (18kDa)	-3.91	18.78	0.00	1.92
ABCE1	ATP-binding cassette, sub-family E (OABP), member 1	-3.78	19.34	0.00	2.40
BLVRB	biliverdin reductase B (flavin reductase (NADPH))	-3.77	22.88	0.01	0.86
PON2	paraoxonase 2	-3.46	18.59	0.01	0.36
PABPC4	poly(A) binding protein, cytoplasmic 4 (inducible form)	-3.43	18.70	0.01	0.27
CPT1A	carnitine palmitoyltransferase 1A (liver)	-3.16	18.32	0.01	0.24
FZD2	frizzled family receptor 2	-2.78	19.21	0.01	1.03

MRC1	macrophage mannose receptor 1-like	-2.61	19.33	0.00	1.51
NMT1	N-myristoyltransferase 1	-2.12	19.75	0.01	0.31

Table 2.6.1. Differential protein expression analyses were performed on normalized tumour versus wound sample data and the results are presented in this table. 16 human ortholog genes were found to be significantly up-regulated in tumour samples (portrayed in red as tumour up-regulated) and 20 ortholog genes significantly up in the wound samples (displayed in blue as tumour down-regulated).

Figure 2.6.1 Volcano plot of tumour versus wound differential gene expression

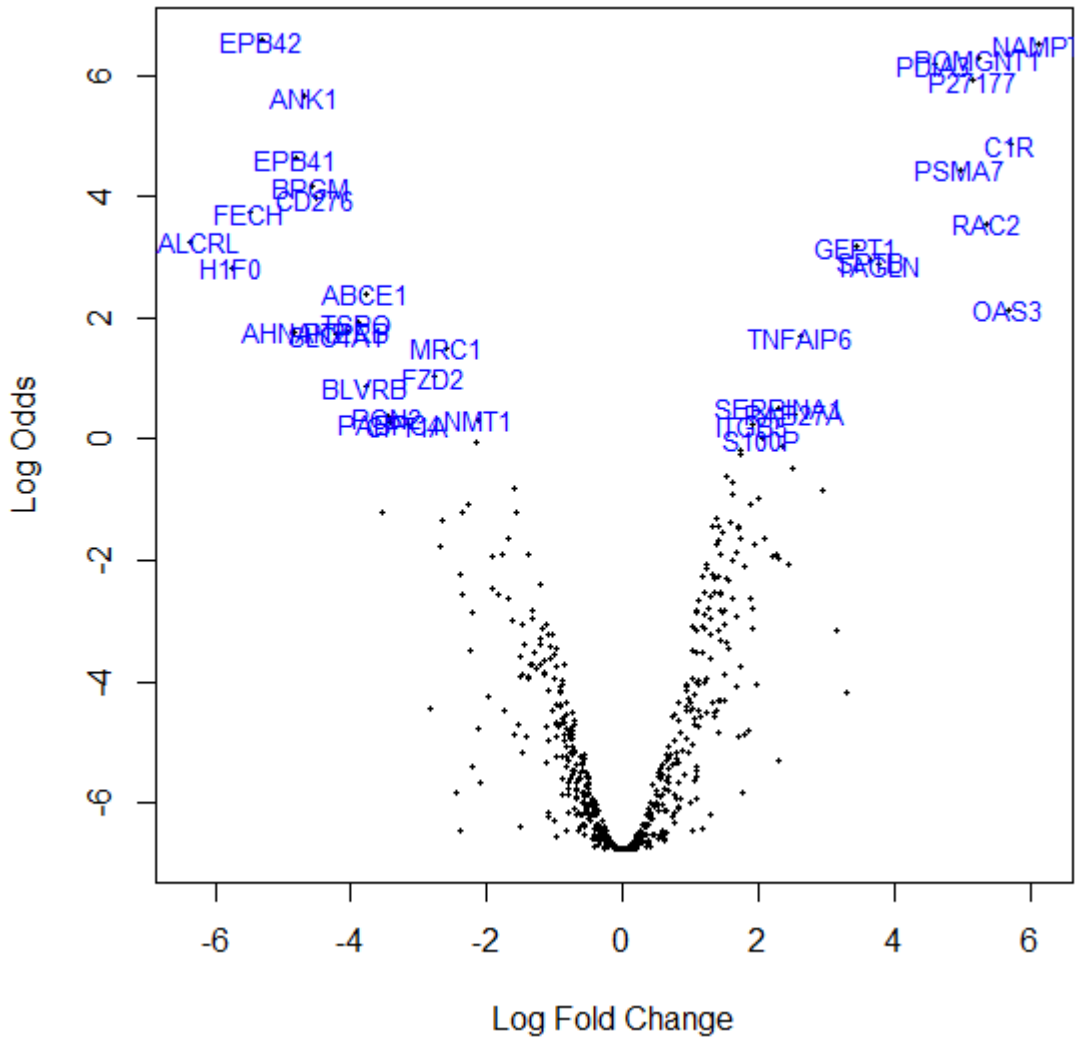


Figure 2.6.1. Differentially expressed genes in the tumour vs. wound samples were sought using Linear Models. 20 genes and 16 genes were found to be up-regulated with a statistical significant positive B-value > 0 (adjusted p-value ≤ 0.02) in the wound and tumour samples respectively (those genes are highlighted in blue on the volcano plot).

2.7 Tumour versus wound keyword searches

To find if genes have been reported in the literature as being involved in wound healing and or glioma tumours, an abstract keyword search was performed using "glioma", "glioblastoma", "wound", "healing" and "granulation" as keyword baits. The results of this can be seen in table 2.7.1. Four genes enriched in tumours (ITGB5, RAC2, C1R and NAMPT) have abstracts containing glioma related keywords and two other tumour enriched genes (RAB27A and S100P), contain wound related keywords. For wound enriched genes, the results were that CALCRL, FECH, EPB41 and PABPC4 had wound related keywords, with PTPRB and TSPO having glioma related keywords.

Table 2.7.1 Tumour and wound enriched genes, abstract keyword search results

Gene	Total abstracts	Matching keywords	glioma	glioblastoma	wound	healing	granulation
NAMPT	213	1		18728403			
C1R	42	1	8227070				
RAC2	78	3	18217210 19423540		10758162		
RAB27A	81	4					18571497 21693760 15548590 19704116
S100P	56	1			18575778		
ITGB5	79	4	11934885 21978494		20404485 11287419		
CALCRL	69	1					12538603
FECH	82	1			19703464		
EPB41	96	1			21750196		
PTPRB	36	2		11313993 12615970			
TSPO	64	2	15769477	18791274			
PABPC4	37	1			9030741		

Table 2.7.1. Genes found to be differentially expressed between tumour and wound were subject to keyword searches for words relevant to glioma and wound healing. 12 out of the 36 genes had articles related to the biology being investigated, which are listed in this table (those not matching a keyword were omitted). Some genes enriched in tumours had articles relating to both glioma and/or wound healing and the vice versa was seen with wound enriched genes. Pubmed identification numbers are reported for an abstract matching a keyword in this table. Tumour enriched genes are in red text and wound enriched are in blue.

2.8 Gene ontology and pathways

Gene ontology and molecular pathways give an insight to the biology and processes occurring in given lists of genes. Gene ontology (GO) biological process for the entire set of orthologs from all tissues was carried out using the online DAVID software. Figure 2.8.1 shows the percent proteins and the most enriched GO biological process categories. To class each ortholog gene as either intracellular, extracellular or membrane, textual gene cellular component for each gene was studied by eye and classified. Figure 2.8.2 shows the results of cellular component analyses where 51% of orthologs are either extracellular or plasma membrane genes. This agrees with the Rybak study who performed a similar approach [53].

Figure 2.8.1 Gene ontology biological process, all orthologs

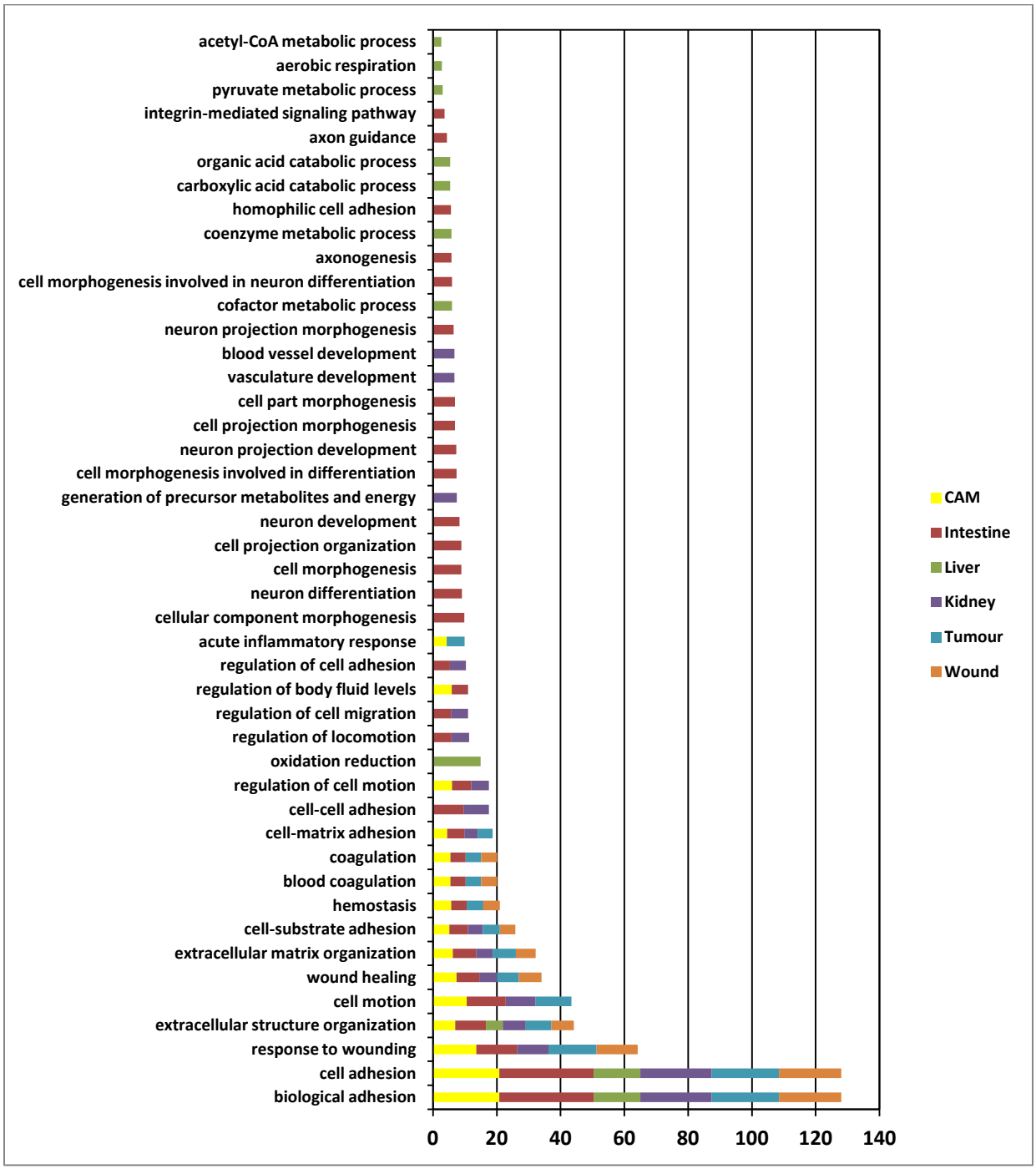


Figure 2.8.1. Global overview of Gene Ontology biological process for all proteins from all tissues. All tissues have evidence of embryonic development and matrix development as shown by the presence of response to wounding, extracellular-matrix organization and cell/biological adhesion ontology categories.

Figure 2.8.2 Gene ontology cellular component

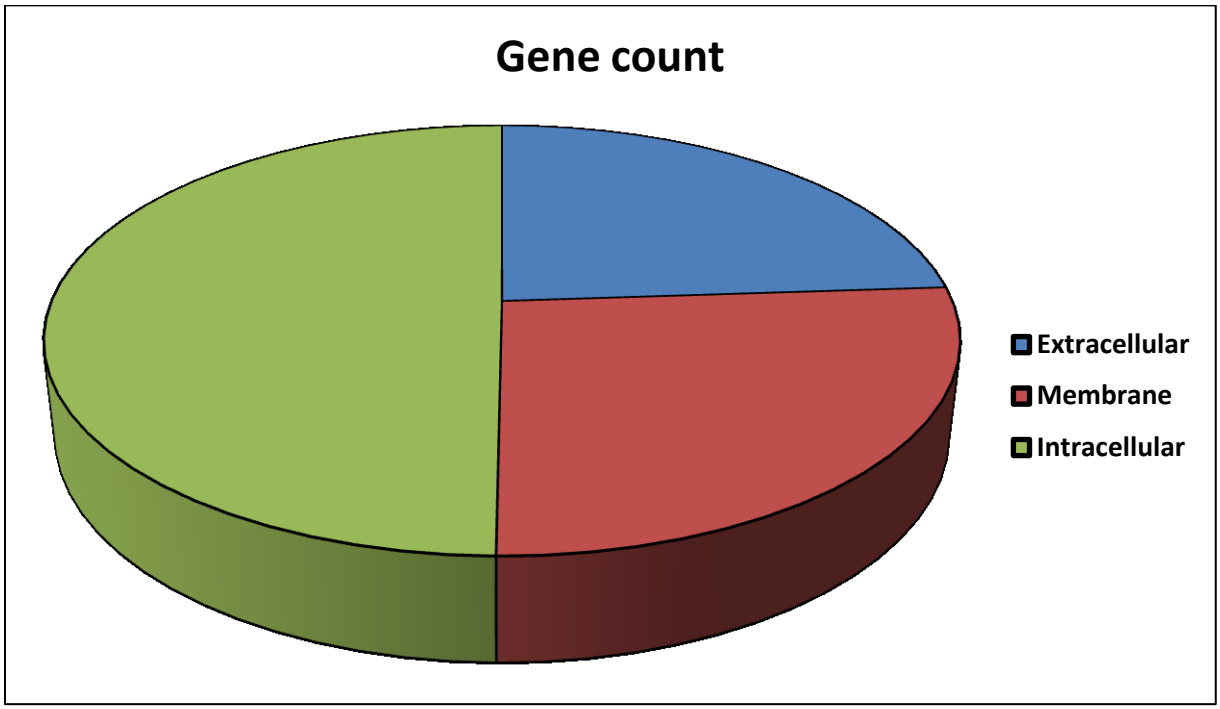


Figure 2.8.2. Global overview of Gene Ontology cellular component for orthologs proteins classed as either intracellular, extracellular or membrane proteins. It is evident that 51% of all proteins are either extracellular or plasma membrane proteins. This agrees with the Rybak study, who performed a similar approach [53].

2.9 Molecular pathways;

Figures 2.9.1 and 2.9.2 shows the pathways enriched in all of the genes present for KEGG and Panther as found by a DAVID enrichment analyses.

Figure 2.9.1 KEGG pathway enrichment from DAVID

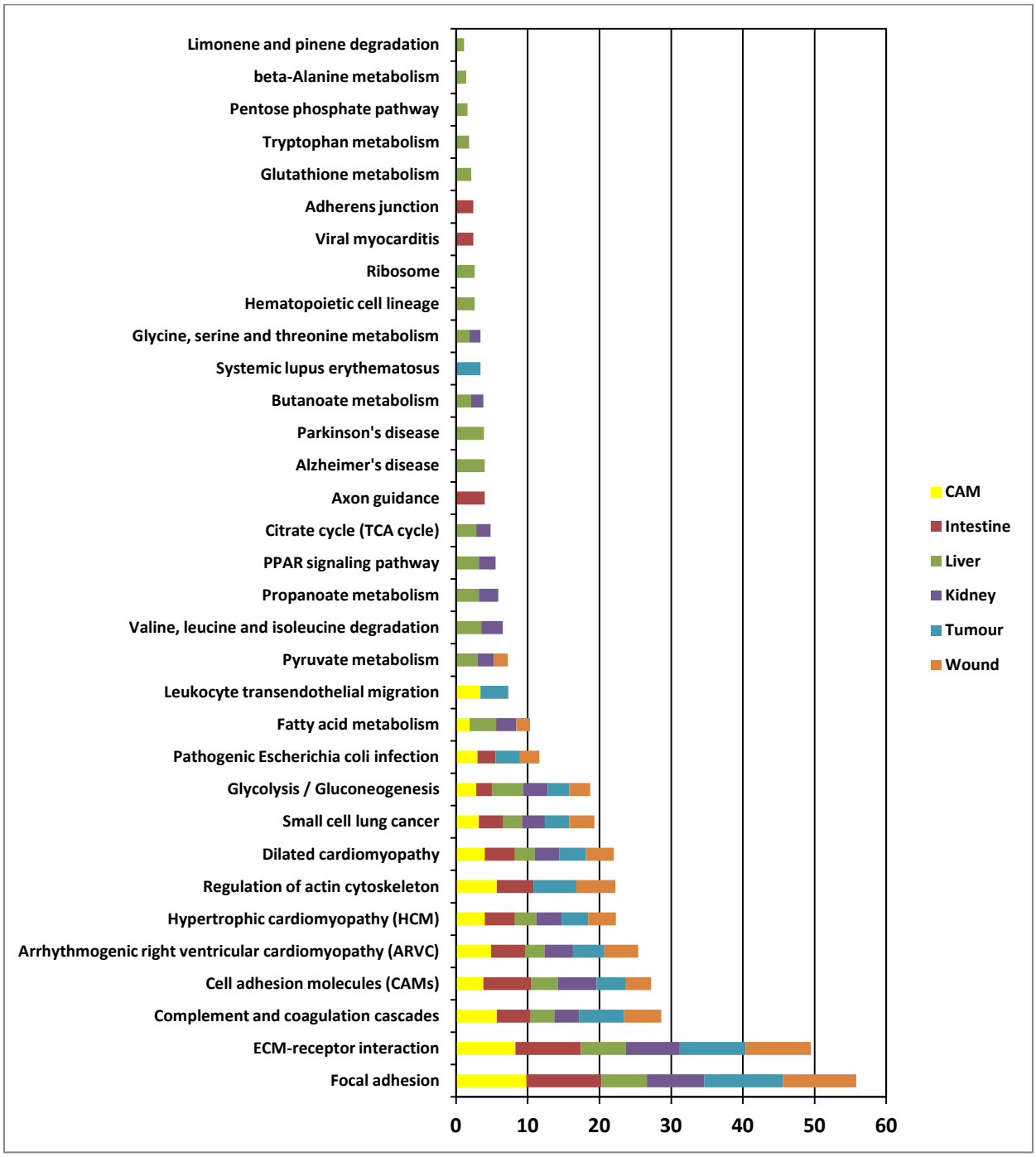


Figure 2.9.1. Global overview of KEGG pathways for all genes found by DAVID.

Figure 2.9.2 Panther pathway enrichment from DAVID

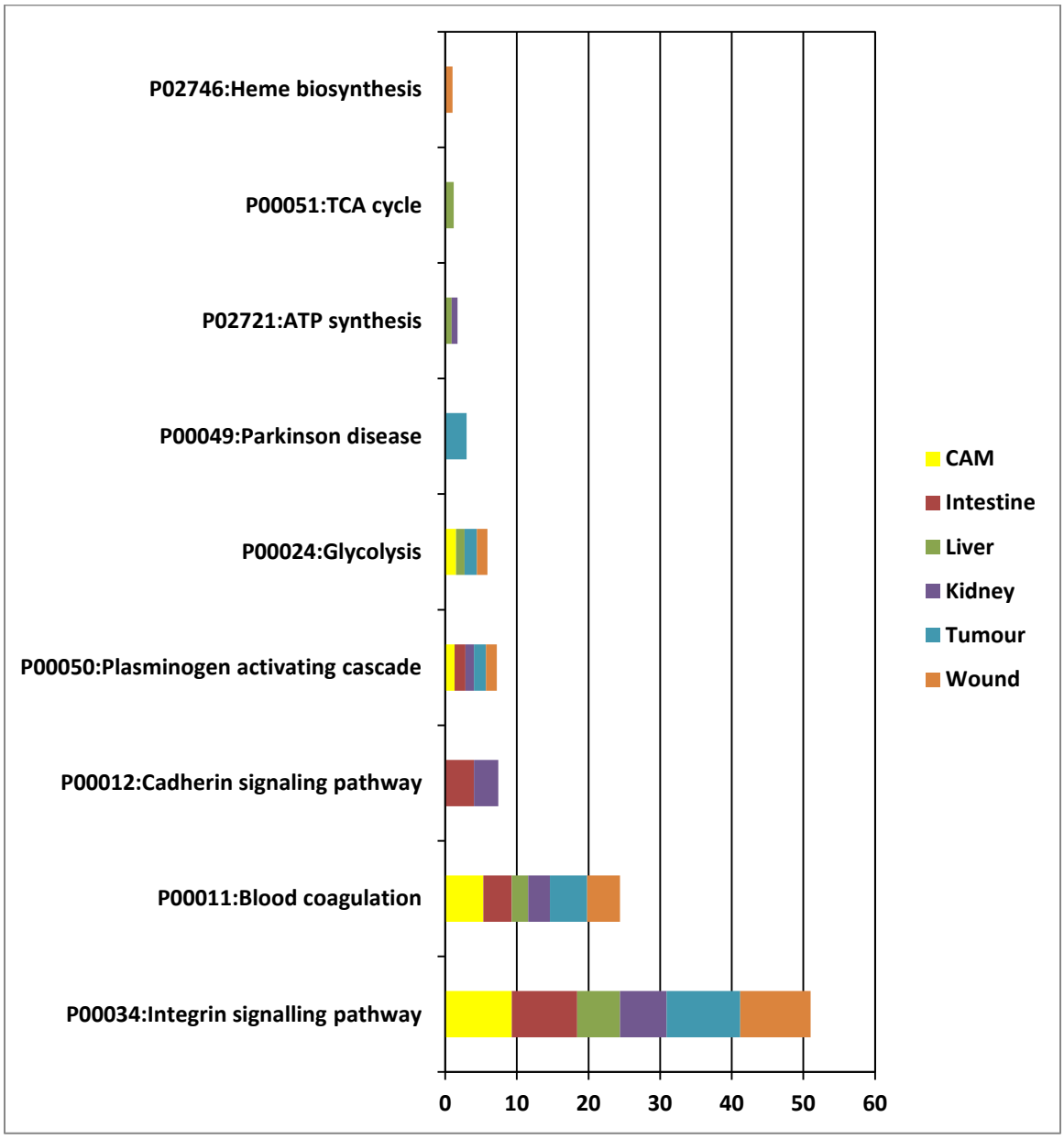


Figure 2.9.2. Global overview of Panther pathways for all genes found by DAVID.

2.10 Identifying human glioma genes from the tumour sample

To find human glioma expressed genes from the tumour sample, the peptides specific to the U87 glioma implanted tumour cells were identified using a program. 2,470 unique peptides from the total tumour sample were matched exactly to the Swissport/Trembl chicken and human database of proteins, which contained 118,018 and 27,333 human and chicken proteins respectively.

The number of peptides matching either or both species was as follows;

- Peptides matching only human = 176 (7.1%)
- Peptides matching only chicken = 1942 (78.6%)
- Peptides that matched both species = 352 (14.3%)

86% of tumour sample peptides were classified specific to either human or chicken genes. 48 genes had peptides specific to human (glioma cells) with 8 genes matching ≥ 7 different human specific peptides. These are listed in table 2.10.1 and 10 genes had promiscuous but human specific peptides (data not shown).

Table 2.10.1 Human specific peptides identify glioma expressed genes

Gene	Number of different peptides matched	Product
KRT1	24	keratin 1
KRT10	18	keratin 10
KRT9	16	keratin 9
KRT2	12	keratin 2
PCCA	9	propionyl CoA carboxylase, alpha polypeptide
GAPDH	7	glyceraldehyde-3-phosphate dehydrogenase
PKM2	7	pyruvate kinase, muscle
VIM	7	Vimentin
ALB	6	Albumin
PC	5	pyruvate carboxylase
ENO1	4	enolase 1, (alpha)
ALDOA	3	aldolase A, fructose-bisphosphate
ANXA5	3	annexin A5
FN1	3	fibronectin 1
KRT14	3	keratin 14
KRT5	3	keratin 5
LDHA	3	lactate dehydrogenase A
MCCC1	3	methylcrotonoyl-CoA carboxylase 1 (alpha)
TGFBI	3	transforming growth factor, beta-induced, 68kDa
TNC	3	tenascin C
AKR1B1	2	aldo-keto reductase family 1, member B1 (aldose reductase)
ANXA1	2	annexin A1
ANXA2	2	annexin A2
COL1A1	2	collagen, type I, alpha 1
HRNR	2	hornerin

HSPD1	2	heat shock 60kDa protein 1 (chaperonin)
ITGA5	2	integrin, alpha 5 (fibronectin receptor, alpha polypeptide)
TPI1	2	triosephosphate isomerase 1
AEBP1	1	AE binding protein 1
ALDOC	1	aldolase C, fructose-bisphosphate
COL4A1	1	collagen, type IV, alpha 1
EEF2	1	eukaryotic translation elongation factor 2
ENO2	1	enolase 2 (gamma, neuronal)
GPI	1	glucose-6-phosphate isomerase
HSP90AB1	1	heat shock protein 90kDa alpha (cytosolic), class B member 1
HSPA5	1	heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa)
LDHAL6B	1	lactate dehydrogenase A-like 6B
LMNA	1	lamin A/C
LOXL1	1	lysyl oxidase-like 1
MYH14	1	myosin, heavy chain 14, non-muscle
MYH9	1	myosin, heavy chain 9, non-muscle
PDIA4	1	protein disulfide isomerase family A, member 4
PDIA6	1	protein disulfide isomerase family A, member 6
PGK1	1	phosphoglycerate kinase 1
PHEX	1	phosphate regulating endopeptidase homolog, X-linked
PRDX4	1	peroxiredoxin 4
SBSN	1	suprabasin
TUBB8	1	tubulin, beta 8

Table 2.10.1. This lists glioma genes identified by peptide matching. 48 genes had peptides specific to human (glioma cells) with 8 genes matching ≥ 7 different human specific peptides. 10 genes had promiscuous but human specific peptides (data not shown).

2.11 Literature searches of glioma involvement of human specific peptide genes

Keyword searches using "glioma" and "glioblastoma" revealed 17 of the 48 genes had published evidence of glioma involvement and these are listed with Pubmed id numbers, in table 2.11.1.

Table 2.11.1 17 genes with glioma published evidence

Gene	Total abstracts	Abstracts matching keywords	Pubmed id of glioma or glioblastoma matching publications
TNC	149	11	18794852 20622113 7694284 11920587 18757408 16292494 11313993 16388320 8548761 19459858 19147558
VIM	238	9	3027087 8381971 12169273 16944923 11082283 18046501 21317457 12084347 18947333
FN1	407	5	20819642 10783396 21980357 10470109 1409674
ANXA2	221	5	20018898 21033036 21976520 10777578 7518469
HSPA5	285	4	10085239 21112319 18708359 18403493
ITGA5	227	4	16569642 20514406 18992284 18464290
COL4A1	122	2	10382266 9506531
COL1A1	459	1	1409674
HSPD1	230	1	20978188
GAPDH	188	1	17597534
TGFBI	159	1	20419098
ANXA1	146	1	19767728
ENO1	108	1	19795461
KRT1	72	1	16944923
KRT10	63	1	16944923
ALDOC	24	1	2209624

Table 2.11.2. This table lists the 17 glioma related genes with publication evidence.

2.12 List of candidate vascular targets from proteomic data

With the novel proteomic approach of targeting membrane and extracellular matrix proteins in an embryo, there is an opportunity to identify potentially novel vascular targets. An AngioScore was first generated that gives a rough estimate on whether known angiogenic related genes were in fact being selected with the proteomic technique. Here, angiogenesis and tumour angiogenesis keywords were searched against total publications for a gene to derive an "AngioScore" (the percentage of publications matching a word to the total number of publications). Table 2.12.1 shows the 43 genes that have an AngioScore ≥ 50 and appendix table A.2.12.1 lists all AngioScores for all orthologs.

Table 2.12.1 Genes with the highest AngioScores

Human Gene	Human Protein	Human Product	GO CC class	Angio Score
KDR	NP_002244	kinase insert domain receptor (a type III receptor tyrosine kinase)	M	94
CDH5	NP_001786	cadherin 5, type 2 (vascular endothelium)	M	90
FLT4	NP_891555	fms-related tyrosine kinase 4	M	86
EDNRA	NP_001948	endothelin receptor type A	M	81
PROCR	NP_006395	protein C receptor, endothelial	M	80
TEK	NP_000450	TEK tyrosine kinase, endothelial	M	77
ECE1	NP_001106819	endothelin converting enzyme 1	M	74
CA9	NP_001207	carbonic anhydrase IX	I	74
wAOC3	NP_003725	amine oxidase, copper containing 3 (vascular adhesion protein 1)	M	70
COL18A1	NP_569712	collagen, type XVIII, alpha 1	E	70
VCAM1	NP_001186763	vascular cell adhesion molecule 1	M	70
PECAM1	NP_000433	platelet/endothelial cell adhesion molecule	M	70
NRP2	NP_003863	neuropilin 2	M	69
SERPINB5	NP_002630	serpin peptidase inhibitor, clade B (ovalbumin), member 5	E	68
TUSC5	NP_758955	tumor suppressor candidate 5	M	67
NRP1	NP_001231902	neuropilin 1	M	66
ANGPT2	NP_001138	angiopoietin 2	E	66
MCAM	NP_006491	melanoma cell adhesion molecule	M	65
CADM1	NP_055148	cell adhesion molecule 1	M	65
IGFBP7	NP_001544	insulin-like growth factor binding protein 7	E	63
ADAM17	NP_003174	ADAM metallopeptidase domain 17	M	63
ANG	NP_001091046	angiogenin, ribonuclease, RNase A family, 5	E	62

TIE1	NP_005415	tyrosine kinase with immunoglobulin-like and EGF-like domains 1	M	62
CDCP1	NP_073753	CUB domain containing protein 1	M	59
ENG	NP_000109	endoglin	M	59
MFI2	NP_005920	antigen p97 (melanoma associated) identified by monoclonal antibodies 133.2 and 96.5	E	58
PLXND1	NP_055918	plexin D1	M	58
LRIG1	NP_056356	leucine-rich repeats and immunoglobulin-like domains 1	M	57
JAM3	NP_116190	junctional adhesion molecule 3	M	57
RECK	NP_066934	reversion-inducing-cysteine-rich protein with kazal motifs	M	56
ANGPTL1	NP_004664	angiopoietin-like 1	E	55
DMBT1	NP_060049	deleted in malignant brain tumors 1	E	55
FGFBP1	NP_005121	fibroblast growth factor binding protein 1	E	55
THBS1	NP_003237	thrombospondin 1	E	55
NME1-NME2	NP_001018146	NME1-NME2 readthrough	I	54
STAB1	NP_055951	stabilin 1	M	53
LYVE1	NP_006682	lymphatic vessel endothelial hyaluronan receptor 1	M	53
SERPINF1	NP_002606	serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1	E	52
PROM1	NP_006008	prominin 1	M	51
ENPP2	NP_001035181	ectonucleotide pyrophosphatase/phosphodiesterase 2	M	51
COLEC12	NP_569057	collectin sub-family member 12	M	50
EDIL3	NP_005702	EGF-like repeats and discoidin I-like domains 3	E	50
KLF2	NP_057354	Kruppel-like factor 2 (lung)	I	50

Table 2.12.1. This table shows the 43 genes that have an AngioScore ≥ 50 and appendix table A.2.12.1 lists all AngioScores for all orthologs.

2.13 cDNA library endothelial enrichment

cDNA libraries have been used in a differential gene expression experiment to identify endothelial and tumour endothelial enriched genes [62]. In this work, we combined the earlier endothelial screen with a Roche 454 RNA-seq hypoxia liver endothelium library to increase the potential of identifying new vascular targets within the proteomic data.

For the full set of orthologs, analyses were performed that matched gene symbols between orthologs and endothelial enriched genes; those that matched were reported. Genes that were membrane or secreted proteins were given preference as these are directly targetable via the blood stream. The criterion used to rank the genes was 1) a gene was log2(FC) ≥ 1 (2 fold) up-regulated in endothelial cells or 2) genes showed an endothelial specific profile by having zero counts in the non-endothelial pool. 81 membrane and 43 extracellular genes were found to have an endothelial enriched cDNA library profile, as well as 81 intracellular genes. Table 2.13.1 contains the full list of endothelial enriched genes.

Table 2.13.1 Putative vascular targets, 205 genes with an endothelial enriched expression profile

Human Gene	Product	GO CC class	Angio Score	log2 (FC)	Endo Specific
CDH5	cadherin 5, type 2 (vascular endothelium)	M	90	8.45	specific
HSD17B2	hydroxysteroid (17-beta) dehydrogenase 2	M	18	8.2	specific
CALCRL	calcitonin receptor-like	M	25	8	specific
ELTD1	EGF, latrophilin and seven transmembrane domain containing 1	M	10	7.99	specific
CD93	CD93 molecule	M	23	7.87	specific
LOC100652938	macrophage mannose receptor 1-like	M	0	7.72	specific
TEK	TEK tyrosine kinase, endothelial	M	77	7.3	specific
SLC12A2	solute carrier family 12 (sodium/potassium/chloride transporters), member 2	M	8	7.04	
PVRL3	poliovirus receptor-related 3	M	23	5.73	specific
NRCAM	neuronal cell adhesion molecule	M	13	5.73	specific
PECAM1	platelet/endothelial cell adhesion molecule	M	70	5.56	
CUBN	cubilin (intrinsic factor-cobalamin receptor)	M	6	5.44	specific
FAT4	FAT tumor suppressor homolog 4 (Drosophila)	M	20	5.08	specific
LYVE1	lymphatic vessel endothelial hyaluronan receptor 1	M	53	4.94	specific
THBD	thrombomodulin	M	43	4.78	specific
PODXL	podocalyxin-like	M	37	4.69	
PTPRB	protein tyrosine phosphatase, receptor type, B	M	25	4.62	
CDH13	cadherin 13, H-cadherin (heart)	M	42	4.35	
INSR	insulin receptor	M	5	4.3	
CD109	CD109 molecule	M	24	4.16	

PLXND1	plexin D1	M	58	4.15	specific
PLXNA4	plexin A4	M	17	4.15	specific
EPHB1	EPH receptor B1	M	35	4.02	specific
KDR	kinase insert domain receptor (a type III receptor tyrosine kinase)	M	94	4	
PTPRG	protein tyrosine phosphatase, receptor type, G	M	31	3.87	specific
ACP2	acid phosphatase 2, lysosomal	M	4	3.7	specific
ENG	endoglin	M	59	3.63	
CD47	CD47 molecule	M	22	3.52	
TMEM119	transmembrane protein 119	M	0	3.51	specific
STT3A	STT3, subunit of the oligosaccharyltransferase complex, homolog A (S. cerevisiae)	M	0	3.5	
PLXNA2	plexin A2	M	10	3.25	
CXADR	coxsackie virus and adenovirus receptor	M	24	3.08	
FABP2	fatty acid binding protein 2, intestinal	M	13	3.03	specific
CNTNAP1	contactin associated protein 1	M	6	3.03	specific
ADAM17	ADAM metallopeptidase domain 17	M	63	2.95	
ADAM10	ADAM metallopeptidase domain 10	M	30	2.93	
IL31RA	interleukin 31 receptor A	M	11	2.72	specific
IL1RAP	interleukin 1 receptor accessory protein	M	10	2.72	specific
FREM2	FRAS1 related extracellular matrix protein 2	M	14	2.72	specific
NRP1	neuropilin 1	M	66	2.68	
ITGA6	integrin, alpha 6	M	27	2.67	
ENPP2	ectonucleotide pyrophosphatase/phosphodiesterase 2	M	51	2.65	
CD9	CD9 molecule	M	28	2.56	
LRRN4	leucine rich repeat neuronal 4	M	13	2.55	
TMEM33	transmembrane protein 33	M	0	2.49	
FLRT2	fibronectin leucine rich transmembrane protein 2	M	0	2.47	
TIE1	tyrosine kinase with immunoglobulin-like and EGF-like domains 1	M	62	2.4	
PTPRS	protein tyrosine phosphatase, receptor type, S	M	9	2.31	specific
MGST3	microsomal glutathione S-transferase 3	M	5	2.15	
MCAM	melanoma cell adhesion molecule	M	65	1.95	
LPHN2	latrophilin 2	M	14	1.93	
IL6ST	interleukin 6 signal transducer (gp130, oncostatin M receptor)	M	14	1.9	
CD36	CD36 molecule (thrombospondin receptor)	M	23	1.88	
LIFR	leukemia inhibitory factor receptor alpha	M	15	1.87	
SDK1	sidekick homolog 1, cell adhesion molecule (chicken)	M	0	1.75	specific
PLA2R1	phospholipase A2 receptor 1, 180kDa	M	18	1.75	specific

NEGR1	neuronal growth regulator 1	M	6	1.75	specific
NAALAD 2	N-acetylated alpha-linked acidic dipeptidase 2	M	11	1.75	specific
LCAT	lecithin-cholesterol acyltransferase	M	10	1.75	specific
FAM171 A1	family with sequence similarity 171, member A1	M	0	1.75	
COLEC12	collectin sub-family member 12	M	50	1.75	specific
STOM	stomatin	M	0	1.66	
IGF1R	insulin-like growth factor 1 receptor	M	30	1.63	
PTGFRN	prostaglandin F2 receptor negative regulator	M	11	1.62	
UPK1B	uroplakin 1B	M	28	1.55	
ITGAV	integrin, alpha V (vitronectin receptor, alpha polypeptide, antigen CD51)	M	45	1.5	
ESAM	endothelial cell adhesion molecule	M	29	1.5	
ITGA4	integrin, alpha 4 (antigen CD49D, alpha 4 subunit of VLA-4 receptor)	M	34	1.4	
MAGT1	magnesium transporter 1	M	4	1.38	
ATP1B3	ATPase, Na+/K+ transporting, beta 3 polypeptide	M	5	1.26	
ITGA5	integrin, alpha 5 (fibronectin receptor, alpha polypeptide)	M	33	1.24	
PTPRM	protein tyrosine phosphatase, receptor type, M	M	23	1.15	
PCDH1	protocadherin 1	M	7	1.04	
SLC26A1	solute carrier family 26 (sulfate transporter), member 1	M	25	0.82	specific
SDK2	sidekick homolog 2 (chicken)	M	0	0.82	specific
ODZ3	odz, odd Oz/ten-m homolog 3 (Drosophila)	M	10	0.82	specific
MMP15	matrix metallopeptidase 15 (membrane-inserted)	M	19	0.82	specific
ITGA9	integrin, alpha 9	M	35	0.82	specific
FGFR3	fibroblast growth factor receptor 3	M	24	0.82	specific
DCHS1	dachsous 1 (Drosophila)	M	17	0.82	specific
BOC	Boc homolog (mouse)	M	7	0.82	specific
ANGPT2	angiopoietin 2	E	66	10.68	specific
MMRN1	multimerin 1	E	42	9.54	specific
VWF	von Willebrand factor	E	30	8.81	specific
SRPX	sushi-repeat containing protein, X-linked	E	44	6.79	specific
MMRN2	multimerin 2	E	13	6.71	specific
PROS1	protein S (alpha)	E	7	5.21	specific
ENPP6	ectonucleotide pyrophosphatase/phosphodiesterase 6	E	0	4.69	specific
EFEMP1	EGF containing fibulin-like extracellular matrix protein 1	E	17	4.46	
CAPZA1	capping protein (actin filament) muscle Z-line, alpha 1	E	6	4.27	

COL15A1	collagen, type XV, alpha 1	E	30	4.15	specific
A2M	alpha-2-macroglobulin	E	9	3.8	
C8orf84	chromosome 8 open reading frame 84	E	0	3.7	specific
C7	complement component 7	E	15	3.7	specific
GPC5	glypican 5	E	6	3.29	specific
THBS1	thrombospondin 1	E	55	3.24	
GPI	glucose-6-phosphate isomerase	E	28	3.24	
TXN	thioredoxin	E	24	3.2	
SPINK5	serine peptidase inhibitor, Kazal type 5	E	4	3.03	specific
MAMDC2	MAM domain containing 2	E	0	3.03	specific
ASPN	asporin	E	3	3.03	specific
SERPINB 5	serpin peptidase inhibitor, clade B (ovalbumin), member 5	E	68	2.72	specific
PON2	paraoxonase 2	E	17	2.68	
CTNNB1	catenin (cadherin-associated protein), beta 1, 88kDa	E	41	2.63	
SERPINB 9	serpin peptidase inhibitor, clade B (ovalbumin), member 9	E	34	2.56	
EMILIN3	elastin microfibril interfacer 3	E	14	2.31	specific
SRPX2	sushi-repeat containing protein, X-linked 2	E	29	2.1	
COL4A1	collagen, type IV, alpha 1	E	28	2.09	
LAMA4	laminin, alpha 4	E	36	2.04	
NUCB2	nucleobindin 2	E	12	2.01	
PTX3	pentraxin 3, long	E	36	1.98	
LAMB1	laminin, beta 1	E	20	1.98	
NID1	nidogen 1	E	10	1.96	
ADAMTS L5	ADAMTS-like 5	E	0	1.75	specific
ACE	angiotensin I converting enzyme (peptidyl-dipeptidase A) 1	E	21	1.75	specific
TINAGL1	tubulointerstitial nephritis antigen-like 1	E	7	1.71	
CTSZ	cathepsin Z	E	12	1.26	
ALCAM	activated leukocyte cell adhesion molecule	E	46	1.2	
MR1	major histocompatibility complex, class I-related	E	6	1.19	
CFH	complement factor H	E	13	1	
ZP3	zona pellucida glycoprotein 3 (sperm receptor)	E	0	0.82	specific
VWA2	von Willebrand factor A domain containing 2	E	0	0.82	specific
LYG2	lysozyme G-like 2	E	0	0.82	specific
CPXM2	carboxypeptidase X (M14 family), member 2	E	0	0.82	specific
ATP5H	ATP synthase, H+ transporting, mitochondrial Fo complex, subunit d	I	0	8.31	specific
SGCE	sarcoglycan, epsilon	I	3	6.04	specific
DPYSL2	dihydropyrimidinase-like 2	I	3	5.42	

TLK1	tousled-like kinase 1	I	0	5.11	
ABCA1	ATP-binding cassette, sub-family A (ABC1), member 1	I	12	5.05	
EIF2S3	eukaryotic translation initiation factor 2, subunit 3 gamma, 52kDa	I	6	5.01	
PAFAH1B2	platelet-activating factor acetylhydrolase 1b, catalytic subunit 2 (30kDa)	I	14	4.96	
SLC25A43	solute carrier family 25, member 43	I	0	4.94	specific
CAV1	caveolin 1, caveolae protein, 22kDa	I	39	4.78	
ALDH1A1	aldehyde dehydrogenase 1 family, member A1	I	31	4.21	
HSPE1	heat shock 10kDa protein 1 (chaperonin 10)	I	18	4.18	
BPHL	biphenyl hydrolase-like (serine hydrolase)	I	13	4.15	specific
PTBP2	polypyrimidine tract binding protein 2	I	8	3.7	
ANXA1	annexin A1	I	25	3.64	
DHTKD1	dehydrogenase E1 and transketolase domain containing 1	I	14	3.53	
NAT1	N-acetyltransferase 1 (arylamine N-acetyltransferase)	I	8	3.51	specific
MYH11	myosin, heavy chain 11, smooth muscle	I	9	3.51	specific
TACC1	transforming, acidic coiled-coil containing protein 1	I	33	3.3	
POFUT2	protein O-fucosyltransferase 2	I	18	3.29	specific
ACTA1	actin, alpha 1, skeletal muscle	I	7	3.29	specific
SULT1C4	sulfotransferase family, cytosolic, 1C, member 4	I	0	3.03	specific
YWHAZ	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide	I	14	2.82	
CA13	carbonic anhydrase XIII	I	0	2.72	specific
HNRNPH1	heterogeneous nuclear ribonucleoprotein H1 (H)	I	11	2.66	
SLC44A1	solute carrier family 44, member 1	I	10	2.65	
DAD1	defender against cell death 1	I	8	2.64	
CPT1A	carnitine palmitoyltransferase 1A (liver)	I	10	2.55	
ADK	adenosine kinase	I	18	2.47	
PSMA2	proteasome (prosome, macropain) subunit, alpha type, 2	I	1	2.46	
CYP2W1	cytochrome P450, family 2, subfamily W, polypeptide 1	I	44	2.31	specific
CEP128	centrosomal protein 128kDa	I	0	2.31	specific
PRPS2	phosphoribosyl pyrophosphate synthetase 2	I	6	2.27	
DECR1	2,4-dienoyl CoA reductase 1, mitochondrial	I	3	2.22	
ATP5J2	ATP synthase, H+ transporting, mitochondrial Fo complex, subunit F2	I	7	2.14	
SSR3	signal sequence receptor, gamma (translocon-	I	0	2.12	

	associated protein gamma)				
ATP5I	ATP synthase, H+ transporting, mitochondrial Fo complex, subunit E	I	5	2.11	
ANXA5	annexin A5	I	18	2.08	
KPNA4	karyopherin alpha 4 (importin alpha 3)	I	1	2.07	
PTGR1	prostaglandin reductase 1	I	23	2.05	
LAP3	leucine aminopeptidase 3	I	0	1.92	
VCL	vinculin	I	15	1.91	
HNRNPK	heterogeneous nuclear ribonucleoprotein K	I	12	1.91	
ACP1	acid phosphatase 1, soluble	I	10	1.88	
MYL6	myosin, light chain 6, alkali, smooth muscle and non-muscle	I	5	1.85	
BPNT1	3'(2'), 5'-bisphosphate nucleotidase 1	I	0	1.84	
MDH1	malate dehydrogenase 1, NAD (soluble)	I	3	1.82	
VIM	vimentin	I	21	1.8	
ACLY	ATP citrate lyase	I	9	1.79	
NIPSNAP 3A	nipsnap homolog 3A (C. elegans)	I	0	1.77	
CORO1C	coronin, actin binding protein, 1C	I	4	1.77	
VSX1	visual system homeobox 1	I	6	1.75	specific
SULT1E1	sulfotransferase family 1E, estrogen-preferring, member 1	I	7	1.75	specific
SLC25A2 2	solute carrier family 25 (mitochondrial carrier: glutamate), member 22	I	0	1.75	specific
S100P	S100 calcium binding protein P	I	38	1.75	specific
PLS3	plastin 3	I	24	1.74	
DLD	dihydrolipoamide dehydrogenase	I	2	1.73	
MYH10	myosin, heavy chain 10, non-muscle	I	7	1.63	
CPM	carboxypeptidase M	I	23	1.61	
TOMM70 A	translocase of outer mitochondrial membrane 70 homolog A (S. cerevisiae)	I	4	1.54	
RPL7	ribosomal protein L7	I	0	1.52	
NDUFS1	NADH dehydrogenase (ubiquinone) Fe-S protein 1, 75kDa (NADH-coenzyme Q reductase)	I	0	1.48	
S100A11	S100 calcium binding protein A11	I	17	1.47	
HADHB	hydroxyacyl-CoA dehydrogenase/3-ketoacyl-CoA thiolase/enoyl-CoA hydratase (trifunctional protein), beta subunit	I	5	1.44	
NPEPPS	aminopeptidase puromycin sensitive	I	8	1.4	
NQO1	NAD(P)H dehydrogenase, quinone 1	I	15	1.38	
FANCI	Fanconi anemia, complementation group I	I	5	1.29	
RAB8B	RAB8B, member RAS oncogene family	I	0	1.26	
HIBCH	3-hydroxyisobutyryl-CoA hydrolase	I	8	1.25	
PPP2R2A	protein phosphatase 2, regulatory subunit B,	I	19	1.24	

	alpha			
IFT81	intraflagellar transport 81 homolog (Chlamydomonas)	I	0	1.23
HMGCS1	3-hydroxy-3-methylglutaryl-CoA synthase 1 (soluble)	I	0	1.23
PRPF4B	PRP4 pre-mRNA processing factor 4 homolog B (yeast)	I	3	1.2
DBT	dihydrolipoamide branched chain transacylase E2	I	3	1.19
SFXN1	sideroflexin 1	I	0	1.18
NADKD1	NAD kinase domain containing 1	I	0	1.15
ST13	suppression of tumorigenicity 13 (colon carcinoma) (Hsp70 interacting protein)	I	16	1.14
RPN2	ribophorin II	I	8	1.14
NUDT5	nudix (nucleoside diphosphate linked moiety X)-type motif 5	I	0	1.14
CPPED1	calcineurin-like phosphoesterase domain containing 1	I	0	1.08
HSDL2	hydroxysteroid dehydrogenase like 2	I	0	1.05
LTA4H	leukotriene A4 hydrolase	I	8	1.01

Table 2.13.1. 205 putative vascular targets found in the chicken proteomic data. 81 membrane (M) and 43 extracellular genes (E) were found to have an endothelial enriched cDNA library profile, as well as 81 intracellular genes (I).

2.14 summary

In this chapter we utilized proteomic data that aimed to extract proteins in contact with the blood supply in a developing chicken embryo. The majority of this chapter was surveying the proteomic data for validation and identifying tissue specific genes. In respect to vascular targeting, many endothelial and tumour endothelial genes were found, confirmed by endothelial cell cDNA expression and AngioScore analyses. The results will be discussed in chapter 6, focusing on the vascular target/endothelial results.

2.15 Acknowledgements of contributions

I would like to acknowledge and thank the Bikfalvi laboratory and Plate-forme Protéomique de la Génopole Toulouse Midi-Pyrénées group. The bioinformatic analyses was carried out on proteomic data experimentally produced in the Bikfalvi laboratory. The mass-spectrometry experiments and initial raw mass-spectrum to protein sequence identifications (Mascot) were performed by Plate-forme Protéomique de la Génopole

Toulouse Midi-Pyrénées group. Some of the work in this chapter has contributed to a manuscript entitled "Mapping the extracellular and membrane proteome associated with the vasculature and the stroma in the embryo", which will be submitted to a proteomic journal (August 2012).

CHAPTER 3; COMBINING 1ST AND 2ND GENERATION SEQUENCED CDNA LIBRARIES TO IDENTIFY VASCULAR TARGETS

3.1 Introduction

Differential gene expression analyses in previous work and work contributing to this thesis ([62]) identified endothelial and tumour endothelial enriched genes using public cDNA and SAGE libraries. In this chapter, to increase vascular target (VT) gene discovery, four enhancements were made. The first was the production of an RNA-seq library of liver Sinusoidal Endothelial Cells (hHSEC) cultured for 24 hours at 1% oxygen that generated an additional 823,750 sequences to add into the endothelial cDNA library pool (using long-read, Roche 454 Titanium 2nd generation sequencing). The second was the addition of liver tumour and normal bulk cDNA libraries. A third improvement was to pool organs, plus foetal and tumour tissues as a single foetal/tumour versus normal angiogenic screen. The final improvement was the introduction of an up to date Reference Sequence project transcriptome (Refseq [109], June 2012), containing 44,203 sequences that includes 6,525 recently annotated non-coding RNAs (4 years have passed since the previous differential gene expression screens). **Please note, appendix tables are given in a separate volume, also downloadable from http://sara.molbiol.ox.ac.uk/userweb/jherbert/tissue_diffex/appendix_files.html.**

3.2 Using two differential gene expression contrasts to identify Vascular Targets

A two-step analysis was performed to predict vascular targets (VTs). The first stage identified endothelial genes by comparing the expression patterns of genes between endothelial and non-endothelial cells. The second stage involved a comparison of bulk tumour/foetal versus bulk normal cDNA libraries to identify genes up-regulated in foetal tissues (undergoing angiogenesis, thus called the angiogenic screen) and tumours.

Putative VTs were genes both endothelial enriched and preferentially expressed in foetal/tumour tissues.

3.3 data collection and differential gene expression

Sanger 1st generation sequenced cDNA libraries are restricted by costs and labour, and so the number of sequences they contain ranges from a single sequence to approximately one hundred thousand Expressed Sequence Tags (ESTs). To add more sequences to the endothelial cell pool, antibody coupled magnetic beads (anti-CDH5) were used to select for sinusoidal liver endothelial cells. These cells were cultured in normal conditions until 80% confluent and then transferred to a 1% hypoxia chamber for 24 hours. Total RNA was sent for Roche 454 Titanium sequencing at Genepool (Edinburgh) and that produced 823,750 sequences (ranging from 200 to 500bp in length). Public endothelial, non-endothelial, bulk normal and bulk foetal/tumour tissues libraries were collected using the Cancer Genome Anatomy Project (CGAP) tissue library browser [110]. The bulk tissues collected were from brain, skin, colon, kidney, lung, liver and prostate. Table 3.3.1 lists the cDNA EST/454 counts for the different tissues and a full list of cDNA library names can be found in Herbert et al. additional files [62] (open access), with the exception of the liver libraries which are listed in table 3.3.2. The methodology employed for the analyses in this work is similar to that described in Herbert et al. [62], except for the EST to gene assignments, which were made with megablast [111] database searching instead of BLAST [112] (more details in methods and Herbert et al. [62]).

The statistics in these analyses were designed by Dov Stekel [62], which combined a generalised likelihood ratio test with a False Discovery Rate (FDR). Briefly, two hypotheses were compared; the null hypothesis, inferring there is no difference in gene expression between the two cDNA library pools and any differences are due to sampling, and an alternative hypothesis stating a genuine difference in gene expression was due to biological effects. The likelihood ratio statistic (R-statistic) was derived by dividing the likelihood of seeing the data under the null hypothesis by the likelihood of seeing the data under the alternative hypothesis. A log2 fold change was also employed, adding a constant (0.1) value to both pool counts and pool sizes to derive a log2 fold change

(preventing any errors of dividing by zero). Putative VT genes were chosen that had a positive (up-regulation) in both the endothelial and angiogenic screens.

Table 3.3.1 Lists of cDNA/454 mRNA transcript libraries

RNA sequence library	Library information	Number of sequences
Endothelial cells	Endothelial cells from CGAP and dbEST databases	31,114
Liver Hypoxia ECs	Roche 454 Titanium RNA-seq library (generated in this work)	823,750
Non-endothelial cells	Non-endothelial cells from CGAP and dbEST databases	136,622
Brain tumour, normal and foetal libraries	Bulk cDNA libraries taken from CGAP and dbEST databases	140,621 100,554 69,862
Liver tumour and normal libraries	Bulk cDNA libraries taken from CGAP and dbEST databases	27,568 31,023
Prostate tumour and normal libraries	Bulk cDNA libraries taken from CGAP and dbEST databases	19,125 68,480
Colon tumour and normal libraries	Bulk cDNA libraries taken from CGAP and dbEST databases	143,025 37269
Skin tumour and normal libraries	Bulk cDNA libraries taken from CGAP and dbEST databases	12,484 33,218
Lung tumour, normal and foetal libraries	Bulk cDNA libraries taken from CGAP and dbEST databases	108,107 82,757 112,690
Kidney tumour, normal and foetal libraries	Bulk cDNA libraries taken from CGAP and dbEST databases	38,519 72,476 2,605

Table 3.3.1. This table lists the endothelial, tumour, foetal and normal tissue mRNA cDNA and Roche 454 expression libraries used to predict the identity of tumour vascular targets. The raw sequence counts of each mRNA library are included.

Table 3.3.2 Additional liver tumour and normal cDNA libraries

Liver bulk cDNA library	Histology	EST count
subtractive library of hepatocellular carcinoma	hepatocellular carcinoma	7
human hepatoblastoma cDNA	hepatoblastoma	7,898
GKA	hepatocellular carcinoma	84
GKB	hepatocellular carcinoma	1547
GKC	hepatocellular carcinoma	17736
GKD	hepatocellular carcinoma	180
GK	hepatocellular carcinoma	80
Human hepatocellular carcinoma subtracted cDNA library	liver tumour	36
L7N800102	normal liver	1,350
L7N800102s1	normal liver	1,702
Stratagene liver (#937224)	normal liver	8,417
GLA	normal liver	93
GLC	normal liver	19,285
GLD	normal liver	176

Table 3.3.2. All bulk tumour, foetal and normal cDNA libraries were divided into 2 pools; tumour/foetal and normal. All bulk cDNA libraries are listed in the additional files of Herbert et al. [62] (open access) except for the liver tumour and normal libraries that are listed in this table.

3.4 Putative Vascular target results

From the differential gene expression contrasts (endothelial cells versus non-endothelial cells and bulk foetal/tumour versus bulk normal tissues) 3,069 genes were found as up-regulated in both contrasts and are listed in appendix table A.3.4.1. To focus on which genes are more likely to be accessible targets via the blood stream, bioinformatics sequence analyses tools were then used through the online website, Center for Biological Sequence Analysis [113], to predict which of these genes contained transmembrane or signal peptide domain sequences. 698 genes were found to have either one or both of these domains and are listed in appendix table A.3.4.2 as vascular targets.

3.4 Vascular targets and possible functional trends

To be able to gain an understanding of any functional trends present in the vascular target list, the complete set of differentially expressed genes and the transmembrane/signal peptide containing VT proteins was analysed independently using the online DAVID resource [114]. This clusters genes with similar functions (selected as Gene ontology biological process) were placed into annotation clusters. The most abundant categories for the whole set of differentially expressed genes (Fig. 3.4.1) were "transcription", "cell cycle", "regulation of transcription" and "metabolic processes", which appear consistent with a developing tumour mass. "Angiogenesis" and "vascular development" categories were also present but at a lower False Discovery Rate (FDR) significance level.

The 698 transmembrane and/or secreted VT genes were separately used as input to DAVID to find enrichment of functional Gene Ontology Biological Process terms (≤ 1e-2 FDR, Fig. 3.4.2). Biological process functions involved in "vascular development" and "cell adhesion" were present, consistent with the predicted angiogenesis and extra-cellular matrix remodelling processes usually occurring in solid tumour growth.

Figure 3.4.1 All differentially expressed genes, Gene ontology Biological Process

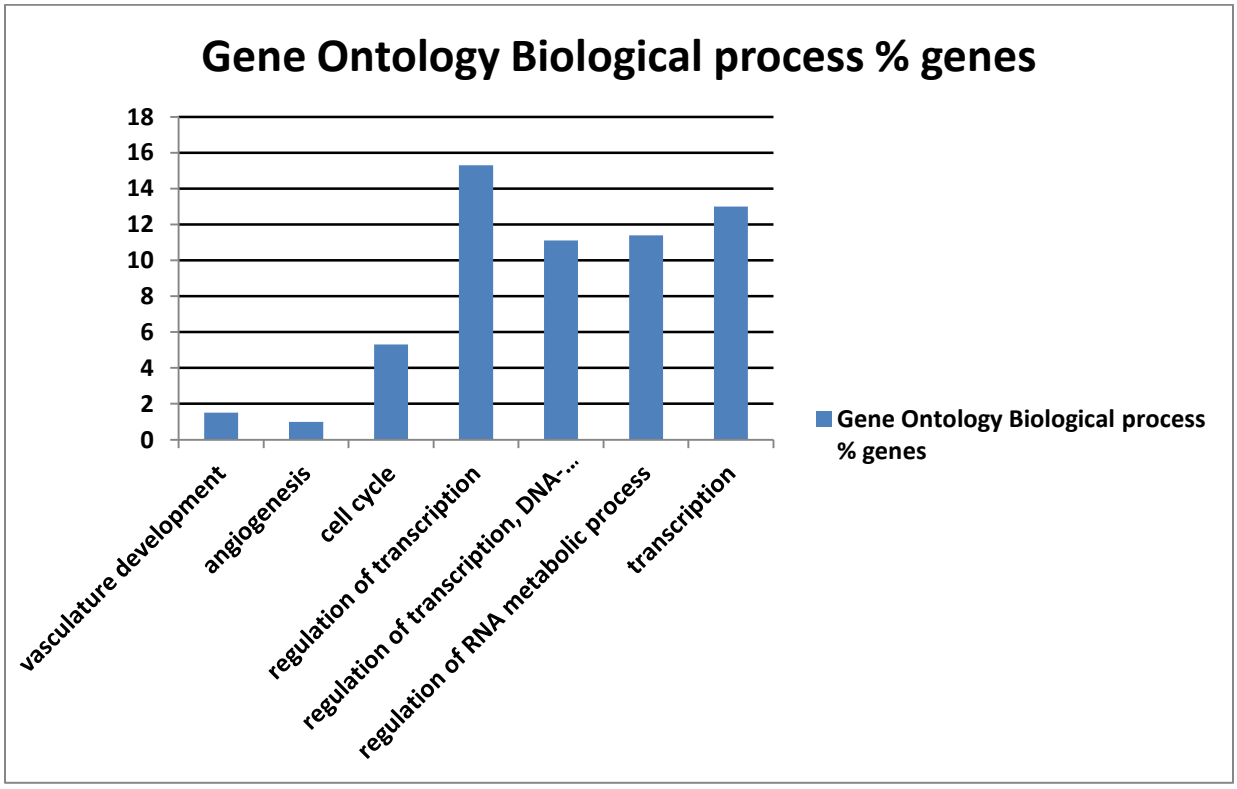


Figure 3.4.1. The 3,069 differentially expressed genes were analysed using the online DAVID resource. The significant terms found enriched (≤ 1e-4 FDR) were "transcription", "cell cycle", "regulation of transcription" and "metabolic processes". "Angiogenesis" and "vasculature development" were also found at a lower significance level (< 1e-1 FDR).

Figure 3.4.2 Gene ontology Biological Process of transmembrane/signal-peptide protein genes

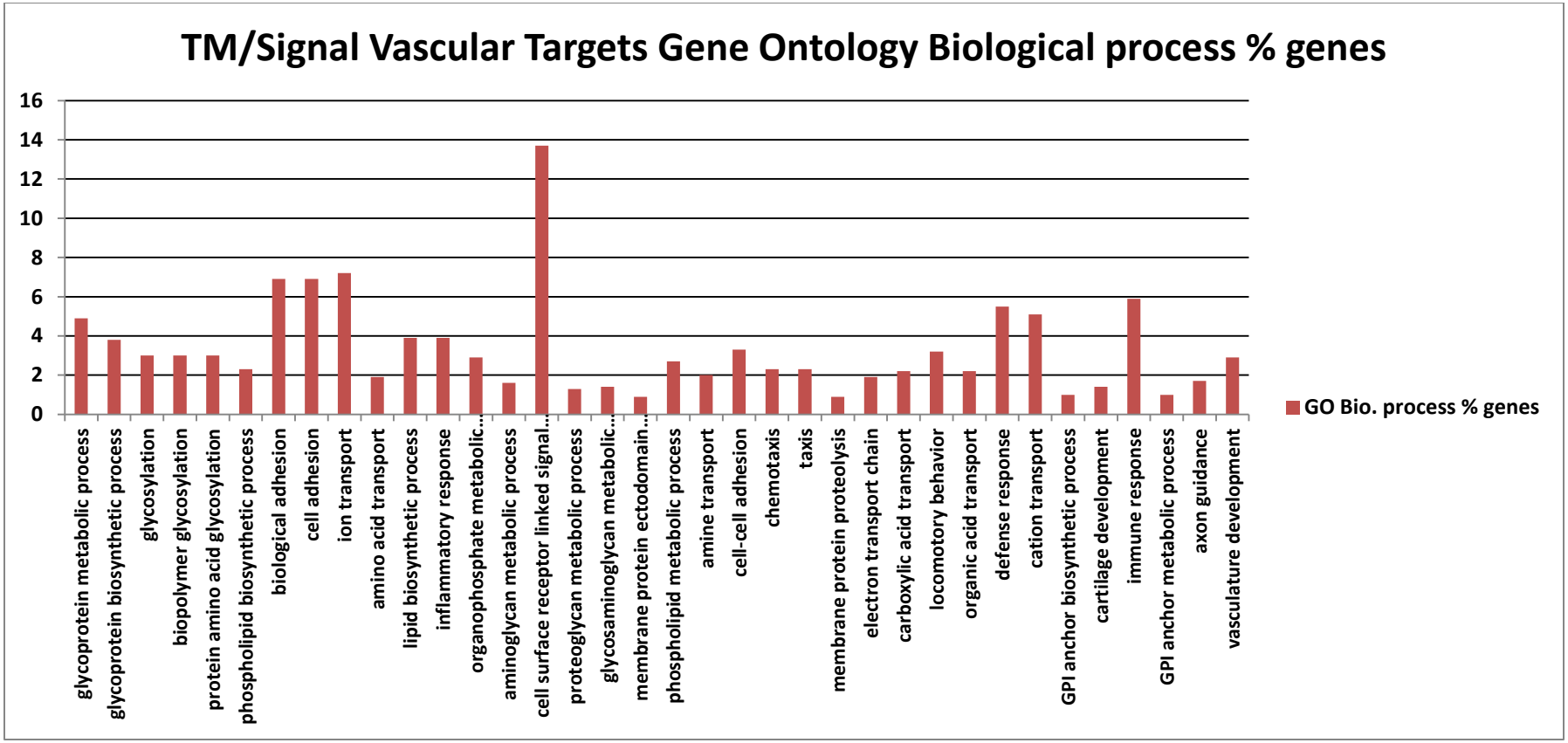


Figure 3.4.2. 698 vascular targets were used as input to DAVID to find enrichment of Gene Ontology Biological Process terms (≤ 1e-2 FDR). Biological processes involved in "vascular development" and "cell adhesion" were present, consistent with predicted angiogenesis and extra-cellular matrix remodelling usually associated with tumour growth.

3.5 VT potential involvement in angiogenesis, tumours and endothelium

Abstract keyword searches were used to generate an "AngioScore" for all differentially expressed genes. An "AngioScore" is calculated as the percentage of abstracts containing a given keyword divided by the total number of abstracts for that gene (only from publications that Entrez Gene at Genbank assign to a gene). A list of the keywords used is as follows (more details in methods): "vascular", "tumour", "tumor", "angiogenic", "angiogenesis", "neovascularization", "neovascularisation", "vasculogenesis", "hypoxia", "endoth" and "VEGF". The notion was to rank genes that had been previously reported to be involved in angiogenesis, tumours and endothelial cell expression. This serves to give confidence in the list of genes found and enable further experiments on other genes that have not been previously researched. A full list of the "AngioScore" keyword results are listed in appendix table A.3.5.1 and the top 24 "AngioScore" genes are listed in table 3.5.1, where there are characterised endothelial genes such as cadherin 5, type 2 (vascular endothelium) and endothelial cell-specific chemotaxis regulator (ECSCR or Endothelial Cell Specific Molecule 2); and also known tumour endothelial (VT) genes: ANGPT2, ESM1 and IGFBP7.

Table 3.5.2 lists a selection 33 VT candidates with the least number of publications to the present date (June 2012). Experimental proof of VT expression profiles would have to be performed to confirm the novel VT candidates.

Table 3.5.1 VT genes with highest scoring "AngioScore"

Gene	Product	Total abstracts	AngioScore
CDH5	cadherin 5, type 2 (vascular endothelium)	115	**90**
ECSCR	endothelial cell-specific chemotaxis regulator	6	**83**
HIF3A	hypoxia inducible factor 3, alpha subunit	21	**71**
ANGPT4	angiopoietin 4	25	**68**
ESM1	endothelial cell-specific molecule 1	25	**68**
ANGPT2	angiopoietin 2	225	**66**
PDPN	podoplanin	96	**66**
IGFBP7	insulin-like growth factor binding protein 7	63	**63**
ADAM17	ADAM metallopeptidase domain 17	194	**63**
TNFRSF1A	tumor necrosis factor receptor superfamily, member 1A	409	**62**
LECT1	leukocyte cell derived chemotaxin 1	15	**60**
PEAR1	platelet endothelial aggregation receptor 1	5	**60**
DLL4	delta-like 4 (Drosophila)	52	**60**
ENG	endoglin	242	**59**
HPSE	heparanase	193	**58**
PLXND1	plexin D1	19	**58**
JAM3	junctional adhesion molecule 3	46	**57**
RECK	reversion-inducing-cysteine-rich protein with kazal motifs	55	**56**
STAB1	stabilin 1	30	**53**
SELE	selectin E	321	**53**
TNFRSF25	tumor necrosis factor receptor superfamily, member 25	39	**51**
S1PR2	sphingosine-1-phosphate receptor 2	32	**50**
IGFLR1	IGF-like family receptor 1	2	**50**
VSIG1	V-set and immunoglobulin domain containing 1	2	**50**

Table 3.5.1. This table shows 24 genes with the highest "AngioScore" from the 698 VTs. Known endothelial and tumour endothelial genes are found in this list; e.g. CDH5 and ECSR are published as being highly endothelial specific and ANGPT2, ESM1 and IGFBP7 are published tumour endothelial markers.

Table 3.5.2 Potential novel VT genes with fewest publications

Gene	Product	Total abstracts	Angio Score
LOC100506250	NPIP-like protein 1-like	0	**0**
C15orf24	chromosome 15 open reading frame 24	0	**0**
TMPRSS11BNL	TMPRSS11B N terminal-like	0	**0**
LOC100509247	hypothetical protein LOC100509247	0	**0**
LOC100506504	hypothetical protein LOC100506504	0	**0**
LOC100652934	hypothetical protein LOC100652934	0	**0**
XKR5	XK, Kell blood group complex subunit-related family, member 5	1	**0**
LOC100507050	hypothetical LOC100507050	1	**0**
LOC147670	hypothetical protein LOC147670	1	**0**
LRRC38	leucine rich repeat containing 38	1	**0**
RPRML	reprimo-like	1	**0**
UPK3BL	uroplakin 3B-like	1	**0**
NANOGNB	NANOG neighbor homeobox	1	**0**
PRR23A	proline rich 23A	1	**0**
IGFLR1	IGF-like family receptor 1	2	**50**
VSIG1	V-set and immunoglobulin domain containing 1	2	**50**
SVOP	SV2 related protein homolog (rat)	2	**0**
MFSD11	major facilitator superfamily domain containing 11	2	**0**
TMEM80	transmembrane protein 80	2	**0**
FAM18A	family with sequence similarity 18, member A	2	**0**
STEAP1B	STEAP family member 1B	2	**0**
C19orf18	chromosome 19 open reading frame 18	2	**0**
C18orf26	chromosome 18 open reading frame 26	2	**0**
MILR1	mast cell immunoglobulin-like receptor 1	2	**0**
TMEM183B	transmembrane protein 183B	2	**0**
NTN5	netrin 5	2	**0**
TTC9B	tetratricopeptide repeat domain 9B	2	**0**
FAM69B	family with sequence similarity 69, member B	3	**33**
C14orf101	chromosome 14 open reading frame 101	3	**0**
TLCD1	TLC domain containing 1	3	**0**
TMEM194B	transmembrane protein 194B	3	**0**
GJD4	gap junction protein, delta 4, 40.1kDa	3	**0**
TMEM179B	transmembrane protein 179B	3	**0**

Table 3.5.2. This table shows the 33 putative VT genes with the fewest publications. The full set of 3,069 genes with publications and "AngioScore" are in appendix table A.3.5.1.

3.6 Summary of VT discovery using combined 1st and 2nd generation sequencing

In summary, 698 putative vascular targets were predicted combining the old Sanger sequencing technologies (cDNA libraries) with the newer 2nd generation sequencing technology of Roche 454 Titanium. Functional enrichment showed tumour growth, whilst angiogenesis related functions and searches of published literature suggested that the gene list could include previously unreported novel vascular targets. During the time of this PhD work, the basis of the cDNA library analyses presented in this chapter was published in BMC Genomics, entitled "A novel method of differential gene expression analysis using multiple cDNA libraries applied to the identification of tumour endothelial genes" [62].

3.7 Serving the scientific community

1) Also, during the work of this thesis, a simple guide to analysing cDNA libraries to identify genes of interest was produced and published in a book called "cDNA libraries: Methods and Applications, Methods in Molecular Biology 729. Springer" [115].
2) An accompanying website was also created where researches can run their own analyses, either through an interactive web page or with provided command line tools [116]. The book chapter website and tool downloads are available at; http://sara.molbiol.ox.ac.uk/userweb/jherbert/

CHAPTER 4; USING RNA-SEQ DEEP SEQUENCING TO FIND PUTATIVE PAN-VASCULAR TARGETS

4.1 Introduction

In the previous 2 chapters, VTs were sought using proteomic, cDNA and 2nd generation sequencing (Roche 454 long read data). In this chapter, pan-VT markers were identified combining RNA-seq data (2nd generation, Illumina and SOLiD short read technologies) from tumour endothelial cells freshly isolated from lung, colon and bladder organs, together with an *in-vitro*, tumour conditioned HUVEC library (tumour Huvec) and the hypoxic Sinusoidal endothelial library (hHsec). Also from this data, a novel endothelial transcriptome was assembled by merging the assemblies of each RNA-seq library. This was subsequently followed by a screen to identify differentially expressed novel intergenic transcripts and an analyses to test for evidence of isoform switching. This was to see if there exist coding isoforms for larger proteins in tumours and, at the same time, to see if shorter proteins are enriched in normal tissues, based on Cufflinks results. If proven, these are potentially similar to those previously reported (TNC [117] and FN1 [118]). The ultimate objective was to identify vascular targets from deep sequencing data of tumour endothelium.

4.2 Sequencing an in-vitro tumour conditioned endothelial transcriptome(Tumour Huvec)

The term hypoxia means a reduced level of oxygen tension within tissues and this can happen in cancer, leading to therapy resistance and a poor prognosis. Hypoxia is a key regulator of tumour growth [119] because the protein level of Hypoxia-inducible factor-1alpha (HIF-1alpha) is induced by hypoxia and this increases VEGF production, which is the most pivotal and potent growth factor in angiogenesis activation. The tumour environment is also acidic and it has been shown *in-vitro* that a low pH does not affect endothelial cell activation in the presence of angiogenic factors such as bFGF or VEGF

[120]. The aim of this experiment was to sequence a tumour endothelial transcriptome from an *in-vitro* source of RNA. If the environment is truly that of a tumour, using HUVECs in a culture dish negates any chance of contaminating tissues such as perivascular or smooth muscle cells in the resulting transcriptome. A good example of this, is the tumour endothelial marker 1, endosialin (TEM1/CD248) [69] which was first thought to be expressed on tumour endothelium but is in fact expressed on surrounding perivascular cells [121, 122].

To simulate the environment of a tumour, passage 0 HUVECs were grown from fresh cords at a Ph of 6.8, 1% oxygen (hypoxia), without glucose and in the presence of the potent angiogenic factor, VEGF. To minimise the effects of individual HUVEC gene expression, 4 cord HUVEC cell isolates were mixed to give a heterogeneous population. The cells were grown in a hypoxic incubator (1% O2) for 24 hours at pH 6.8, in glucose free media and with VEGF. The RNA was extracted using a Qiagen RNeasy kit and the 18 and 28s rRNA integrity tested on a 1% gel, which were intact (see Fig. 4.2.1).

4.3 Up-regulation of the hypoxia related gene, DLL4

As a control, a gene known to be up-regulated by VEGF and hypoxia [123, 124] is delta-like 4 (Drosophila), (DLL4). Therefore, to give evidence of a tumour environment transcriptome, DLL4 was tested versus a Normoxia control (4% oxygen, Ph 7.4, glucose and restricted growth factors 10% FCS). This was done using Quantitative PCR (qPCR), using flotillin 2 as a house-keeping gene. The relative expression of DLL4 to the control gene FLOT2 in triplicate was quantified and the results are shown in figure 4.3.1. DLL4 showed up-regulation in tumour conditioned HUVEC (tumour Huvec) versus the normal conditioned counterpart and the sample was sent to the University of Edinburgh (Genepool) for 3 lanes of Illumina GII, 50bp paired-end RNA-seq transcriptome sequencing.

Figure 4.2.1 RNA Agarose gel test for mRNA integrity

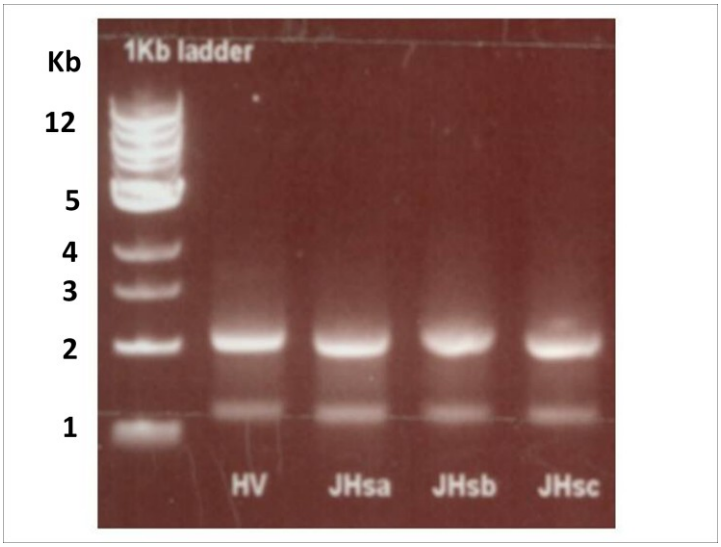


Figure 4.2.1. This figure shows the results of loading total RNA isolated from Human Umbilical Cord samples onto a 1% Agarose gel. High quality and intact mRNA was inferred from the clearly visible 18s and 28s rRNA bands.

Figure 4.3.1 Testing for induction of the hypoxia gene DLL4

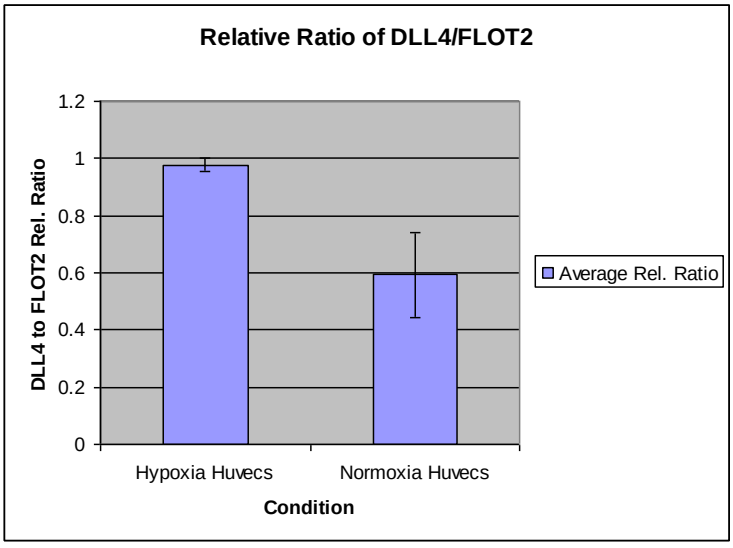


Figure 4.3.1. DLL4 was validated as being up-regulated in the tumour conditioned HUVEC versus HUVEC cultured under normal conditions (in the presence of glucose, reduced growth factors, 4% oxygen and a pH of 7.4).

4.4 Collection of normal healthy adult transcriptome libraries

To contrast the tumour Huvec and hHsec transcriptomes, samples representing normal adult healthy tissues were taken from a publication where human splice variants had been surveyed [125]. The corresponding author was contacted to ascertain whether the tissues originated from normal and healthy adults. This was confirmed and the tissues used here were from the brain, cortex, lung, liver, heart and muscle (Sequence Read Archive (SRA) accession numbers: Lung = SRR014265, heart = SRR014263, liver = SRR014264, brain = SRR036966, cerebral cortex = SRR014262 and muscle = SRR014266).

4.5 Vascular targets from a Tumour Huvec and hHsec versus normal human tissue screen

An early measure of RNA expression for short read next generation sequencing was devised by Mortazavi et al. [126], who invented the term Reads Per Kilobase of exon model, per Million reads mapped (RPKM). This measure of gene expression took into account differing gene transcript lengths and performed a rough normalizing step as RNA-seq libraries were sequenced to different depths (total numbers of reads sequenced). Using the Refseq transcriptome as a measure of gene length (the average was taken when multiple isoforms were present), RPKM values were calculated for the tumour Huvec, hHsec and normal adult tissue libraries to be used in an initial 2nd generation sequencing differential gene expression experiment. Fold change to the logarithm base 2 (log2(FC)) was used to rank the genes, using the average RPKM of tumour Huvec + hHsec compared to the average RPKM of all normal tissues. A $\log2(FC) \geq 3$ was used as a cut-off for selecting differentially expressed VT genes versus normal tissues and 753 genes were found using this criterion (appendix table A.4.5.1). Please note, 4 of these genes have now been removed from the NCBI (LOC100127894, LOC100130764, MIR1274B, and MTND4P). The top 252 genes in this list are classed as membrane or extracellular by either signal peptide, membrane domain or cellular component from Gene Ontology (Entrez Gene). Table 4.5.1 lists the top 33 genes with the highest endothelium enrichment versus normal bulk tissues. Several known vascular related targets are listed; MMP1, ESM1, ANGPT2 and ANXA1.

Table 4.5.1 Potential VT found from the RPKM analyses

Log2(FC) tumour huvec and hHsec combined		Product	Best TM	Signal peptide
13.37	TMSB4X	thymosin beta 4, X-linked	0	extracellular
9.38	**ESM1**	endothelial cell-specific molecule 1	0	signal
9.07	MMP1	matrix metallopeptidase 1 (interstitial collagenase)	0	signal
8.00	SERPINE1	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1	0	signal
7.62	RELL1	RELT-like 1	2	signal
7.50	CAPZA1	capping protein (actin filament) muscle Z-line, alpha 1	0	extracellular
7.29	CLIC1	chloride intracellular channel 1	0	membrane
7.07	THBS1	thrombospondin 1	0	extracellular
6.98	CLDN11	claudin 11	4	signal
6.96	AXL	AXL receptor tyrosine kinase	1	signal
6.89	EFEMP1	EGF containing fibulin-like extracellular matrix protein 1	0	signal
6.83	NUDT19	nudix (nucleoside diphosphate linked moiety X)-type motif 19	0	signal
6.78	SLC35F2	solute carrier family 35, member F2	9	
6.54	ANGPT2	angiopoietin 2	0	signal
6.38	PTX3	pentraxin 3, long	0	signal
6.32	ANXA1	annexin A1	0	extracellular
6.30	LOXL2	lysyl oxidase-like 2	0	signal
6.23	SEP15	15 kDa selenoprotein	0	signal
6.21	MARCKS	myristoylated alanine-rich protein kinase C substrate	0	membrane
6.20	TFPI2	tissue factor pathway inhibitor 2	0	signal
6.19	CRIM1	cysteine rich transmembrane BMP regulator 1 (chordin-like)	1	signal
6.12	GDF6	growth differentiation factor 6	0	signal
6.10	ECSCR	endothelial cell-specific chemotaxis regulator	1	signal
6.09	CTGF	connective tissue growth factor	0	signal
6.07	SULF1	sulfatase 1	0	extracellular
6.04	HHIP	hedgehog interacting protein	0	signal
6.04	TM4SF18	transmembrane 4 L six family member 18	4	signal
6.01	MMRN1	multimerin 1	0	membrane
6.01	EDN1	endothelin 1	0	signal
5.97	SLC12A2	solute carrier family 12	12	membrane

		(sodium/potassium/chloride transporters), member 2		
5.95	HSPG2	heparan sulfate proteoglycan 2	0	extracellular
5.89	MGP	matrix Gla protein	1	signal
5.77	MMP10	matrix metallopeptidase 10 (stromelysin 2)	0	Signal

Figure 4.5.1. On this table are the 33 VT's with the highest log2(FC) of endothelium versus normal bulk tissues from the RPKM analyses. Several known vascular related targets are listed; MMP1, ESM1, ANGPT2 and ANXA1.

4.6 Using short read 2nd generation sequencing data to find pan Vascular Targets

Depending on the endothelial environment, different repertoires of genes are being expressed. Within different organ tumours, it is hypothesized that both organ specific and general vascular genes are being expressed due to the tumour environment of the organ. In this study, we sought to find those specifically expressed in a tumour environment of multiple tumours. Given general markers of tumour endothelium and the fact that endothelium is in contact with blood, gives the opportunity to deliver a multiple tissue vascular target drug directly into the blood stream as a general cancer therapeutic. Next generation sequencing facilitates snapshots of mRNA repositories in cells types of interest. Here we use the *in-vitro* tumour environment Huvec RNA-seq library combined with RNA-seq libraries from tumour endothelial cells freshly isolated from lung, colon and bladder tumours using Ulex europaeus agglutinin 1 coupled magnetic beads, to identify pan-TEM or pan-VTs.

The RNA-seq paired-end libraries were sequenced on either a SOLiD4 or an Illumina sequencing machine. The different samples are listed in the bullet point list below:

- *in-vitro*; Tumour conditioned Human Umbilical Vein Endothelial Cells (Tumour Huvec), Illumina GII sequencing
- Bead isolated lung tumour endothelium x2 (one SOLiD4 sequenced and the other Illumina Hiseq sequenced)
- Bead isolated lung normal endothelium (SOLiD4)
- Bead isolated tumour bladder endothelium (SOLiD4)
- Bead isolated tumour colon endothelium (SOLiD4)
- Bead isolated normal colon endothelium (SOLiD4)
- Public domain, bulk lung normal Illumina GI library from an adult human [125]

4.7 High quality and intact RNA attained before sequencing

High quality and intact RNA from endothelial cells was isolated from the fresh tumour and normal lung, bladder and colon tissues using Ulex europaeus agglutinin 1 (UEA-1) coated magnetic beads. A schematic diagram of the approach is displayed in figure 4.7.1. The high quality of RNA can be seen in figure 4.7.2. The top panel of this figure shows the results of an Agilent Bioanalyser, which generates an RNA integrity number (RIN), dependent on the quality of RNA, and RIN values ≥ 7 were attained for all samples sequenced. To ascertain that successful endothelial cell isolation was occurring during extraction, a QPCR test was performed comparing the expression levels of an endothelial cell marker in both EC isolates and bulk tissue controls. Significant up-regulation of the endothelial marker CD31 (PECAM1) was seen in isolated endothelium versus bulk tissue (bottom panel of figure 4.7.2), suggesting high endothelial cell enrichment.

Figure 4.7.1 Tumour endothelial cell isolation from lung, bladder and colon

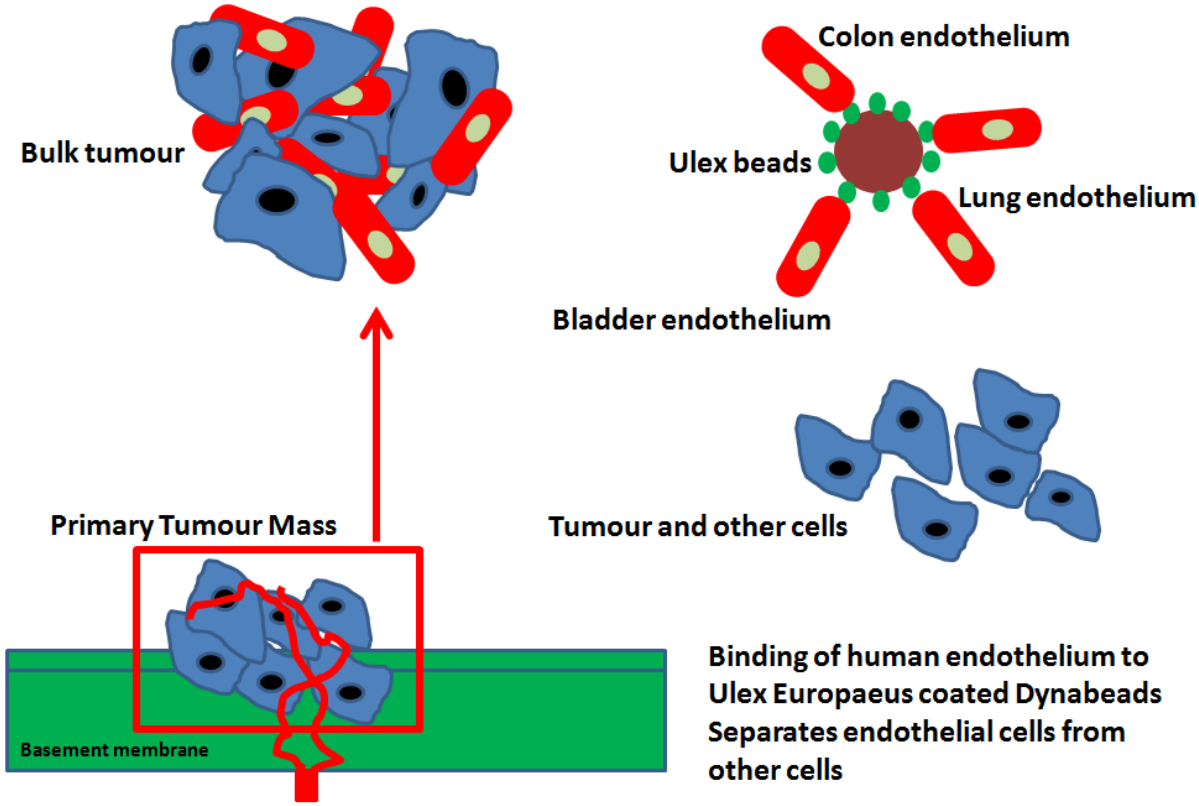


Figure 4.7.1; This figure portrays the procedure used to isolate enriched populations of endothelial cells from bulk tumour tissues (same as applied to normal tissues). Fresh tumours were immediately stored on ice and transported to the laboratory for endothelial cell isolation using Ulex Europaeus protein coated beads that bind to sugar molecules on the surface of endothelium.

Figure 4.7.2 Validation of endothelial cell enrichment and RNA integrity

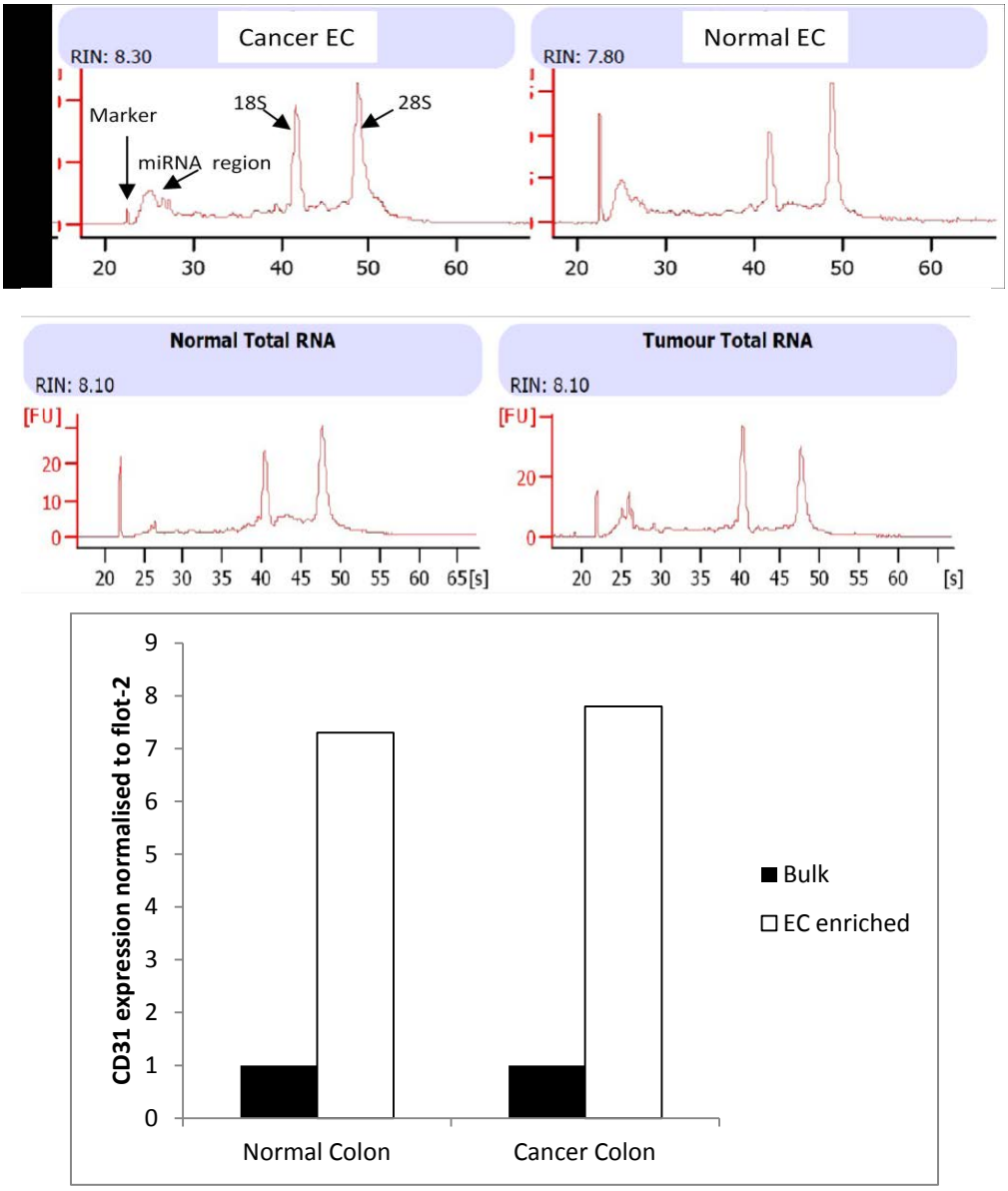


Figure 4.7.2. The top panel of this figure portrays the RNA integrity assessed using an Agilent Bioanalyser. 18s/28s RNA remained intact following the isolation procedure and RIN values ≥ 7 were attained for all samples sequenced. The bar chart shows QPCR validation of the enrichment of CD31 expression in bead isolated endothelial cells from normal and tumour tissues versus that of bulk tissue. FLOT2 was used as a housekeeping control.

4.8 Bioinformatics pipeline to analyse RNA-seq data

Figure 4.8.1 gives a diagrammatical overview of the bioinformatic differential gene expression experiments carried out to find putative VT's using short read SOLID4 and Illumina Hiseq sequencers. A more detailed protocol is listed in the next bullet list, which is based on a Nature Protocol [127].

- First a Gene Transfer Format (GTF) reference file (Refseq GTF) was chosen as a reference genome annotation file to guide transcript assemblies [128].
- To determine fragment sizes of sequenced fragments, bowtie 0.12.7 [102] was run in paired end mode and parsed with a Perl program to establish fragment lengths and length standard deviations of the different samples.
- The Tophat algorithm [103], which is able to map reads spanning an intron, was run to map all reads from each sample separately onto the human genome, version hg19.
- Each set of mapped reads for all samples were separately assembled into a transcriptome using the Cufflinks program (using the Refseq GTF file as a guide) from the Cufflinks suite of programs [104].
- A tumour and normal endothelial transcriptome was then assembled using the Cuffmerge program (from Cufflinks) to merge all the different transcript assemblies.
- Differential gene expression was then performed using the Cuffdiff program (also from Cufflinks) and output files parsed to find differentially expressed genes, novel genes and those with a potential inserted coding exon.

The success statistics of Tophat read mapping are given in table 4.8.1. All libraries were paired end reads except for the public SRA data, accession SRR014265. At best, a total of 82% of fragments from the Illumina Hiseq library were mapped to the genome, which was the RNA-seq library with the longest read length of 101bp. The normal lung

endothelium library was the library with the highest percentage (56%) of reads mapped using the SOLiD technology (paired end, 50bp and 35bp).

Figure 4.8.1 Overview of 2nd generation sequence experiments performed

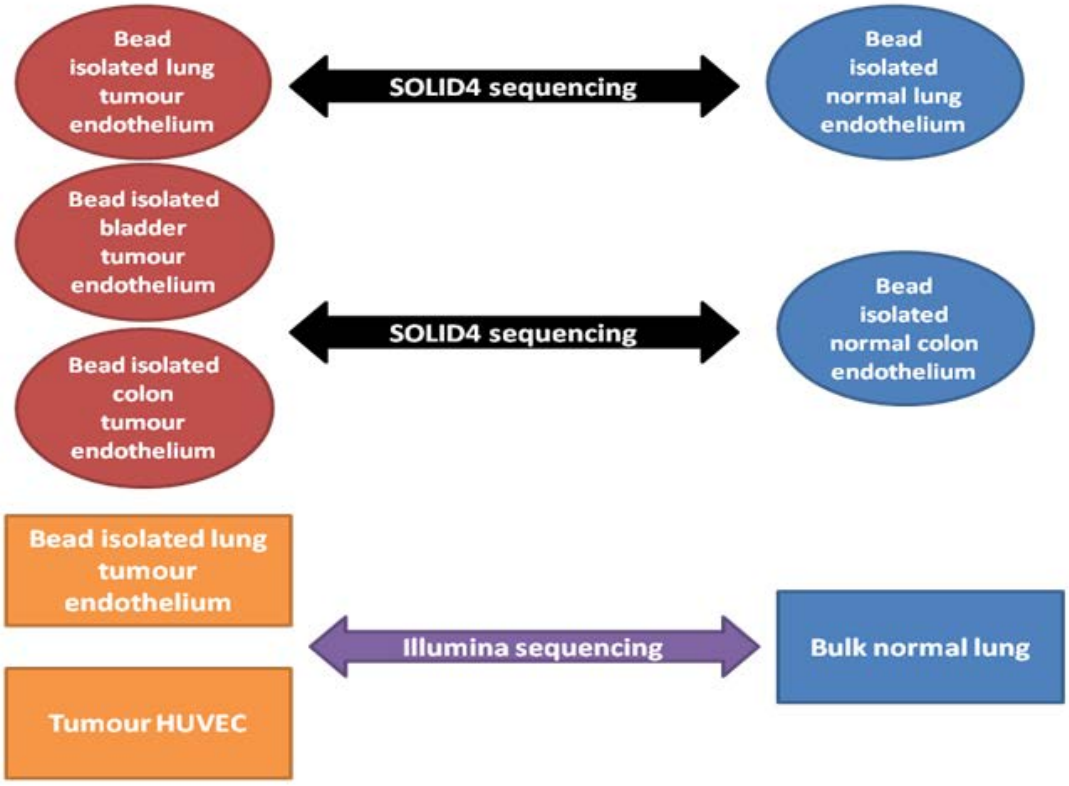


Figure 4.8.1. This figure overviews the experimental design for 7 samples of endothelial cells isolated from various organs and sequenced on a mix of 2nd generation sequencing technologies. Red and blue ovals depict bead isolated endothelium subject to SOLiD4 sequencing, orange squares to Illumina Hiseq/GII sequencing and an Illumina GI library taken from the Gene Expression Omnibus database (GEO). Note; Tumour HUVEC is HUVEC cultured under tumour conditions for 24 hours before library preparation and sequencing.

Table 4.8.1 Tophat fragment mapping statistics for all RNAseq sample libraries

Technology	Tissue	Total fragments	Number of mapped fragments	% fragments mapped	Read length	Paired or single end	Source
Illumina GAI	Normal bulk lung	25862057	18204695	70	32	single	Publication SRR014265
Illumina GAII	Tumour conditioned HUVEC	24469421	16508654	67	50 - 50	paired	Genepool
Illumina Hiseq	Tumour Endo Lung	84197750	69247065	82	101 - 101	paired	Genepool
SOLiD4	Tumour Endo colon	86542525	42402875	49	50 - 35	paired	Birmingham
SOLiD4	Tumour Endo lung	79086167	34649254	44	50 - 35	paired	Birmingham
SOLiD4	Tumour Endo bladder	76774924	31875963	42	50 - 35	paired	Birmingham
SOLiD4	Normal Endo lung	84430299	47478222	56	50 - 35	paired	Birmingham

SOLiD4	Normal Endo colon	79929637	21564021	27	50 - 35	paired	Birmingham

Table 4.8.1. Each RNAseq library was mapped to the human genome (hg19) using Tophat and the number of fragments that mapped is given. All libraries were paired end reads except for the public data from the Sequence Read Archive accession number SRR014265. At best, a total of 82% of fragments from the Illumina Hiseq library were mapped to the genome, which were the longest read length of 101bp. The normal endothelium library from SOLID4 was the library with the highest percentage of reads mapped using SOLiD (paired end, 50bp and 35bp). Source = where the sequencing was performed. Publication = Pan et al. [125], Genepool = sequencing service at Edinburgh University and Birmingham = Technology hub sequencing and bioinformatics, University of Birmingham.

4.9 Endothelial enrichment in the RNA-seq libraries

In previous studies that used magnetic coupled antibodies to isolate endothelium, a measure of enrichment or endothelial cell purity was made by measuring gene expression levels of certain cell type specific genes. For instance, St. Croix et al. divided 8 genes into 4 groups of two (Selection, Epithelial, Hematopoietic and Endothelial groups) and measured the Serial Analysis of Gene Expression (SAGE) library tag abundances for each of those genes, verifying their endothelial cell enrichment [69]. As Cuffdiff produces a measure of gene expression (FPKM, which is the same as RPKM, except for the word fragment is used instead of read; as you get 2 reads per paired end sequenced fragment), a similar analyses was carried out; for this a log2 fold change of endothelial markers versus markers of other cells was used. The 7 endothelial specific/enriched genes were PECAM1, CDH5, MCAM, TIE1, ANPEP, ECSCR and ROBO4 and the markers of other cell types: EPCAM, KRT8 and CDH1 as markers of epithelium, CD14, CD68 and ITGAM for Hematopoietic markers, and PDGFRA as a marker of pericytes. As can be clearly seen in figure 4.9.1, there is significant endothelial enrichment, particularly in the tumour Huvec *in-vitro* model system. Positive log2(fold changes) were also seen for the Lung tumour (SOLiD4), Lung Tumour (Illumina), Bladder tumour (SOLiD4) and Lung normal (SOLiD4) bead isolated endothelium.

Figure 4.9.1 Endothelial cell markers contrasted with markers of other cells

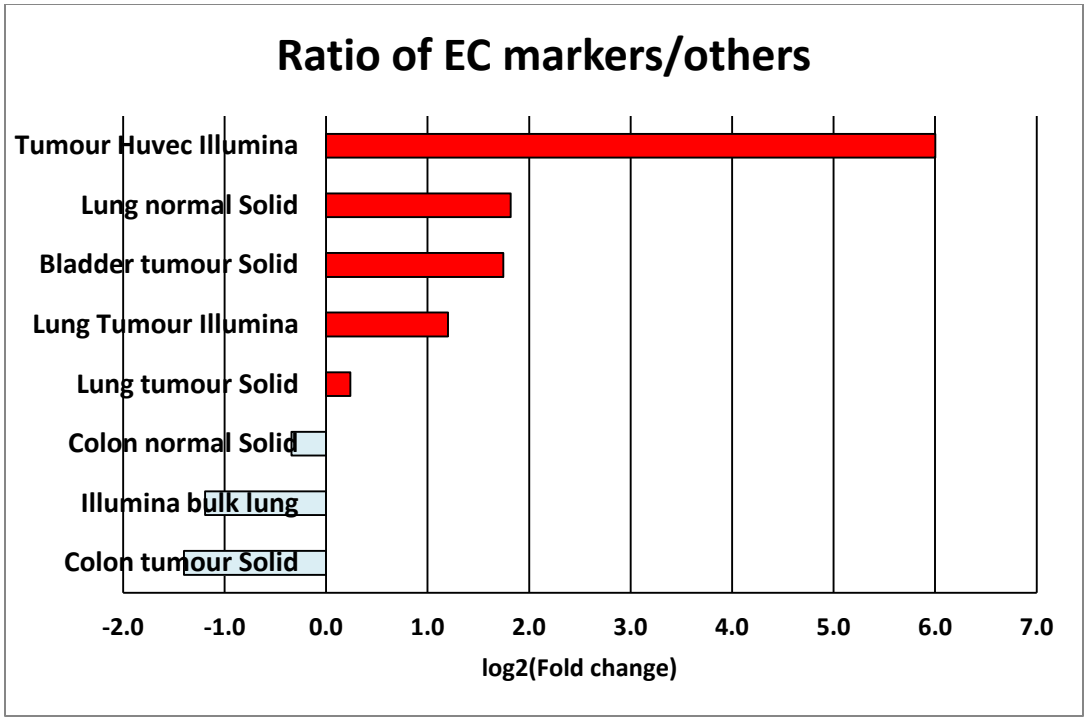


Figure 4.9.1. This figure shows the log2(FPKM fold change ratio) of seven markers of endothelial cells versus seven markers of other cell types. The seven endothelial markers were: PECAM1, CDH5, MCAM, TIE1, ANPEP, ECSCR and ROBO4, and the markers from other cell types were: EPCAM, KRT8 and CDH1 as markers of epithelium: CD14, CD68 and ITGAM as Hematopoietic Markers, and PDGFRA as a marker of pericytes. The *in-vitro* tumour environment Huvec was found the purest source of endothelial cells. The purest bead isolated endothelium were those from the normal lung and bladder tumour tissues.

4.10 Differentially expressed, enriched genes in tumour EC

Differentially expressed genes were sought employing the Cuffdiff algorithm version 1.3.0, performing five different contrasts as follows;

- Tumour lung EC (SOLiD4) versus normal lung EC (SOLiD4)
- Tumour lung EC (Illumina) versus bulk normal lung (Illumina, public)
- Tumour bladder EC (SOLiD4) versus normal colon EC (SOLiD4)
- Tumour colon EC (SOLiD4) versus normal colon EC (SOLiD4)
- Tumour conditioned HUVEC versus bulk normal lung (Illumina, public)

The output of Cuffdiff is a series of files, one of which portrays the differentially expressed genes. A criterion was chosen that defined whether a gene was significantly differentially expressed; a gene was enriched in tumour endothelial cells if the log2(fold change) was ≥ 2 in tumour endothelium versus normal endothelium. For the five individual contrasts, the number of differentially expressed genes was as follows:

- Lung tumour EC = 2,477 differentially enriched genes (SOLiD4)
- Lung tumour EC = 409 differentially enriched genes (Illumina)
- Bladder tumour EC = 768 differentially enriched genes (SOLiD4)
- Colon tumour EC = 570 differentially enriched genes (SOLiD4)
- Tumour Huvec = 215 differentially enriched genes (Illumina)

The full list of genes symbols for the above contrasts can be seen in appendix table A.4.10.1, where a total of 3,922 non-redundant genes were enriched in tumour endothelium using the log2(FC) ≥ 2 for all 5 contrasts combined.

4.11 pan-Vascular Target genes; genes enriched in multiple contrasts

Instead of finding tissue specific enriched vascular targets, pan-vascular-targets (pan-VT) were identified using a Venn diagram approach. Appendix table A 4.10.1 has a column stating the number of contrasts a gene was differentially expressed in, and a Venn

diagram in figure 4.11.1 displays the count overlaps. Zero genes were found to be enriched in all five contrasts, 3 genes were found in four contrasts, 61 genes in three contrasts, 386 genes in two contrasts and the remaining 3,472 in a single contrast. Since, from the five contrasts, two were lung endothelial contrasts using different sequencer technologies, these were combined to make a four contrast Venn diagram. This approach identified a pan-VT for all tissues types contrasted and this gene was Endothelial Specific Molecule 1 (ESM1); the Venn results can be seen in figure 4.11.2 where ESM1 was in all four contrasts, 23 genes were up-regulated in three contrasts, 358 in two contrasts and 3,540 genes in a single contrast. ESM1 was the most prominent pan VT found in these analyses, which is exemplified when looking at the histogram of reads mapped to the ESM1 exons from the RNA-seq data. There is a striking difference between the results of tumour (Figure 4.11.3) and normal (Figure 4.11.4) endothelium.

As it is preferable for VT genes to be easily accessible via the blood stream, two tables of genes listing VT of multiple contrasts can be viewed in tables 4.11.1 and 4.11.2, which list membrane and extracellular putative VT genes respectively.

Figure 4.11.1 Venn diagram showing overlaps between all five differential gene expression contrasts

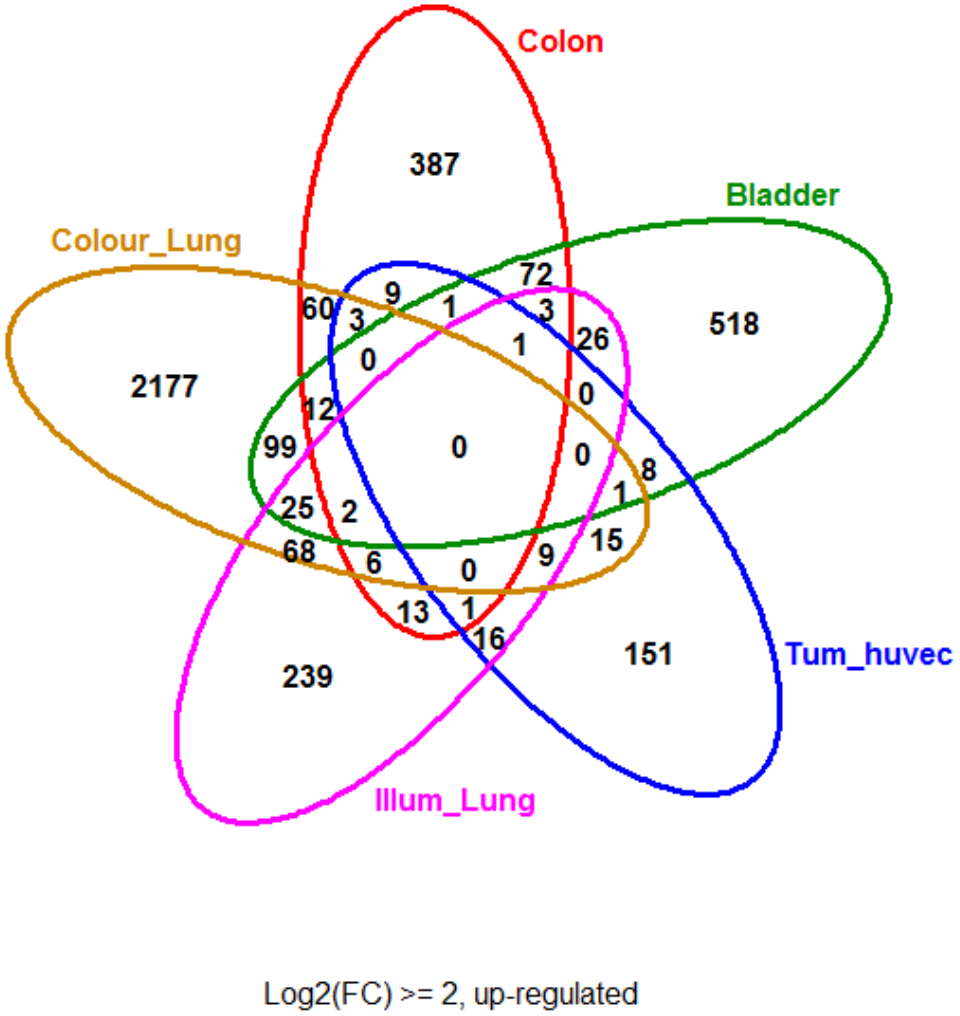


Figure 4.11.1. Pan-Tumour endothelial enriched or pan-VT genes were sought that had log2 fold change of two or more up-regulation in tumour versus normal in multiple contrasts. Each coloured circle represents one differential gene expression contrast. A total of 3,922 genes were found from all contrasts having log2(FC) ≥ 2. The number of genes up-regulated in more than one contrast is as follows: 386 genes were found up-regulated in two contrasts, 61 genes in three contrasts and 3 genes in four contrasts. There were no genes found log2(FC) ≥ 2 for all five contrasts and 3,472 genes were enriched in one contrast only.

Figure 4.11.2 Venn diagram showing overlaps between four differential gene expression contrasts (lung contrasts combined)

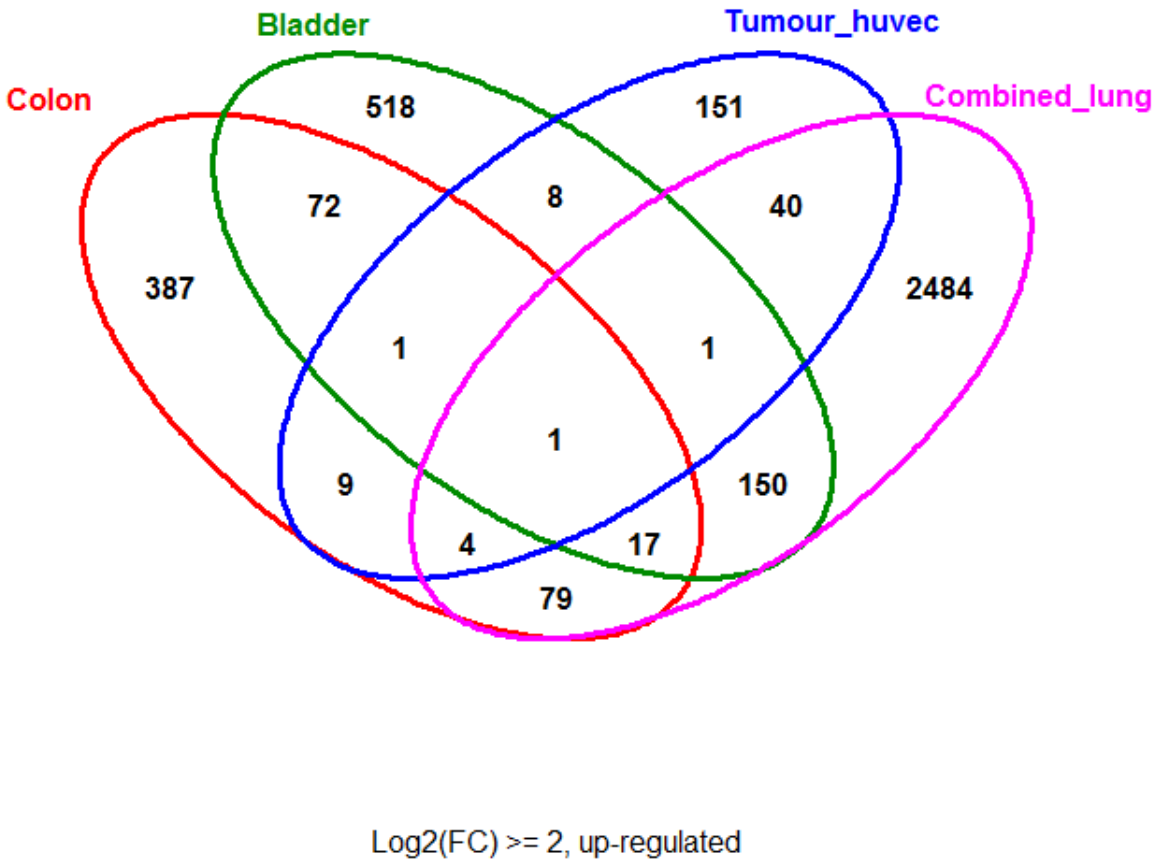


Figure 4.11.2. This figure shows the overlapping up-regulated genes with log2(FC) ≥ 2 from four tissue contrasts. In this figure, the two lung differential gene expression experiments were combined to give a single result (the more up-regulated of the two). Each coloured circle represents one differential gene expression contrast. Endothelial Specific Molecule 1 (ESM1) is the only gene found in all four tissue contrasts and is shown to be a pan-VT of all tissue types/organs used in this work. 23 genes were up-regulated in three contrasts, 358 in two contrasts and 3,540 genes in a single contrast.

Figure 4.11.3 ESM1 pan vascular target up-regulated in tumour endothelium of lung, bladder, colon and tumour Huvec

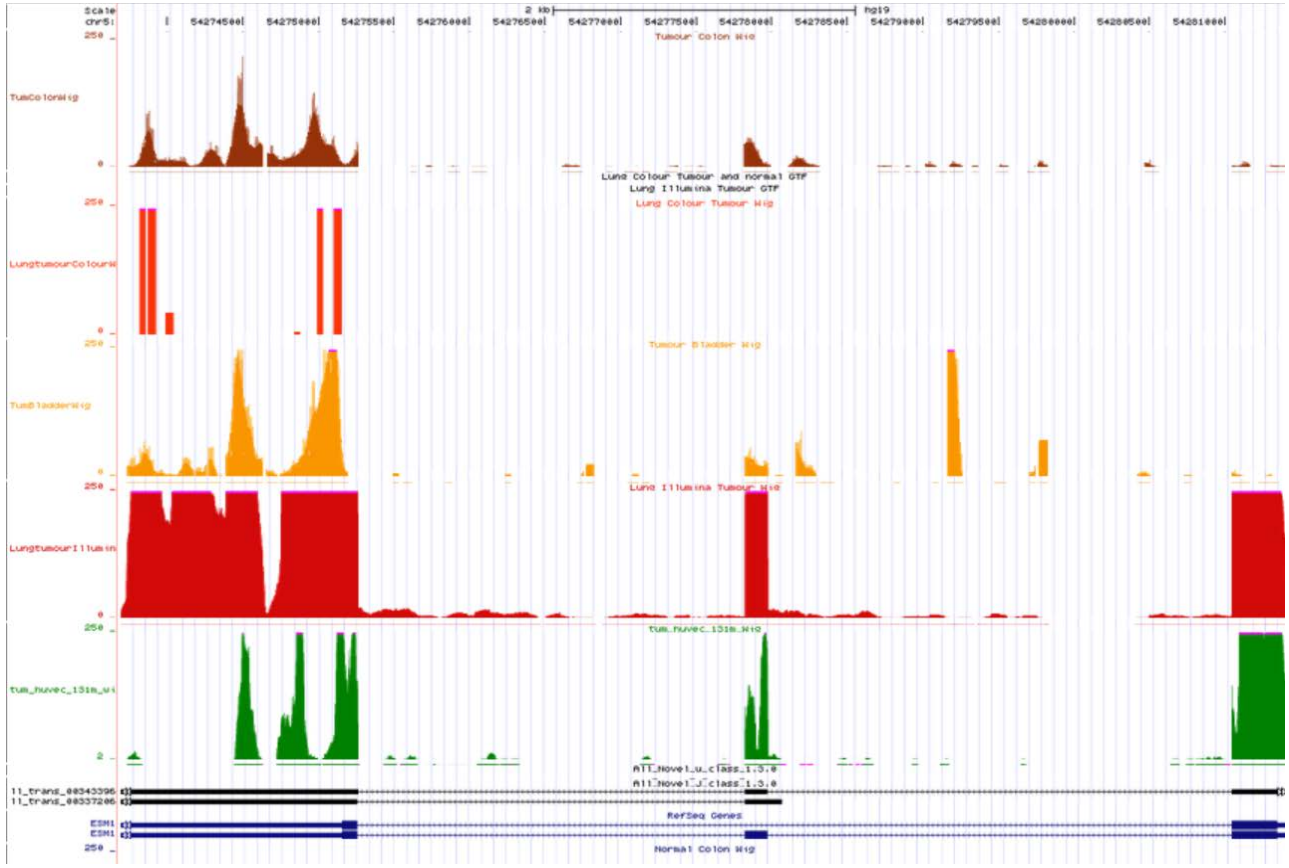


Figure 4.11.3; This shows a genomic overview of ESM1 and a histogram of reads mapping to the ESM1 exons. It is evident that this gene is highly expressed in tumour endothelium of lung, colon and bladder organs, as well as the tumour conditioned Huvec.

Figure 4.11.4 ESM1 pan vascular target very little expression in normal endothelium of lung and colon

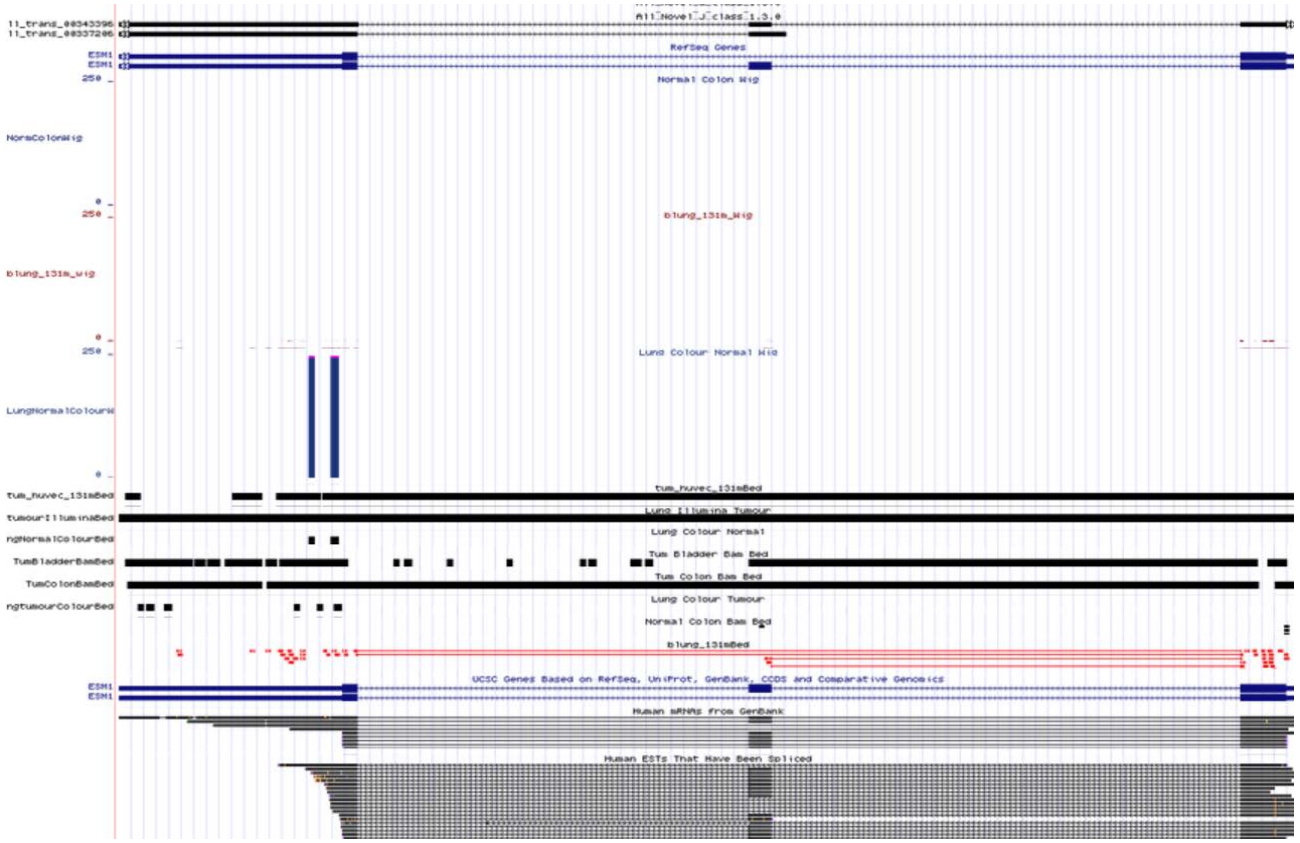


Figure 4.11.4; In contrast to tumour endothelium, ESM1 expression is almost absent from the normal colon and lung samples.

Table 4.11.1 Pan-vascular targets; membrane domain containing genes that were found enriched in two or more RNA-seq contrasts

No. of contrasts	Gene	Product	Angio log2(FC)	Endo log2(FC)	Class
2	MYCT1	myc target 1	4.97	8.47	M
3	FJX1	four jointed box 1 (Drosophila)	2.04	-0.43	M
3	TNFRSF10D	tumor necrosis factor receptor superfamily, member 10d, decoy with truncated death domain	3.17	-1.87	M
2	GP1BB	glycoprotein Ib (platelet), beta polypeptide	7.83		M
2	GLCE	glucuronic acid epimerase	0.61	6.73	M
2	SELE	selectin E	0.02	6.08	M
2	DLL4	delta-like 4 (Drosophila)	0.39	3.7	M
2	INSR	insulin receptor	-0.57	4.3	M
2	DTNA	dystrobrevin, alpha	0.88	2.72	M
2	GPR143	G protein-coupled receptor 143	0.84	2.31	M
2	VMP1	vacuole membrane protein 1	0	2.64	M
2	TMEM41A	transmembrane protein 41A	0.29	1.98	M
2	FCGR2A	Fc fragment of IgG, low affinity IIa, receptor (CD32)	0.25	1.75	M
2	CLDN3	claudin 3	1.3		M
2	PAQR7	progestin and adipoQ receptor family member VII	-0.01	1.23	M
2	GPX8	glutathione peroxidase 8 (putative)	-0.94	2.04	M
2	FAM171A1	family with sequence similarity 171, member A1	-0.7	1.75	M
2	UPK1B	uroplakin 1B	-0.8	1.55	M
2	ADAM12	ADAM metallopeptidase domain 12	1.82	-1.71	M
2	MAL2	mal, T-cell differentiation protein 2 (gene/pseudogene)	1.49	-3.61	M
2	GPRC5A	G protein-coupled receptor, family C, group 5, member A	1.07	-3.88	M
2	MS4A1	membrane-spanning 4-domains, subfamily A, member 1	1.26	-6.1	M
3	DARC	Duffy blood group, chemokine receptor	-0.03		M
3	BIRC6	baculoviral IAP repeat containing 6	-0.18	-0.21	M
3	S1PR3	sphingosine-1-phosphate receptor 3	0.76	-1.66	M
3	BAMBI	BMP and activin membrane-bound inhibitor homolog (Xenopus laevis)	0.29	-1.68	M
3	CD200	CD200 molecule	0.29	-2.64	M

3	ERRFI1	ERBB receptor feedback inhibitor 1	-1.47	-2.06	M
3	GALNT14	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 14 (GalNAc-T14)	-0.89	-3.04	M
3	CLDN1	claudin 1	-0.78	-8.23	M
3	TMEM27	transmembrane protein 27	-5.09	-7.59	M
2	GRIN2D	glutamate receptor, ionotropic, N-methyl D-aspartate 2D	0.02	0.82	M
2	CXorf61	chromosome X open reading frame 61		0.82	M
2	NOX5	NADPH oxidase, EF-hand calcium binding domain 5	0.79		M
2	PROM2	prominin 2	-0.22	0.82	M
2	F2R	coagulation factor II (thrombin) receptor	-0.16	0.65	M
2	SLC4A11	solute carrier family 4, sodium borate transporter, member 11	-0.44	0.82	M
2	CXCL9	chemokine (C-X-C motif) ligand 9	0.2		M
2	PROCR	protein C receptor, endothelial	0.4	-0.36	M
2	GPR157	G protein-coupled receptor 157			M
2	TMEM116	transmembrane protein 116	-0.99	0.82	M
2	TSPAN6	tetraspanin 6	-0.49	0.21	M
2	SLC16A1	solute carrier family 16, member 1 (monocarboxylic acid transporter 1)	-0.01	-0.55	M
2	MPZL3	myelin protein zero-like 3	-0.7		M
2	SLC39A4	solute carrier family 39 (zinc transporter), member 4	-1.08		M
2	FOLH1	folate hydrolase (prostate-specific membrane antigen) 1	-1.18		M
2	C20orf24	chromosome 20 open reading frame 24	0.35	-1.78	M
2	CXCR7	chemokine (C-X-C motif) receptor 7	-0.04	-1.68	M
2	CLDN10	claudin 10	-1.44	-0.42	M
2	F2RL1	coagulation factor II (thrombin) receptor-like 1	-0.17	-1.8	M
2	MUC20	mucin 20, cell surface associated	0.26	-2.24	M
2	TPBG	trophoblast glycoprotein	-0.05	-2.42	M
2	ADRA2B	adrenergic, alpha-2B-, receptor	-2.6		M
2	SDC4	syndecan 4	-1.65	-0.96	M
2	GDPD5	glycerophosphodiester phosphodiesterase domain containing 5	-0.25	-2.64	M
2	KCTD11	potassium channel tetramerisation domain containing 11	-0.44	-2.64	M
2	ABCG2	ATP-binding cassette, sub-family G (WHITE), member 2	0.39	-3.61	M
2	SLC6A8	solute carrier family 6 (neurotransmitter transporter, creatine), member 8	0.26	-3.61	M

2	IER3	immediate early response 3	-0.19	-3.37	M
2	DNAJA4	DnaJ (Hsp40) homolog, subfamily A, member 4	-1.31	-3.57	M
2	RASL11B	RAS-like, family 11, member B	0.61	-7.03	M
2	CEACAM6	carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross reacting antigen)	0.97	-7.59	M
2	ITGB6	integrin, beta 6	0.79	-7.59	M
2	C3orf52	chromosome 3 open reading frame 52	-0.29	-7.03	M
2	CLCA2	chloride channel accessory 2	-1.02	-7.03	M
2	MUC1	mucin 1, cell surface associated	0.72	-9.56	M
2	CLDN4	claudin 4	-0.92	-9.3	M
2	TACSTD2	tumor-associated calcium signal transducer 2	-0.83	-9.67	M
2	RNF128	ring finger protein 128	-0.36	-11.21	M

Table 4.11.1; There were 69 genes predicted to have transmembrane domains that were enriched in tumour endothelium from two or more contrasts.

Table 4.11.2 Pan-vascular targets; secreted/extracellular genes that were found enriched in two or more RNA-seq contrasts

No. of contrasts	Gene	Product	Angio log2(FC)	Endo log2(FC)	Class
3	MMP1	matrix metallopeptidase 1 (interstitial collagenase)	1.38	12.12	E
3	ANGPT2	angiopoietin 2	2.32	10.68	E
3	MMP10	matrix metallopeptidase 10 (stromelysin 2)	1.14	7.45	E
2	ADAMTS6	ADAM metallopeptidase with thrombospondin type 1 motif, 6	6.22	5.01	E
2	ADAMTS18	ADAM metallopeptidase with thrombospondin type 1 motif, 18	6.22	4.5	E
2	PTX3	pentraxin 3, long	6.33	1.98	E
2	CCL20	chemokine (C-C motif) ligand 20	3.4	4.15	E
2	COL15A1	collagen, type XV, alpha 1	1.2	4.15	E
4	**ESM1**	endothelial cell-specific molecule 1	0.26	8.35	E
3	COL10A1	collagen, type X, alpha 1	4.97		E
3	SAA2-SAA4	SAA2-SAA4 readthrough	2.76		E
3	COL11A1	collagen, type XI, alpha 1	1.48		E
3	LOXL2	lysyl oxidase-like 2	2.15	-1.8	E
3	PGF	placental growth factor	3.08	-3.57	E
3	MMP11	matrix metallopeptidase 11 (stromelysin 3)	2.66	-7.03	E
2	MMRN1	multimerin 1	0.97	9.54	E
2	APLN	apelin	0.78	7.75	E
2	CST1	cystatin SN	7.43		E
2	CRLF1	cytokine receptor-like factor 1	6.87		E
2	DKK2	dickkopf homolog 2 (Xenopus laevis)	4.97		E
2	TNF	tumor necrosis factor	3.69		E
2	KLK12	kallikrein-related peptidase 12	3.69		E
2	IL1B	interleukin 1, beta	5.45	-2.64	E
2	INS-IGF2	INS-IGF2 readthrough	2.76		E
2	S100A8	S100 calcium binding protein A8	-0.26	2.84	E
2	EREG	epiregulin	-1.23	3.29	E
2	FAM3B	family with sequence similarity 3, member B	1.94		E
2	S100A11	S100 calcium binding protein A11	0.41	1.47	E
2	SERPINE1	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1	-1.05	2.83	E
2	LYPD6B	LY6/PLAUR domain containing 6B	1.77		E

2	ULBP2	UL16 binding protein 2	1.51	0.05	E
2	C2CD4A	C2 calcium-dependent domain containing 4A	1.41	-0.17	E
2	SFRP2	secreted frizzled-related protein 2	1.14		E
2	CRTAC1	cartilage acidic protein 1	1.12		E
2	MMP2	matrix metallopeptidase 2 (gelatinase A, 72kDa gelatinase, 72kDa type IV collagenase)	1.41	-1.02	E
2	F2RL2	coagulation factor II (thrombin) receptor-like 2	-2.45	1.75	E
2	IGF2	insulin-like growth factor 2 (somatomedin A)	2.86	-3.92	E
2	PLXDC1	plexin domain containing 1	1.79	-3.2	E
2	PRG2	proteoglycan 2, bone marrow (natural killer cell activator, eosinophil granule major basic protein)	4.65	-7.59	E
2	EGFL6	EGF-like-domain, multiple 6	2.98	-7.03	E
2	INHBB	inhibin, beta B	1.7	-6.1	E
2	NOV	nephroblastoma overexpressed gene	1.44	-8.31	E
2	COL1A1	collagen, type I, alpha 1	1.71	-9.18	E
2	MMP12	matrix metallopeptidase 12 (macrophage elastase)	2.63	-11.05	E
4	IL8	interleukin 8	-0.23	-1.71	E
3	HSPH1	heat shock 105kDa/110kDa protein 1	-0.36	0.23	E
3	SERPINE2	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 2	0.53	-2.03	E
3	BPIFA1	BPI fold containing family A, member 1	-1.53		E
3	C9orf47	chromosome 9 open reading frame 47	-2.26	-6.1	E
3	GDF15	growth differentiation factor 15	0.38	-9.3	E
3	SPP1	secreted phosphoprotein 1	-1.19	-9.16	E
2	DEFB1	defensin, beta 1	0.81	0.82	E
2	CEACAM5	carcinoembryonic antigen-related cell adhesion molecule 5	0.45	0.82	E
2	IGFBP7	insulin-like growth factor binding protein 7	0.64	0.17	E
2	NMB	neuromedin B	0.61		E
2	EFNA1	ephrin-A1	0.27	0.26	E
2	SFRP4	secreted frizzled-related protein 4	0.15		E
2	PRSS22	protease, serine, 22	0.13		E
2	INS	insulin			E
2	MAGEA9B	melanoma antigen family A, 9B			E
2	TGIF2-C20ORF24	TGIF2-C20orf24 readthrough			E
2	CILP2	cartilage intermediate layer protein 2			E

2	HTRA3	HtrA serine peptidase 3	-0.06		E
2	AGT	angiotensinogen (serpin peptidase inhibitor, clade A, member 8)	-0.11		E
2	BOLA1	bolA homolog 1 (E. coli)	-0.13		E
2	HTRA1	HtrA serine peptidase 1	0.27	-0.51	E
2	ANGPTL4	angiopoietin-like 4	0.38	-0.74	E
2	IL23A	interleukin 23, alpha subunit p19	-0.44		E
2	CXCL17	chemokine (C-X-C motif) ligand 17	-0.7		E
2	PPIA	peptidylprolyl isomerase A (cyclophilin A)	0.12	-0.92	E
2	INHBA	inhibin, beta A	0.29	-1.23	E
2	HSPD1	heat shock 60kDa protein 1 (chaperonin)	0.84	-1.9	E
2	HSPG2	heparan sulfate proteoglycan 2	0.15	-1.28	E
2	IL11	interleukin 11	-0.13	-1.36	E
2	ADM	adrenomedullin	-0.37	-1.3	E
2	BAIAP3	BAI1-associated protein 3	-1.69		E
2	STC1	stanniocalcin 1	-1.82	-0.08	E
2	PDGFD	platelet derived growth factor D	-0.23	-2.24	E
2	C1orf63	chromosome 1 open reading frame 63	0.03	-2.64	E
2	IL1RN	interleukin 1 receptor antagonist	-0.06	-2.64	E
2	S100A9	S100 calcium binding protein A9	0.8	-6.1	E
2	IGFBP3	insulin-like growth factor binding protein 3	0.31	-6.13	E
2	ACP6	acid phosphatase 6, lysophosphatidic	-0.17	-6.1	E
2	EPOR	erythropoietin receptor	0.68	-7.03	E
2	VWA1	von Willebrand factor A domain containing 1	0.65	-7.03	E
2	SAA2	serum amyloid A2	-0.62	-6.1	E
2	EGFLAM	EGF-like, fibronectin type III and laminin G domains	-0.63	-6.1	E
2	ANXA8	annexin A8	-1.26	-6.1	E
2	SPESP1	sperm equatorial segment protein 1	0.42	-8	E
2	S100A2	S100 calcium binding protein A2	-0.91	-7.03	E
2	CXCL3	chemokine (C-X-C motif) ligand 3	-1.85	-6.1	E
2	RNF215	ring finger protein 215	-2.24	-6.1	E
2	GDF3	growth differentiation factor 3	-5.66	-3.04	E
2	SFN	stratifin	0.67	-9.56	E
2	LTF	lactotransferrin	-0.37	-8.57	E
2	SAA1	serum amyloid A1	-3	-6.1	E
2	SAA4	serum amyloid A4, constitutive	-2.36	-7.03	E
2	AZGP1	alpha-2-glycoprotein 1, zinc-binding	-1.58	-8.79	E
2	SPOCK3	sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 3	-4.38	-7.59	E

Table 4.11.2; 99 extracellular/secreted protein genes were significantly enriched in two or more contrasts, which included ESM1 that was enriched in all four organ tumours tested. It also included genes that had been previously reported as tumour angiogenesis markers: ANGPT2, IGFBP7, IL8, PLXDC1, EREG, MMP1 and MMP2.

4.12 pan-Vascular Targets Biological Process functional enrichment

To investigate any functional representational groups within the 450 genes that were found enriched in tumour endothelium in two or more contrasts, the Database for Annotation, Visualization and Integrated Discovery (DAVID, [129]) was used, selecting Gene Ontology, Biological Process. This identified biological processes such as blood vessel/vascular development, angiogenesis, response to wounding, regulation of cellular proliferation, ectoderm/epidermis development and regulation of cytokine production. Table 4.12.1 has the full list of the 26 enriched functional categories, with an FDR $\leq$ 2%.

Table 4.12.1 Functional enrichment of 450 genes, tumour endothelial enriched for two or more contrasts

GO Biological Process Term	Genes	Count	%	P-Value	Benjamini FDR
response to wounding	DARC, ELF3, S100A8, S100A9, ADM, ANXA8, CCL20, CXCL3, CXCL9, F2R, F2RL1, F2RL2, CCNB1, CYP1A1, EREG, GP1BB, IGF2, INS-IGF2, ITGB6, IL1RN, IL1B, IL11, , IL8, MAP1B, MMRN1, PTX3, PROCR, SPP1, SELE, SERPINE1, SAA1, SAA2, SAA4, SPRR3, S1PR3, TNF	36	8.4	1.40E-08	3.10E-05
ectoderm development	ELF3, SOX9, COL1A1, EVPL, EREG, FOXQ1, IRF6, KRT13, KRT14, KRT15, KRT16, KRT17, KRT5, KRT6A, LEF1, OVOL1, PTGS2, SPRR1A, SPRR3, SFN	20	4.7	1.40E-07	1.50E-04
epidermis development	ELF3, SOX9, COL1A1, EVPL, EREG, FOXQ1, IRF6, KRT13, KRT14, KRT15, KRT16, KRT17, KRT5, LEF1, OVOL1, PTGS2, SPRR1A, SPRR3, SFN	19	4.4	2.10E-07	1.50E-04
regulation of cell proliferation	BNIPL, CEBPA, KLF5, NKX3-1, S100A11, SERTAD1, SOX9, ADM, AZGP1, AGT, BIRC6, F2R, CDKN2D, CDKN3, DHRS2, DLX5, EREG, GAS8, HES1, HHEX, INSR, IGFBP3, IGFBP7, IRF6, IL1B, IL11, IL8, KRT5, KRT6A, MCTS1, MMP12, PNP, PPARG, PGF, KCTD11, PTGS2, SCIN, SERPINE1, S1PR3, SFN, TNF, MYCN	43	10	2.10E-07	1.20E-04
response to vitamin	CDKN2D, CYP1A1, CYP24A1, INSR, IGFBP7, IL1B, MEST, MAP1B, MUC1, PPARG, PTGS2, SPP1	12	2.8	2.50E-07	1.10E-04

response to nutrient	UGT1A9, UGT1A10, UGT1A4, UGT1A6, UGT1A8, UGT1A7, UGT1A3, UGT1A1, UGT1A5, COL1A1, CDKN2D, CYP1A1, CYP24A1, INSR, IGFBP7, IL1B, MEST, MAP1B, MUC1, PPARG, PTGS2, SPP1, STC1	15	3.5	3.70E-06	1.30E-03
negative regulation of cell proliferation	BNIPL, CEBPA, NKX3-1, S100A11, ADM, AZGP1, AGT, F2R, CDKN2D, CDKN3, DHRS2, EREG, GAS8, IGFBP3, IGFBP7, IRF6, IL1B, IL8, KRT5, PPARG, PTGS2, SCIN, SFN, TNF	24	5.6	8.80E-06	2.70E-03
inflammatory response	DARC, ELF3, S100A8, S100A9, CCL20, CXCL3, CXCL9, F2R, ITGB6, IL1RN, IL1B, IL23A, IL8, PTX3, SPP1, SELE, SAA1, SAA2, SAA4, S1PR3, TNF	22	5.1	1.80E-05	4.80E-03
response to vitamin D	CDKN2D, CYP24A1, INSR, IL1B, PTGS2, SPP1	6	1.4	2.10E-05	5.10E-03
defense response	DARC, ELF3, S100A8, S100A9, CCL20, CXCL3, CXCL9, F2R, CCNL2, DEFB1, HIST1H2BG, HIST2H2BE, INHBA, INHBB, ITGB6, IL1RN, IL1B, IL23A, IL8, LTF, BPIFA1, PTX3, PPARG, PRG2, SPP1, SELE, SAA1, SAA2, SAA4, S1PR3, TNF	32	7.5	2.90E-05	6.20E-03
wound healing	ANXA8, F2R, F2RL1, F2RL2, CCNB1, EREG, GP1BB, IGF2, IL1B, IL11, MMRN1, PROCR, SERPINE1, SAA1, SAA2, SPRR3	16	3.7	3.20E-05	6.20E-03
regulation of growth	BNIPL, HTRA1, HTRA3, SERTAD1, TSPYL2, AGT, CDKN2D, ESM1, FOXS1, GDF3, INHBA, INSR, IGF2, IGFBP3, IGFBP7, MCTS1, MAP1B, NOV, PPARG, KCTD11, SPP1, SFN	22	5.1	3.60E-05	6.50E-03

regulation of cell growth	HTRA1, HTRA3, SERTAD1, TSPYL2, AGT, CDKN2D, ESM1, INHBA, IGF2, IGFBP3, IGFBP7, MAP1B, NOV, PPARG, SPP1, SFN	16	3.7	3.80E-05	6.30E-03
response to extracellular stimulus	UGT1A9, ADM, COL1A1, CDKN2D, CYP1A1, CYP24A1, INSR, IGFBP7, IL1B, MEST, MAP1B, MUC1, PPARG, PTGS2, SFRP2, SPP1, STC1	17	4	4.40E-05	6.90E-03
response to nutrient levels	UGT1A9, ADM, COL1A1, CDKN2D, CYP1A1, CYP24A1, INSR, IGFBP7, IL1B, MEST, MAP1B, MUC1, PPARG, PTGS2, SPP1, STC1	16	3.7	4.50E-05	6.50E-03
blood vessel development	KLF5, LMO2, ADRA2B, ANGPT2, ANGPTL4, AGT, ANXA2P3, CXCL17, COL1A1, COL15A1, DLL4, EREG, IL1B, IL8, JMJD6, MMP2, PGF, PLXDC1	18	4.2	4.70E-05	6.30E-03
vasculature development	KLF5, LMO2, ADRA2B, ANGPT2, ANGPTL4, AGT, ANXA2P3, CXCL17, COL1A1, COL15A1, DLL4, EREG, IL1B, IL8, JMJD6, MMP2, PGF, PLXDC1	18	4.2	6.30E-05	8.00E-03
regulation of cytokine production	AGT, F2R, EREG, HSPD1, INHBA, INHBB, IGF2, IGF2BP3, IL1B, PPARG, RNF128, SAA1, SAA2, S1PR3, TNF	15	3.5	6.90E-05	8.30E-03
epithelial cell differentiation	ELF3, AGT, DLX5, EVPL, EREG, IRF6, KRT14, KRT5, PPARG, SPRR1A, SPRR3, SFN, UPK1B	13	3	6.90E-05	7.80E-03
regulation of response to external stimulus	AGT, ANXA2P3, F2R, F2RL1, GPX2, IGF2, IL8, PPARG, PTGS2, SPP1, SELE, SERPINE1, SAA1, SAA2	14	3.3	7.00E-05	7.60E-03
blood coagulation	ANXA8, F2R, F2RL1, F2RL2, GP1BB, IL11, MMRN1, PROCR, SERPINE1, SAA1, SAA2	11	2.6	1.10E-04	1.10E-02
Coagulation	ANXA8, F2R, F2RL1, F2RL2, GP1BB, IL11, MMRN1, PROCR, SERPINE1, SAA1, SAA2	11	2.6	1.10E-04	1.10E-02

bone development	SOX9, COL1A1, CYP24A1, FHL2, HSPG2, IGF2, IGFBP3, JUND, MMP2, PTGS2, SPP1, STC1	12	2.8	1.10E-04	1.10E-02
Angiogenesis	KLF5, ADRA2B, ANGPT2, ANGPTL4, ANXA2P3, CXCL17, COL15A1, DLL4, EREG, IL1B, IL8, PGF, PLXDC1	13	3	1.40E-04	1.40E-02
Hemostasis	ANXA8, F2R, F2RL1, F2RL2, GP1BB, IL11, MMRN1, PROCR, SERPINE1, SAA1, SAA2	11	2.6	1.70E-04	1.60E-02
regulation of cell cycle	CKS2, H2AFX, SERTAD1, TSPYL2, ANLN, CDK1, CCNB1, CDKN2D, CDKN3, CYP1A1, EREG, INHBA, INSR, IGF2, IL1B, PIM3, PTGS2, SFN, TNF, UBE2C	20	4.7	2.10E-04	1.80E-02

Table 4.12.1; This table lists the top 26 GO Biological Processes, functional enrichment categories found to be enriched in the 450 genes found up-regulated in tumour endothelium with greater or equal to two contrasts. Functional enrichment was seen for processes relating to tumour/tissue growth and tumour angiogenesis.

4.13 Genes with published links to tumour angiogenesis

To test for known involvement in tumour and/or angiogenesis, an AngioScore term was derived by keyword matching in abstract searches of the 3,922 genes found (the same as performed in the cDNA library chapter). The full list of results can be found in appendix table A.4.13.1 and the top 39 genes with the highest AngioScore of ≥ 60 are shown in table 4.13.1.

Table 4.13.1 39 RNA-seq contrast genes with the highest AngioScore

Gene	Number of contrasts	Product	Class	Angio Score
DLEU2L	1	deleted in lymphocytic leukemia 2-like	intracellular	**100**
LOC613037	1	nuclear pore complex interacting protein pseudogene	intracellular	**100**
CHEK2P2	1	checkpoint kinase 2 pseudogene 2	intracellular	**100**
VEGFA	1	vascular endothelial growth factor A	extracellular	**95**
VEGFC	1	vascular endothelial growth factor C	extracellular	**93**
HIF1A	1	hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)	intracellular	**90**
EDN1	1	endothelin 1	extracellular	**87**
PROCR	2	protein C receptor, endothelial	membrane	**80**
HIC1	2	hypermethylated in cancer 1	intracellular	**79**
VEGFB	1	vascular endothelial growth factor B	extracellular	**76**
IRAK1BP1	1	interleukin-1 receptor-associated kinase 1 binding protein 1	intracellular	**75**
NOS3	1	nitric oxide synthase 3 (endothelial cell)	intracellular	**74**
CA9	1	carbonic anhydrase IX	intracellular	**74**
VASH1	1	vasohibin 1	extracellular	**73**
COL18A1	1	collagen, type XVIII, alpha 1	extracellular	**70**
VCAM1	1	vascular cell adhesion molecule 1	membrane	**70**
CD248	1	CD248 molecule, endosialin	membrane	**70**
PTTG1	1	pituitary tumor-transforming 1	intracellular	**69**
PGF	3	placental growth factor	extracellular	**68**
ESM1	4	endothelial cell-specific molecule 1	extracellular	**68**
ANGPT2	3	angiopoietin 2	extracellular	**66**
PDPN	1	podoplanin	membrane	**66**
MCAM	1	melanoma cell adhesion molecule	membrane	**65**
LIPG	1	lipase, endothelial	extracellular	**65**
TNFRSF12A	1	tumor necrosis factor receptor superfamily, member 12A	membrane	**65**
MIR210	1	microRNA 210	intracellular	**65**
PTK6	1	PTK6 protein tyrosine kinase 6	intracellular	**64**
NUAK1	1	NUAK family, SNF1-like kinase, 1	intracellular	**64**
IGFBP7	2	insulin-like growth factor binding protein 7	extracellular	**63**
TRAF5	1	TNF receptor-associated factor 5	intracellular	**63**
MAGEA4	1	melanoma antigen family A, 4	extracellular	**63**
MIR126	1	microRNA 126	intracellular	**63**
MEN1	1	multiple endocrine neoplasia I	intracellular	**62**

TNF	2	tumor necrosis factor	extracellular	**60**
EGFL7	1	EGF-like-domain, multiple 7	extracellular	**60**
MAGEA2	1	melanoma antigen family A, 2	extracellular	**60**
CCDC34	1	coiled-coil domain containing 34	intracellular	**60**
MIR650	1	microRNA 650	intracellular	**60**
DLL4	2	delta-like 4 (Drosophila)	membrane	**60**

Table 4.13.1; This table ranks RNA-seq enriched genes based on the AngioScore. 39 genes were found to have an AngioScore ≥ 60.

4.14 RNA-seq, ranking by a cDNA endothelial screen

RNA-seq tumour endothelial enriched genes were cross checked with endothelial screen results from the cDNA library analyses. The aim of this was to best target the genes that are most likely expressed by endothelial cells. The results can be seen in table 4.14.1, which shows 35 RNA-seq endothelial enriched genes based on cDNA library analyses results of log2(FC) ≥ 6 enriched in endothelial cells. Notably, 9 of these genes were also found in more than one RNA-seq contrast.

Table 4.14.1 Ranking RNA-seq genes with a cDNA endothelial cell expression evidence

Gene	Number of contrasts	Product	Endothelial log2(FC)
MMP1	3	matrix metallopeptidase 1 (interstitial collagenase)	12.12
ANGPT2	3	angiopoietin 2	10.68
CLDN11	1	claudin 11	10.65
FLJ41200	1	hypothetical LOC401492	9.89
MMRN1	2	multimerin 1	9.54
HHIP	1	hedgehog interacting protein	9.22
LINC00152	1	long intergenic non-protein coding RNA 152	9.21
MYCT1	2	myc target 1	8.47
ESM1	**4**	**endothelial cell-specific molecule 1**	8.35
ATP5H	1	ATP synthase, H+ transporting, mitochondrial Fo complex, subunit d	8.31
ELTD1	1	EGF, latrophilin and seven transmembrane domain containing 1	7.99
APLN	2	Apelin	7.75
RHOJ	1	ras homolog gene family, member J	7.74
PLK1S1	1	polo-like kinase 1 substrate 1	7.65
B3GALTL	1	beta 1,3-galactosyltransferase-like	7.57
MMP10	3	matrix metallopeptidase 10 (stromelysin 2)	7.45
NUDT19	1	nudix (nucleoside diphosphate linked moiety X)-type motif 19	7.4
GIPC2	1	GIPC PDZ domain containing family, member 2	7.19
C3orf58	1	chromosome 3 open reading frame 58	7.13
LPAR6	1	lysophosphatidic acid receptor 6	7.04
BMX	1	BMX non-receptor tyrosine kinase	7.02
GALNTL2	1	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase-like 2	7.01
LRRC70	1	leucine rich repeat containing 70	6.93
CRIM1	1	cysteine rich transmembrane BMP regulator 1 (chordin-like)	6.87
GLCE	2	glucuronic acid epimerase	6.73
MMRN2	1	multimerin 2	6.71
ZNF521	1	zinc finger protein 521	6.61
LAMP3	1	lysosomal-associated membrane protein 3	6.57
TMEM59L	1	transmembrane protein 59-like	6.52
CLEC14A	1	C-type lectin domain family 14, member A	6.18
SELE	2	selectin E	6.08
MIER1	1	mesoderm induction early response 1 homolog	6.08

		(Xenopus laevis)	
TM6SF1	1	transmembrane 6 superfamily member 1	6.08
LOC100507254	1	hypothetical LOC100507254	6.04
TNFSF4	1	tumor necrosis factor (ligand) superfamily, member 4	6.01

Table 4.14.1; 35 endothelial enriched genes based on cDNA library analyses results of log2(FC) ≥ 6 enriched in endothelial cells. 9 of these genes were also found in more than one RNA-seq contrast.

4.15 RNA-seq, ranking by a cDNA angiogenic screen

RNA-seq tumour endothelial enriched genes were cross checked with foetal/tumour/angiogenic screen results from the cDNA library analyses. This screen aimed to dissect out any contaminating normal tissue expressed gene passengers found in the RNA-seq results. The results of this can be seen in table 4.15.1, which shows 45 foetal/tumour/angiogenic enriched genes based on cDNA library analyses results of log2(FC) >= 5 enriched in foetal/tumour/angiogenic bulk tissues versus normal bulk tissues. 8 of these genes were also found in more than one RNA-seq contrast.

Table 4.15.1 Ranking RNA-seq genes with a cDNA angiogenic/tumour expression evidence

Gene	No. of RNA-seq contrasts	Product	Angio log2(FC)
HBG1	1	hemoglobin, gamma A	8.21
GP1BB	2	glycoprotein Ib (platelet), beta polypeptide	7.83
TUBB2B	1	tubulin, beta 2B	7.58
MMP9	1	matrix metallopeptidase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase)	7.48
CST1	2	cystatin SN	7.43
SLC35D3	1	solute carrier family 35, member D3	7.21
ETV4	1	ets variant 4	7.15
FCGR1A	1	Fc fragment of IgG, high affinity Ia, receptor (CD64)	7.09
SEMA5B	1	sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5B	7.09
TROAP	1	trophinin associated protein (tastin)	6.87
CRLF1	2	cytokine receptor-like factor 1	6.87
AURKB	1	aurora kinase B	6.72
ZCCHC18	1	zinc finger, CCHC domain containing 18	6.54
FAM83D	2	family with sequence similarity 83, member D	6.44
LDLRAD3	1	low density lipoprotein receptor class A domain containing 3	6.44
PTX3	2	pentraxin 3, long	6.33
ADAMTS6	2	ADAM metallopeptidase with thrombospondin type 1 motif, 6	6.22
ADAMTS18	2	ADAM metallopeptidase with thrombospondin type 1 motif, 18	6.22
KIF20A	1	kinesin family member 20A	6.22
IVL	1	Involucrin	6.22
HOXB2	1	homeobox B2	5.96
FABP6	1	fatty acid binding protein 6, ileal	5.96
ZNF134	1	zinc finger protein 134	5.81
DLGAP5	1	discs, large (Drosophila) homolog-associated protein 5	5.81
CECR2	1	cat eye syndrome chromosome region, candidate 2	5.81
PCDHB4	1	protocadherin beta 4	5.81
DLX2	1	distal-less homeobox 2	5.81
LHX6	1	LIM homeobox 6	5.64
GPR84	1	G protein-coupled receptor 84	5.64
PCDHB10	1	protocadherin beta 10	5.64

ZMYND19	1	zinc finger, MYND-type containing 19	5.64
MIR612	1	microRNA 612	5.64
MARCH4	1	membrane-associated ring finger (C3HC4) 4	5.45
IL1B	2	interleukin 1, beta	5.45
BMP8B	1	bone morphogenetic protein 8b	5.45
NCKAP5L	1	NCK-associated protein 5-like	5.45
NHLH2	1	nescient helix loop helix 2	5.45
S100A7	1	S100 calcium binding protein A7	5.45
POU5F1B	1	POU class 5 homeobox 1B	5.45
AKIRIN2-AS1	1	AKIRIN2 antisense RNA 1 (non-protein coding)	5.45
DAPL1	1	death associated protein-like 1	5.45
EPHX4	1	epoxide hydrolase 4	5.45
CDC25A	1	cell division cycle 25 homolog A (S. pombe)	5.23
TLX1	1	T-cell leukemia homeobox 1	5.23
PI3	1	peptidase inhibitor 3, skin-derived	5.12

Table 4.15.1; Ranking RNA-seq 2nd generation sequence data using angiogenic/tumour cDNA screen. 45 genes with an angiogenic screen results of log2(FC) ≥ 5 enriched in foetal/angiogenic/tumour bulk tissues versus normal bulk tissues were found. 8 of these genes were also found in more than one RNA-seq contrast.

4.16 RNA-seq, ranking by combined endothelial and angiogenic cDNA library screens

In the previous two analyses, the cDNA library screens of endothelial cells versus normal non-endothelial cells and the foetal/tumour/angiogenic bulk tissues versus normal bulk tissues were independently scrutinised to produce rankings of the RNA-seq contrast results. In this third approach, both cDNA libraries screens were combined to produce a tumour endothelial confirmed approach. So not only are genes putatively enriched in freshly isolated tumour endothelium, they are also up-regulated in both endothelial cells and foetal/tumour/angiogenic tissues from the cDNA library analyses. The results of this third incorporation of cDNA libraries can be seen in table 4.16.1 (in both endothelial and foetal cDNA screens, only a minimum log2(FC) ≥ 1 was considered). 126 genes were found using this criterion. 15 of these genes were found to be tumour enriched in more than on RNA-seq contrast. 20 genes are membrane, 16 extracellular and the remaining 90 are intracellular.

Table 4.16.1 Ranking RNA-seq enriched genes with tumour/angiogenesis and endothelial cDNA expression evidence

No. of RNA-seq contrasts	Gene	Product	Angio log2(FC)	Endo log2(FC)	Class
3	MMP1	matrix metallopeptidase 1 (interstitial collagenase)	1.38	12.12	E
3	ANGPT2	angiopoietin 2	2.32	10.68	E
3	MMP10	matrix metallopeptidase 10 (stromelysin 2)	1.14	7.45	E
2	MYCT1	myc target 1	4.97	8.47	M
2	ADAMTS6	ADAM metallopeptidase with thrombospondin type 1 motif, 6	6.22	5.01	E
2	ADAMTS18	ADAM metallopeptidase with thrombospondin type 1 motif, 18	6.22	4.5	E
2	PTX3	pentraxin 3, long	6.33	1.98	E
2	CCL20	chemokine (C-C motif) ligand 20	3.4	4.15	E
2	DIAPH3	diaphanous homolog 3 (Drosophila)	1	5.89	I
2	HIST1H4C	histone cluster 1, H4c	1.99	3.77	I
2	COL15A1	collagen, type XV, alpha 1	1.2	4.15	E
2	SNORA52	small nucleolar RNA, H/ACA box 52	2.76	1.75	I
2	MYCN	v-myc myelocytomatosis viral related oncogene, neuroblastoma derived (avian)	1.86	2.31	I
2	FAM43A	family with sequence similarity 43, member A	1.14	2.56	I
2	C4orf46	chromosome 4 open reading frame 46	1.48	1.62	I
1	FLJ41200	hypothetical LOC401492	1.2	9.89	I
1	LY96	lymphocyte antigen 96	4.25	5.73	E
1	MOV10L1	Mov10l1, Moloney leukemia virus 10-like 1, homolog (mouse)	4.25	4.6	I
1	ZNF600	zinc finger protein 600	4.97	3.87	I
1	LOC100507254	hypothetical LOC100507254	2.76	6.04	I
1	C2orf27A	chromosome 2 open reading frame 27A	4.65	3.7	I
1	GALNTL2	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase-like 2	1.32	7.01	I
1	ZNF134	zinc finger protein 134	5.81	2.31	I
1	TNFSF4	tumor necrosis factor (ligand) superfamily, member 4	1.77	6.01	E

1	TMEM59L	transmembrane protein 59-like	1.22	6.52	M
1	CENPW	centromere protein W	2.18	5.49	I
1	SOX4	SRY (sex determining region Y)-box 4	1.37	5.98	I
1	MARCH4	membrane-associated ring finger (C3HC4) 4	5.45	1.75	I
1	ZNF630	zinc finger protein 630	3.69	3.51	I
1	LHX6	LIM homeobox 6	5.64	1.4	I
1	CENPI	centromere protein I	4.25	2.72	I
1	ZNF267	zinc finger protein 267	2.35	4.5	I
1	ZNF304	zinc finger protein 304	2.76	4.02	I
1	TTC13	tetratricopeptide repeat domain 13	3.87	2.72	M
1	PLA2G10	phospholipase A2, group X	4.65	1.75	E
1	MYCL1	v-myc myelocytomatosis viral oncogene homolog 1, lung carcinoma derived (avian)	4.65	1.75	I
1	CCDC23	coiled-coil domain containing 23	1.36	5.01	I
1	INTS7	integrator complex subunit 7	1.56	4.78	I
1	HRH1	histamine receptor H1	1.05	5.21	M
1	DAPK1	death-associated protein kinase 1	1.91	4.28	I
1	PLXNA4	plexin A4	2.02	4.15	M
1	LOC283143	hypothetical LOC283143	2.18	3.87	I
1	CCNE1	cyclin E1	3.32	2.72	I
1	ZNF610	zinc finger protein 610	1.99	4.02	I
1	CSF2RA	colony stimulating factor 2 receptor, alpha, low-affinity (granulocyte-macrophage)	1.82	4.15	E
1	THAP2	THAP domain containing, apoptosis associated protein 2	1.7	4.15	I
1	RPSAP9	ribosomal protein SA pseudogene 9	1.56	4.28	I
1	ZNRD1	zinc ribbon domain containing 1	1.86	3.87	I
1	C2CD4B	C2 calcium-dependent domain containing 4B	3.98	1.75	I
1	CYP21A1P	cytochrome P450, family 21, subfamily A, polypeptide 1 pseudogene	4.65	1.04	I
1	FRAT2	frequently rearranged in advanced T-cell lymphomas 2	2.4	3.29	I
1	WDR89	WD repeat domain 89	1.08	4.6	I
1	NANP	N-acetylneuraminic acid phosphatase	1.24	4.39	I
1	MGST2	microsomal glutathione S-transferase 2	1.86	3.76	I
1	FAM26E	family with sequence similarity 26, member E	1.56	3.87	M
1	HEY1	hairy/enhancer-of-split related with	1.86	3.51	I

		YRPW motif 1			
1	IPO11	importin 11	2.04	3.19	I
1	TNFRSF25	tumor necrosis factor receptor superfamily, member 25	1.91	3.29	M
1	AKR7A2P1	aldo-keto reductase family 7, member A2 pseudogene 1	2.76	2.31	I
1	SACS	spastic ataxia of Charlevoix-Saguenay (sacsin)	2.48	2.51	I
1	TMEM119	transmembrane protein 119	1.36	3.51	M
1	ZNF23	zinc finger protein 23 (KOX 16)	1	3.87	I
1	SLC44A5	solute carrier family 44, member 5	1.77	3.03	M
1	PDGFB	platelet-derived growth factor beta polypeptide	2.31	2.44	E
1	STAC	SH3 and cysteine rich domain	1.05	3.7	I
1	ENG	Endoglin	1.11	3.63	M
1	SLC36A4	solute carrier family 36 (proton/amino acid symporter), member 4	1.67	3.03	M
1	LOC654342	lymphocyte-specific protein 1 pseudogene	2.87	1.81	I
1	TAF3	TAF3 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 140kDa	2.87	1.75	I
1	KANSL2	KAT8 regulatory NSL complex subunit 2	2.69	1.93	I
1	MED11	mediator complex subunit 11	1.33	3.29	I
1	PDCL3	phosducin-like 3	1	3.6	I
1	LOC100128822	hypothetical LOC100128822	1.82	2.72	I
1	GOLGA6L5	golgin A6 family-like 5 (pseudogene)	1	3.51	I
1	TIPRL	TIP41, TOR signaling pathway regulator-like (S. cerevisiae)	2.4	2.1	I
1	CEP128	centrosomal protein 128kDa	2.18	2.31	I
1	ADNP2	ADNP homeobox 2	1.84	2.58	I
1	NSUN6	NOP2/Sun domain family, member 6	1.38	3.03	I
1	SLC36A1	solute carrier family 36 (proton/amino acid symporter), member 1	1.02	3.29	I
1	C2orf68	chromosome 2 open reading frame 68	1	3.31	I
1	CA13	carbonic anhydrase XIII	1.56	2.72	I
1	GLT25D1	glycosyltransferase 25 domain containing 1	2.12	2.15	I
1	PLK4	polo-like kinase 4	2.5	1.75	I
1	MT1A	metallothionein 1A	1.51	2.72	I
1	PLXNB3	plexin B3	1.91	2.31	M

1	IGFLR1	IGF-like family receptor 1	1.84	2.31	M
1	ADAP2	ArfGAP with dual PH domains 2	1.82	2.31	I
1	LOC283663	hypothetical LOC283663	1.7	2.31	I
1	RANBP1	RAN binding protein 1	1.4	2.47	I
1	GPATCH1	G patch domain containing 1	1.14	2.72	I
1	ZNF420	zinc finger protein 420	1.51	2.31	I
1	TXNDC5	thioredoxin domain containing 5 (endoplasmic reticulum)	1.29	2.53	I
1	MTCP1NB	mature T-cell proliferation 1 neighbor	2.06	1.75	I
1	SLC44A3	solute carrier family 44, member 3	1.08	2.72	M
1	SPRY1	sprouty homolog 1, antagonist of FGF signaling (Drosophila)	1.46	2.31	I
1	SULF2	sulfatase 2	1.76	2	E
1	PDPN	Podoplanin	1.86	1.75	M
1	KLHL4	kelch-like 4 (Drosophila)	1.2	2.4	I
1	SULT1E1	sulfotransferase family 1E, estrogen-preferring, member 1	1.84	1.75	I
1	COL4A1	collagen, type IV, alpha 1	1.47	2.09	E
1	SMOX	spermine oxidase	1.77	1.75	I
1	NDUFB2	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 2, 8kDa	1.01	2.51	I
1	FUK	Fucokinase	1.2	2.31	I
1	LOC388152	hypothetical LOC388152	1.67	1.75	I
1	CDKAL1	CDK5 regulatory subunit associated protein 1-like 1	1.08	2.31	M
1	MRPS24	mitochondrial ribosomal protein S24	1.27	2.09	I
1	NCOA2	nuclear receptor coactivator 2	1.56	1.75	I
1	RPL26L1	ribosomal protein L26-like 1	1.54	1.75	I
1	TMEM241	transmembrane protein 241	1.94	1.32	M
1	PCDHA9	protocadherin alpha 9	1.51	1.75	M
1	OLFML1	olfactomedin-like 1	1.48	1.75	E
1	AAED1	AhpC/TSA antioxidant enzyme domain containing 1	1.48	1.75	I
1	DIAPH2	diaphanous homolog 2 (Drosophila)	1.41	1.81	I
1	GALNT10	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 10 (GalNAc-T10)	1.66	1.43	I
1	JAG2	jagged 2	1.33	1.75	M
1	TOMM6	translocase of outer mitochondrial membrane 6 homolog (yeast)	1.64	1.4	I
1	FANCI	Fanconi anemia, complementation group I	1.7	1.29	I
1	ZNF618	zinc finger protein 618	1.29	1.62	I

1	URGCP-MRPS24	URGCP-MRPS24 readthrough	1.44	1.46	I
1	CSTB	cystatin B (stefin B)	1.59	1.26	I
1	ATP5L2	ATP synthase, H+ transporting, mitochondrial Fo complex, subunit G2	1.05	1.75	M
1	ZNF83	zinc finger protein 83	1.07	1.72	I
1	HIF1A	hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)	1.12	1.49	I
1	ADAT2	adenosine deaminase, tRNA-specific 2	1.29	1.25	I
1	C17orf69	chromosome 17 open reading frame 69	1.28	1.04	I
1	CKS1B	CDC28 protein kinase regulatory subunit 1B	1.08	1.02	I

Table 4.16.1; 126 genes were found to have, for both an endothelial and angiogenic results, a log2(fold change) ≥ 1. 15 of these genes were found to be tumour enriched in more than on RNA-seq contrast. 20 genes are membrane, 16 extracellular and the remaining 90 are intracellular for cellular component. The key to the final column (Class): M = membrane, E = extracellular and I = intracellular.

4.17 Endothelial transcriptome, tumour and normal

Transcriptome assemblies were built using Cufflinks 1.3.0 program on each of the 9 different RNA-seq samples used in this work. The resulting multiple Cufflink transcriptomes were merged together into a single assembly with the Cuffmerge program [104], where each transcript was partitioned into one of several classes. The different classes and the number of different transcripts in each class can be gleaned from table 4.17.1. One novel intergenic transcript had 51 exons (accession number TCONS_00079877) and in total there were 21,351 potential novel, new genome transcripts found. For known genes, 20,948 potential new isoforms were predicted. For the whole set of novel and new isoform transcriptome FASTA sequences, please see http://sara.molbiol.ox.ac.uk/userweb/jherbert/.

Table 4.17.1 Endothelial transcriptome, a table showing differently classed transcripts produced with Cufflinks

Priority	Code	Description	Count
1	=	Transcript matches the intron chain and exons agree with a known gene	**38,554**
2	c	Transcript is contained within a transcript of a known gene	**0**
3	j	This is potentially a novel isoform of a known gene as, at least, one splice junction is shared with a reference transcript	**20,948**
4	e	A single exon transcript that overlaps both an exon and a reference intron. This suggest a possible pre-mRNA.	**0**
5	i	An intronic transcript	**20**
6	o	Transcript overlaps an exon of a reference transcript	**584**
7	p	A PCR run-on, possibly connecting two genes	**1**
8	r	Transcript overlaps with lower-case bases in genome and is classed as a repeat.	**0**
9	u	Novel intergenic transcript	**21,351**
9.5	**U**	Long novel intergenic transcript with ≥ 10 exons (presence of multiple splice sites)	**19**
10	x	Transcript that overlaps an exon but on the opposite strand	**1,707**
11	s	Mapping errors where an intron of a transcript maps over a reference intron but on the opposite strand.	**34**
12	.	Multiple classifications of the above but is deciphered in the tracking file.	**0**

Table 4.17.1; Different classes of transcript and counts; the above table is based on the that of Cufflinks web-page manual, with the addition of frequency of differently classed transcripts and a new class, a capital U, standing for novel intergenic transcripts that have 10 or more exons (see Cufflinks manual [130]).

4.18 Differentially expressed novel transcripts

The output of the Cuffmerge program was a Gene Transfer Format (GTF) file that was subsequently used in differential gene expression experiments using Cuffdiff that included novel transcripts from the 7 merged transcriptome assemblies. Precedence was given to multiple exon novel transcripts with a log2(FC) ≥ 1 enrichment in tumours, as well as choosing transcripts that had the largest average expression (transcripts expressed at a low level were ignored). This filtering was performed manually and candidates were further investigated using BLAST (similarity and homology searches), BLAT with the University of California Santa Cruz (UCSC) Genome browser for exon architecture, and functional domain prediction using InterProScan [131]. One such promising novel transcript candidate was transcript TCONS_00068524, which had a log2(FC) ≥ 2 in both bladder tumour endothelial cells and tumour conditioned HUVEC (see figure 4.18.1). An enrichment was also seen in tumour colon and tumour lung endothelium (Illumina) but not in the SOLiD4 tumour lung. This transcript had an average expression FPKM of 157.

The nucleotide and 3rd frame translation sequences of TCONS_00068524 are given below.

>**TCONS_00068524 mRNA sequence** chr6:144521416-144522188;

TAGCACCGCTCAGGCTGCACTGCGCTTGCTGCAAGCTTCAGGCTCCTGCTAAG
CTAGTGCTGCATCGTCTCCCTTCAGCCGCCATCATGATTATCTACCGCGACCT
CATCAGCCACGATGAGATGTTTTCCGACATCTACAAGATCTGGGAGATCATG
GACGGGCTGTACCTGGAGGTGGAGGGGAAGATGGTCAAAAGCACAGTAATC
ACTGGTGTTGATATTTTCTTTAATTGTGAAAACATGAATCCAGATGGCATGGT
TGCTCTATTGGACTACCGTGAAGATGGTGTGACCCCATATATGATTTTCTTTA
AGGATGATTTAGAAATGGAAAAATGTTAACAAATTTGGCAATTACTTTGATC
TATCACCTGTCATTGTAACTGGCTTCTGCTTGTCATCCACACAACACCAGGAC
TTAAGACAAATTGGGGCTGATGTCATCTTGAGCTCTTCATTTATTTTGACCGT
GATTTATTTGGAGTGGAGGCATTGTTTTTAAGAAAAAC

>**TCONS_00068524** (Third frame translation of the novel transcript. The longest open reading frame is in green to the first stop codon, which is the frame that matches a Refseq protein sequence)

XXXXXXLAPLRLHCACCKLQAPAKLVLHRLPSAAIMIIYRDLISHDEMFSDIYKI
WEIMDGLYLEVEGKMVKSTVITGVDIFFNCENMNPDGMVALLDYREDGVTPYM
IFFKDDLEMEKC*QIWQLL*SITCHCNWLLLVIHTTPGLKTNWG*CHLELFIYFDR
DLFGVEALFLRK

A BLAST search of the 6 frame translation of this sequence to the Refseq database of proteins found similarities to a tumour induced protein in an Orangutan species (Sumatran orangutan, Pongo abelii), accession number XP_002831390. A three exon sequence architecture was derived by aligning the nucleotide sequence to the human genome (HG19) using BLAT at the UCSC genome browser. The results of this can be seen in figure 4.18.2 in the top panel. The bottom panel shows the results of the functional domain scan as performed using InterProScan.

Figure 4.18.1 Differentially expressed novel transcript TCONS_00068524

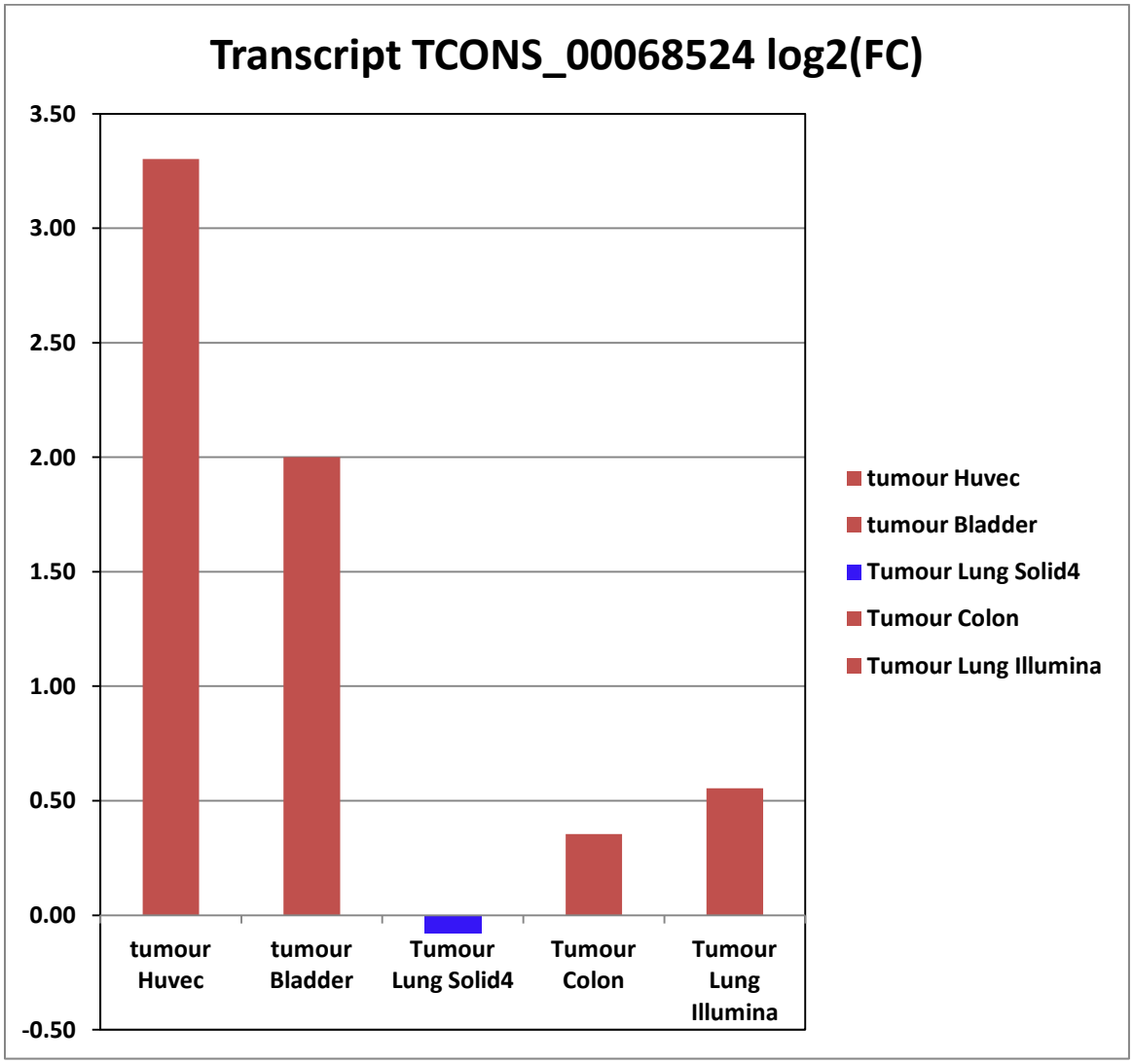


Figure 4.18.1; A differentially expressed novel transcript, TCONS_00068524, was identified up-regulated in tumour conditioned HUVEC and tumour bladder endothelial cells; enrichment of expression was also seen in tumour endothelial cells of lung and colon (but slight down regulation in tumour lung, SOLiD4). Log2-fold change of tumour versus normal endothelium expression is reported.

Figure 4.18.2 TCONS_00068524 genome and domain architecture

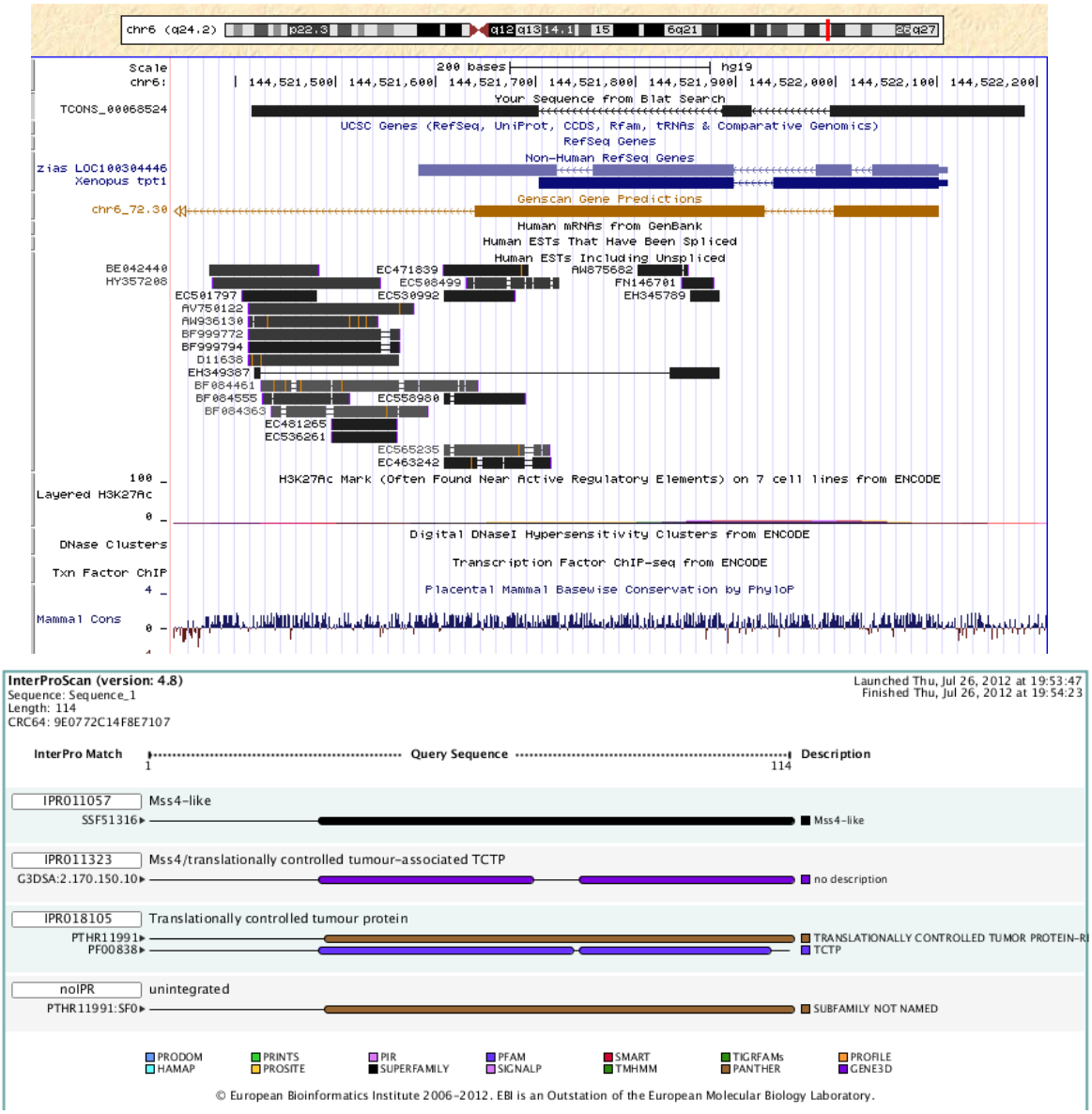


Figure 4.18.2; Top panel; UCSC genome browser genomic architecture overview of the 3 exon differentially expressed novel transcript. Bottom panel; functional domain prediction of the translation of the novel transcript, TCONS_00068524. The results from InterProScan at the EBI database show conserved functional domains of mammalian translational controlled tumour protein (TCTP) in the novel transcript TCONS_00068524.

4.19 Isoform switching; does this occur in tumour endothelium?

The extracellular matrix around tumours contains specific forms of Fibronectin 1(FN1) and Tenascin C (TNC), which are structurally different to those expressed under normal physiological conditions due to differential splicing events [132]. These differential isoforms can affect the behaviour of tumour cells, with the TNC isoform having a pro-migratory and proliferative effect on tumour cells at sites of breast cancer [133]. FN1 expresses an extra-domain B (EDB) that is found enriched on tumour vasculature but also in the lymph nodes of lymphoma patients; an EDB specific antibody has been used as a targeted cancer therapy for lymphoma disease [134].

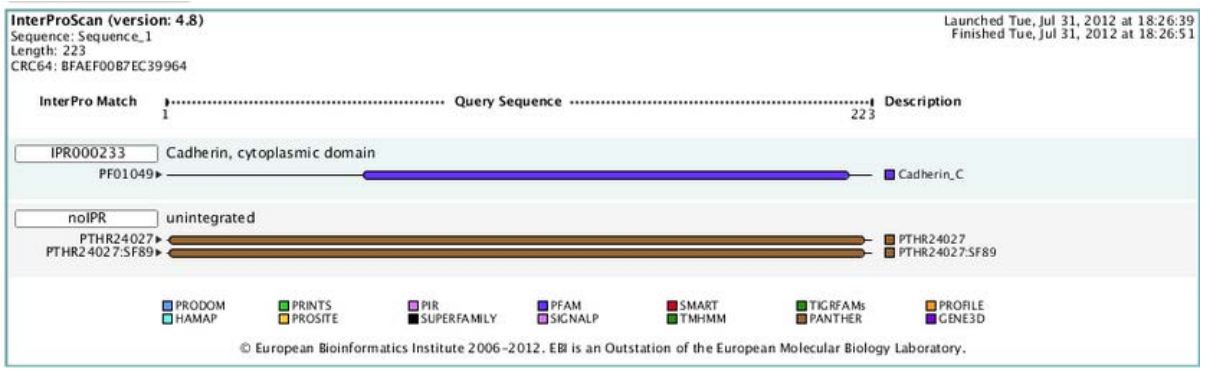
In this work it was interesting to validate isoforms using Cufflinks where two protein sequences from the same gene appeared tumour and normal enriched respectively. Isoform switching genes were chosen where the longer protein isoform was up-regulated in tumour endothelium and the shorter protein enriched in normal endothelium. Both the Refseq and Ensembl transcriptomes were used in conjunction with Cufflinks [104] to identify possible candidates. Table 4.19.1 lists the 9 isoform switch candidates, which includes the FPKM expression level and the length of each protein isoform. Figure 4.19.1 shows the functional/structural domain architecture of the long and short forms of Cadherin 5, type 2 (vascular endothelium, CDH5), (Ensembl based). The expression results suggest a longer form of CDH5 is enriched in tumour Huvec and a shorter protein, only containing the cytoplasm Cadherin domain, is normal lung enriched.

Table 4.19.1 Cufflinks isoform switch candidate genes

Gene	Tumour Up protein	Protein length	Tumour Up RPKM	Normal up protein	Protein length	Normal Up RPKM	Tissue contrasts the differences were found
CDH5	ENSP00000437691	525	120_9	ENSP00000461880	223	0_16	Tumour Huvec
PKM2	ENSP00000334983	531	219_45	ENSP00000413887	458	2_293	Lung (Illumina)
SPARCL1	ENSP00000414856	664	240_6	ENSP00000438188	127	0_21	Lung (SOLiD4)
PPP1R15A	ENSP00000200453	674	415_32	ENSP00000443279	632	0_86	Lung (Illumina)
VCAM1	NP_001069	739	28_5	NP_001186763	677	0_8	Lung (Illumina)
CD99	NP_002405	185	25_0	NP_001116370	169	0_38	Lung (SOLiD4)
CTLA4	NP_005205	223	5_0	NP_001032720	174	0_8	Bladder
PMEPA1	NP_064567	287	26_3	NP_954638	252	1_12	Lung (Illumina)
C2orf74	NP_001137431	194	9_0	NP_001137432	115	0_19	Lung (SOLiD4) Tumour Huvec
FN1	NP_997647	2477		NP_997643	2330		EDB domain Control

Table 4.19.1; This table lists the genes with Cufflinks evidence of isoform switching. The expression levels of each isoform are given with two FPKM values separated by an underscore character ("_"); an example is 100_2, which means in tumour endothelium, this isoform had an expression level of 100 FPKM and in normal, 2 FPKM. The expression in tumour endothelium is the first number in both "Tumour Up FPKM" and "Normal Up FPKM" headed columns in the table. In total, 9 potential isoform switching genes were identified, with Fibronectin 1 also listed as a control.

Figure 4.19.1 CDH5 domain overview, long and short isoforms



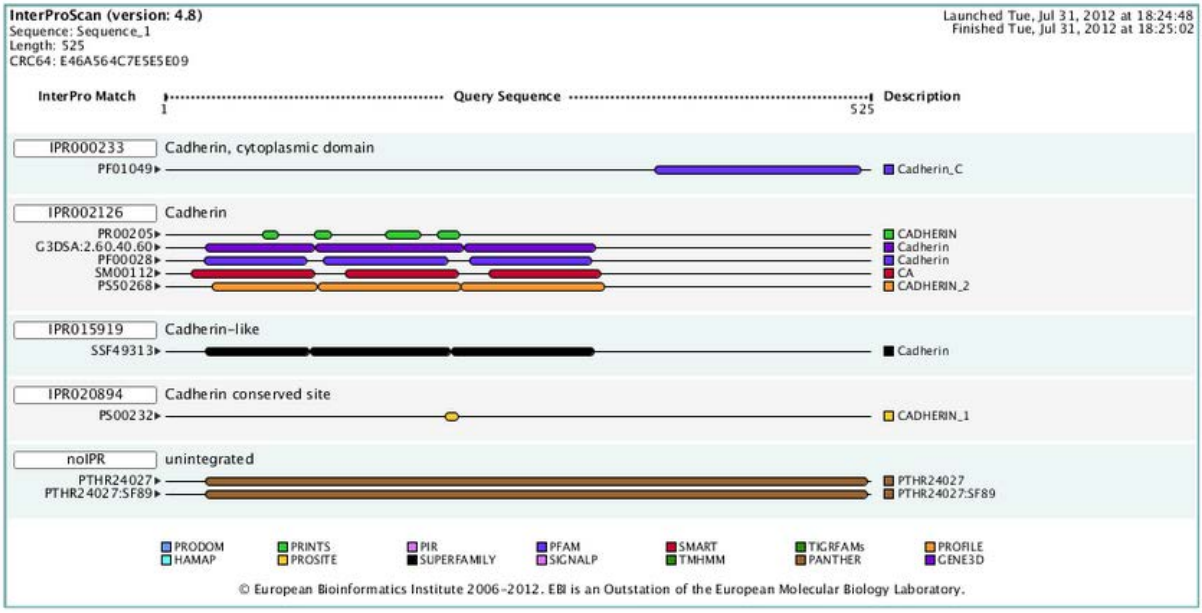


Figure 4.19.1; The functional and structural domain architecture of the long and short isoforms of CDH5 predicted to be differentially expressed in tumour Huvec versus normal bulk lung. The longer isoform (bottom panel) contains both the Cadherin repeats that are extracellular and in the cytoplasm; this isoform was up-regulated in tumour Huvec. The shorter protein, up-regulated in normal lung, only contains the cytoplasm Cadherin domain (top panel).

4.20 Results of isoform switching

For all isoforms in table 4.19.1, primers were designed that span intron boundaries that are specific to each isoform, with a generic reverse primer suitable for both isoforms. Lung tumour and normal endothelial cDNA isolates were used as input to SYBR green qPCR. With the exception of VE-cadherin (CDH5), CTLA4 and the positive control, FN1, no significant isoform switching or differential expression was evidenced. This is a surprising result, as, for example, according to Cufflinks the results for gene PPP1R15A has considerable FPKM expression differences between isoforms. Cufflinks uses a statistical model to assign expression measures to different isoforms of genes, based on read to genome alignments from tophat [103, 104]. It is possible that this model is not always accurate in assigning quantitative measures to different isoforms.

On a more positive note, there was a clear difference in expression between the long and short forms of CDH5 according to the qPCR results. It is not isoform switching but the small version appears to be 16 fold up-regulated in normal versus tumour, whereas the long form is 4 fold up-regulated in normal. It is possible that the short form of CDH5 is switched off in tumour endothelial cells because of the high cycle threshold number for this gene. The longer isoform is present in tumour and normal but is down regulated in tumour endothelium. One interesting note is that CDH5 is on a region of chromosome 16 which suffers loss of heterozygosity in some cancers [135]. However, as endothelial cells are apparently genetically stable, this an unlikely cause of down-regulated expression in tumour endothelium. Another possibility is the fact that Cadherins are involved in cell to cell adhesion, which is destabilised during invasion and metastasis of tumour cells. A hallmark of endothelial cells around tumours is that vessels are crooked and fragile, which gives a feasible reason as to why both isoforms of CDH5 are down-regulated in angiogenic primed tumour endothelium versus normal, manifested by weaker cell to cell junctions.

Cytotoxic T-Lymphocyte Antigen 4 (CTLA4) shows evidence of isoform switching. The larger protein is 3 fold up-regulated in tumour endothelium versus normal endothelium,

whilst the shorter protein shows modest up-regulation in normal endothelium. Searching literature does not find evidence of CTLA4 expressed in endothelium. A search of the Unigene cluster for this gene also shows no evidence of vascular expression and analyses using Cancer Genome Anatomy Project SAGE libraries also failed to produce any endothelial gene expression evidence. This gene has been reported to be a T-cell receptor, on the surface of T-helper immune system cells. When activated by a ligand, it initiates an inhibitory signal to the T-cell, which ultimately leads to repression of the immune response [136]. This has a possible connection to tumour growth, as suppression of the immune system against cancer cells and surrounding tissues will aid growth and metastasis. It is possible some T-cells were co-isolated with the tumour endothelium.

4.21 Endothelial organ specific expression

Although not directly related to vascular targeting, it is of scientific interest to look at endothelial expression in specific organs or tissue compartments. This was carried out using an RPKM analyses (see above) for lung, bladder, colon, liver and tumour conditioned HUVEC. For each compartment, gene expression RPKM values were compared to the average of all other endothelial cell types using a log2 fold change ratio. A cut-off of 4 fold change was applied. A further Venn overlap analyses was employed to identify any genes had high ratios in two or more tissues (there were 102 genes like this, Figure 4.21.1). From the combined RPKM and Venn analyses, the breakdown of compartment enriched genes was;

- Lung enriched = 1,168 genes
- Liver enriched = 2,190 genes
- Colon enriched = 1,527 genes
- Bladder enriched = 2,438 genes
- Tumour conditioned HUVEC = 1,956 genes

The full list of 9,279 genes for each compartment, together with AngioScore, Endothelial cDNA library analyses and Angiogenic cDNA library screen, can be found in a download Excel file at the following link, selecting the download file called "Endothelial compartment/organ enriched genes (RPKM)";

http://sara.molbiol.ox.ac.uk/userweb/jherbert/tissue_diffex/appendix_files.html
A full transcriptome RPKM sheet is also in this file so that a user can look up a gene of interest to see its endothelial expression profile across different organs.

Figure 4.21.1 Overlap of compartment enriched endothelial expression

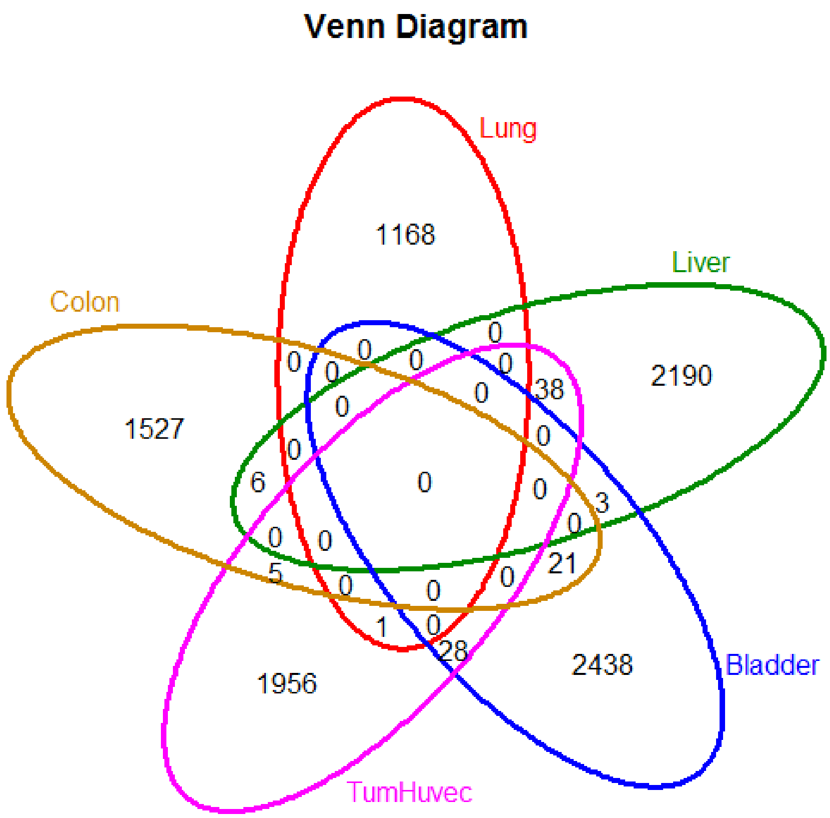


Figure 4.21.1; Combining log ratios of compartments versus all others with a Venn analyses finds 9,279 compartment enriched genes.

4.22 Summary

For the first time, short read RNA-seq data of bead-isolated tumour endothelium from multiple organs has been used to identify pan-VT targets with deep sequencing technology. To aid selection of best VT's, several different rank orders of genes were generated based on cDNA library and published data. A new endothelial transcriptome has been constructed, which identifies differentially expressed novel transcripts. From results here, there was no compelling evidence for the existence of isoform switching at the expressed transcript level within tumour endothelium. However, CDH5 and FN1 showed differential regulation between tumour and normal endothelium. In summary, a valuable list of vascular targets has been produced, which can be taken by others for validation. A proportion of the candidates genes are currently undergoing experimental validation and it is hoped a manuscript will be ready for the end of the year.

4.23 Acknowledgement of contributions

All the bioinformatics analyses were carried out by the author of this thesis, which is also true of the tumour environment HUVEC (cord processing, endothelium isolation, RNA extraction and qPCR). The author would like to thank Xiaodong Zhuang and Forhad Ahmed who performed the fresh tumour endothelial cell isolation prior to sequencing and the SYBR green qPCR and other qPCR data.

CHAPTER 5; IDENTIFYING VASCULAR TARGETS IN LUNG CANCER USING MICROARRAY DATA FROM MULTIPLE PATIENT TUMOUR ENDOTHELIUM

5.1 Introduction

In previous chapters, vascular targets were sought with sequencing technologies where the cost prohibited analyses of multiple biological replicates. In this chapter, biological replicates and patient variability were accounted for in an analysis of 4 patient samples of non-small cell lung cancer (NSCLC). Tumour and adjacent normal lung endothelial cells were isolated from fresh bulk tissues and these were then hybridised on to the latest Agilent gene expression array chips to identify differentially expressed genes. Standard bioinformatics tools were then used to decipher which genes would be the most targetable via the blood supply. **Please note, appendix tables are given in a separate volume, also downloadable from http://sara.molbiol.ox.ac.uk/userweb/jherbert/tissue_diffex/appendix_files.html.**

5.2 Sample collection, preparation and bioinformatics analyses

Lung tumour and distant normal tissues were obtained immediately after surgery with patient consent. Tissues were weighed, minced and then digested for 1.5 hours with Dulbecco's modified Eagle's medium containing the enzyme Collagenase. The digested supernatant was filtered through a 70 micron filter and Ulex europaeus lectin coated beads, which bind L-fucose sugar molecules that were then used to positively select out endothelial cells [74]. Total RNA was extracted using the RNA-easy minikit of Qiagen and cRNA was generated and subject to amplification and labelling. The amplified cRNA was hybridized onto Agilent whole human gene expression microarrays. The Bioconductor packages preprocessCore [137] and Limma were used to background subtract and Quantile normalize probe signal intensities, prior to differential gene expression analyses using Linear Models for Microarrays (Limma) [108]. 1- Correlation

distances, 2d-clustering and principle component analyses were performed in R/Bioconductor [138].

5.3 Patient and sample variability

Gene expression heterogeneity between samples was expected as they were collected and processed from different patients and on different days. Several global measures of gene expression (microarrays in this case) collect an extreme higher number of variables compared to the number of samples being assayed. Because of this high dimensionality of microarray gene expression data, visualization of global gene expression differences between different samples and how they are related to each other is difficult. One way this is made possible is by using a principle component analysis that minimizes the dimensions of the data whilst keeping a maximal amount of the variation contained within the data [139, 140]. It identifies straight line directions that have the maximal amount of variation and these are called the principle components. Samples are able to be separated along the principle components using a low numbers of variables that make it possible to group/visualise similarities and differences between the different samples and sample groups. In these circumstances, the variability between samples and sample groups can be seen as an advantage as genes significantly up-regulated will be so in spite of sample differences. Therefore, any gene that is consistently up-regulated in tumour endothelium versus normal will be a general marker of NSCLC but other genes that are up-regulated in a subset of tumours or grades are sample specific and will not be identified. Figures 5.3.1 and 5.3.2 show principle component analyses on global gene expression and on the most significantly differentially expressed genes respectively. This provides evidence that there exists some general tumour endothelial markers of NSCLC. A Dendrogram of 1 - correlation co-efficient of global gene expression data shows that tumour sample 1 (T1) is clearly different to other tumour and all normal samples (Figure 5.3.3). Further evidence of clear gene expression differences is exemplified when performing Hierarchical clustering on the most statistically significantly differentially expressed genes. This results in clear distinctions between sample groups and gene expression levels (Figure 5.3.4).

Figure 5.3.1 Principle Component Analyses of global gene expression data

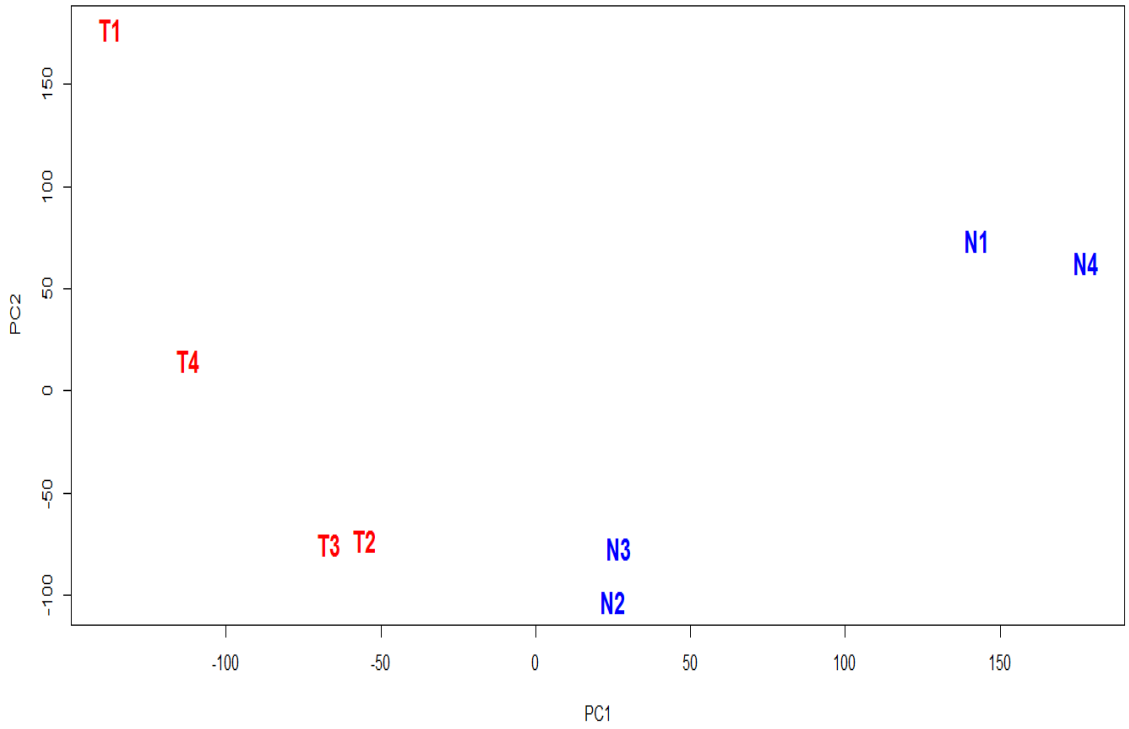


Figure 5.3.1; A principle component analysis (PCA) plot of samples using all of the gene expression data (normalized probe intensities and replicate probes averaged). As with the correlation distance analyses (see figure 5.3.3), it is clear tumour sample 1 (T1) is very different from all other samples. Along the first principle component, it is evident that normal samples N1 and N4 are different from normal samples N2 and N3 at a global level of gene expression. The variability shown by the different samples will lead to identification of genes that are consistently regulated despite the heterogeneity in samples. Tumour samples T1 to T4 are coloured in red and blue labels N1 to N4 are the normal samples.

Figure 5.3.2 Principle Component Analyses of the most differentially expressed genes

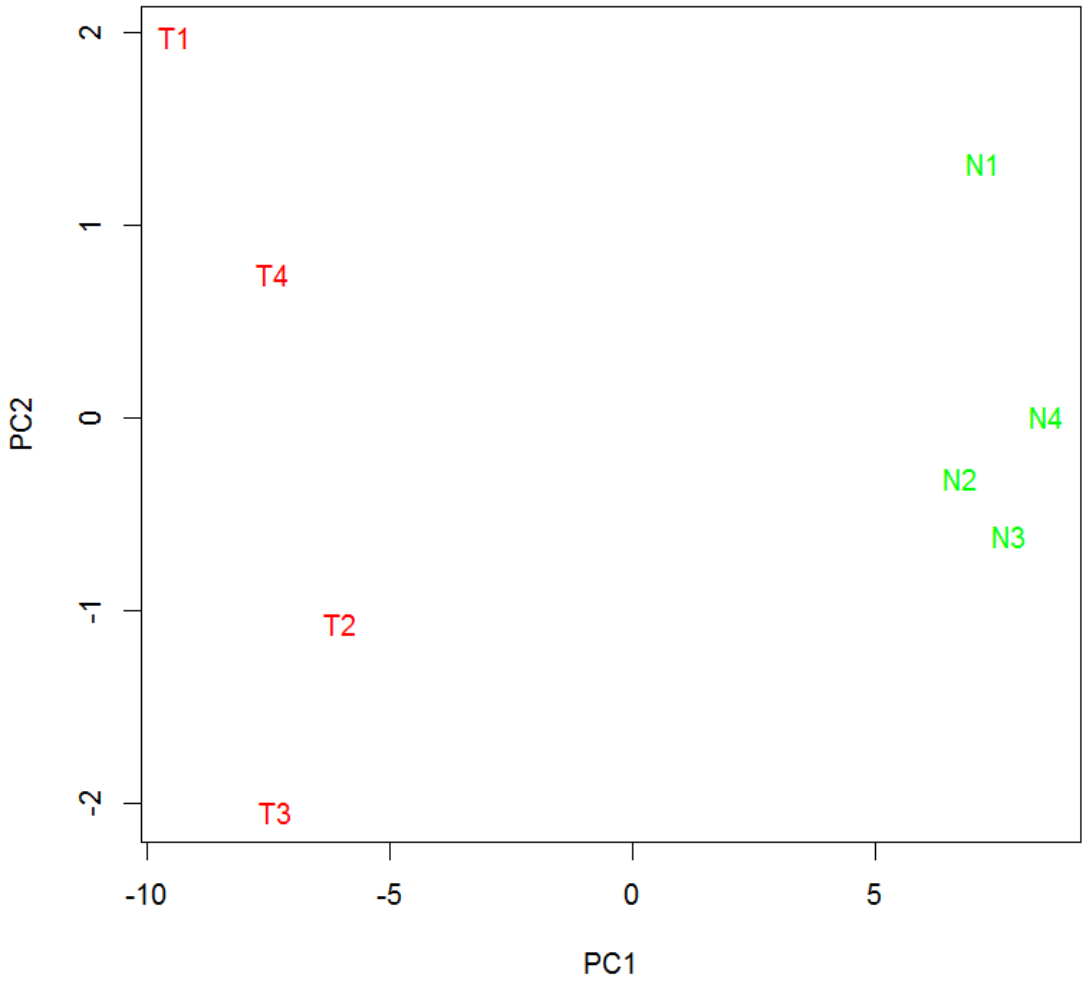


Figure 5.3.2; A principle component analysis (PCA) plot of samples using only the most significantly differentially expressed genes. The first principle component (PC1) shows the largest variation of data in one dimension and there is a clear distinction between the tumour and normal sample groups; despite sample heterogeneity as shown when using the complete gene expression set of data, this plot provides evidence of significantly differentially expressed genes regardless of sample heterogeneity. Tumour samples T1 to T4 are coloured in red and in green are the N1 toN4 labelled normal samples.

Figure 5.3.3 Distance calculated as 1- correlation of global gene expression as a dendrogram of samples

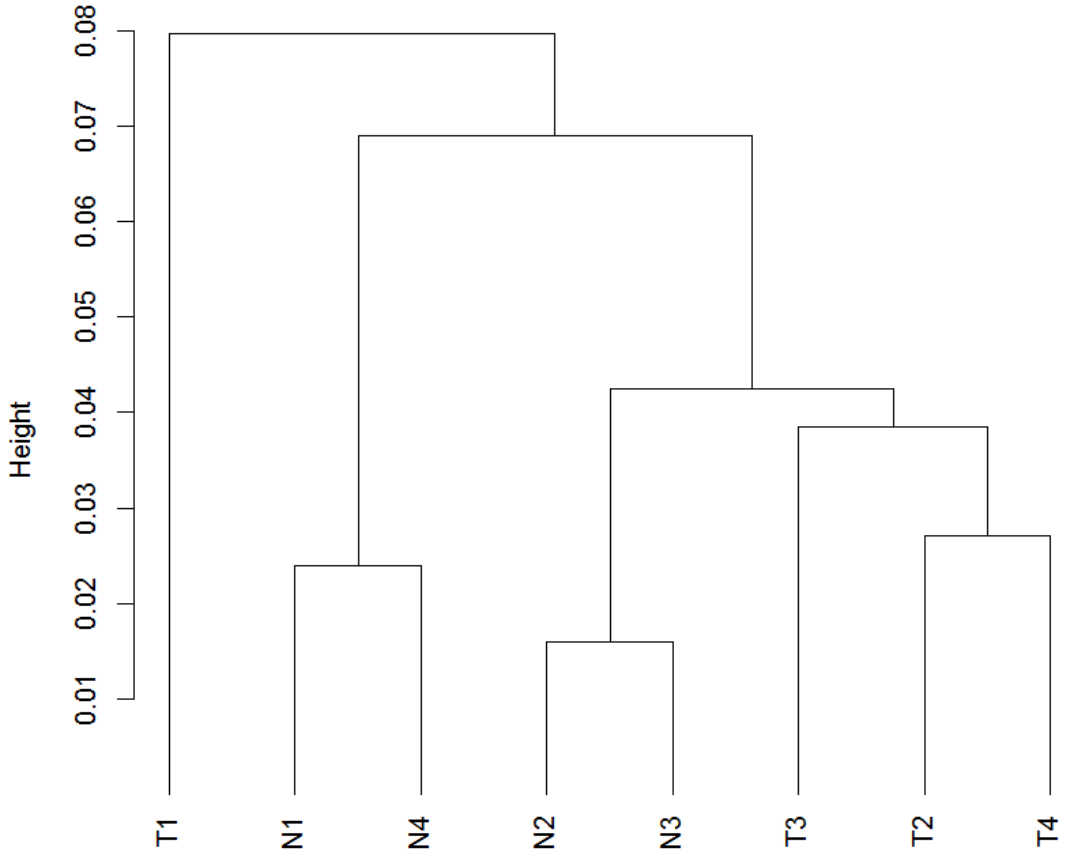


Figure 5.3.3; Distances between different samples can be calculated with various methods, for example Euclidean distance. Here a 1 minus Pearson Correlation Co-efficient was used to generate a Dendrogram. It is clearly visible that tumour sample 1 (T1) is very different to the other tumour samples and all normal samples.

Figure 5.3.4 Hierarchical clustering of the top 70 differentially expressed genes

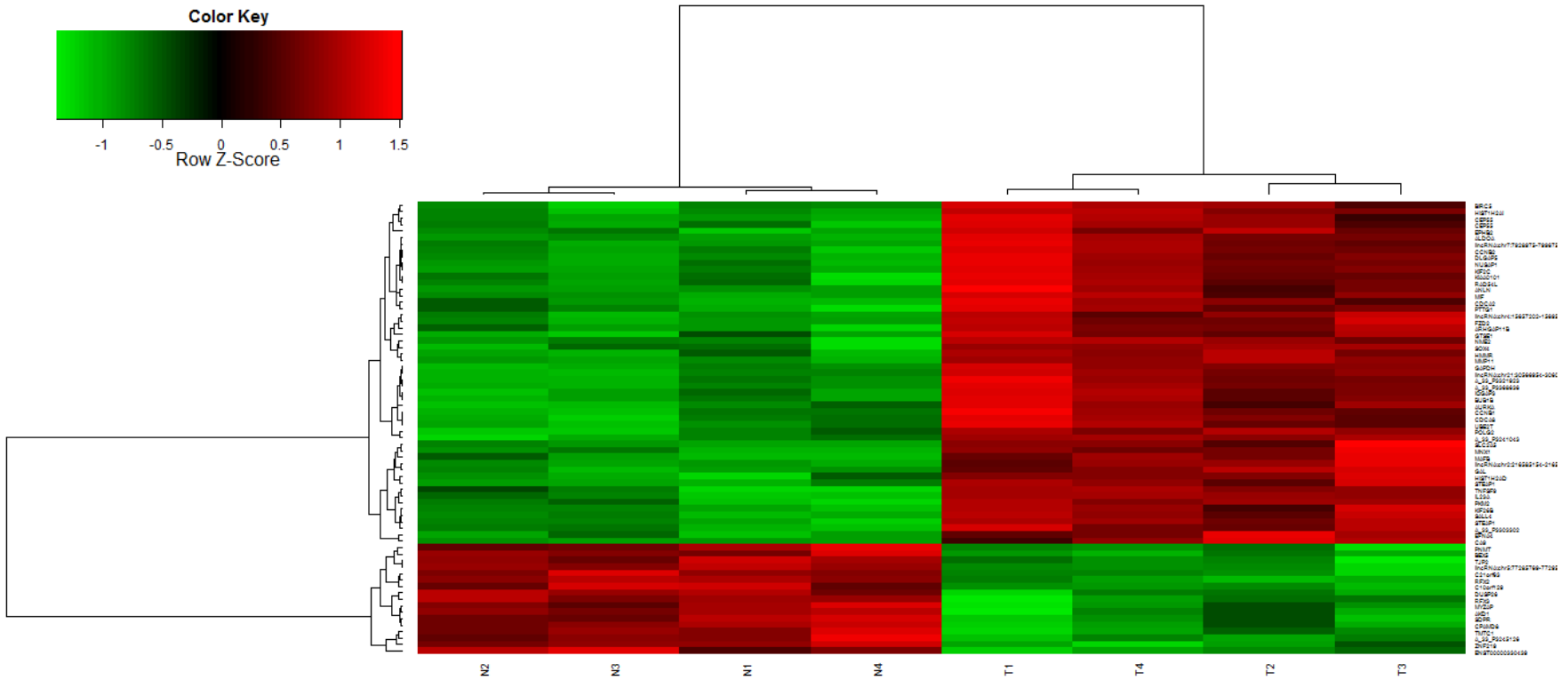


Figure 5.3.4; The top 70 most significantly differentially expressed genes were subject to 2 dimensional clustering using 1 - correlation as a distance measure and the Ward method of hierarchical clustering. There is a clear distinction between sample groups and differentially expressed genes.

5.4 The most consistent and significantly differentially expressed genes

Differentially expressed genes were analysed using the Linear Models for Microarray Bioconductor package [108, 138]. This produced a table of results that included the following for each probe/gene: a moderated T-statistic (where the standard errors were worked out and moderated using all genes) and a B statistic. The B statistics is the logarithm of odds that a gene is differentially expressed. The exponential function in Excel returns e (which is a constant number based on the rate of growth that all constantly growing processes posses [141]), to the power of a provided number (in this case the B statistic). The B statistic of the most significantly differentially expressed gene was 8.44 (MMP11). The odds of this gene being differentially expressed is; the e constant to the power of 8.44 is 4,629, and the probability that MMP11 is differentially expressed in this data set is 4,629/(1+4,629) is 0.99; or a 99% chance that MMP11 was differentially expressed between tumour and normal samples (up-regulated in this case). When a B statistic of zero is found, there is a 50% chance the gene is differentially expressed [142]; genes determined as statistically significantly differentially expressed were those with a positive B value and a multiple correction (Benjamini and Hochberg [143]) adjusted p-value ≤ 0.04. The total number of genes found at this significance level was 747. The number of genes from this that were up-regulated in tumour endothelium versus normal endothelium = 343 and these are listed in table 5.4.1 (284 of these are known genes), along with cDNA library results, AngioScore and cellular component. In terms of the best candidates to validate experimentally based on access via blood supply, the breakdown of the cellular component is as follows: 36 genes were found to be extracellular, 45 membrane and the remaining 262 intracellular.

Table 5.4.1 Significantly differentially expressed genes

Systematic Name	Gene	Log2 FC	adj. P.Val	Endo Log2(FC)	Angio Log2(FC)	Angio Score	Cell comp.
NM_004887	CXCL14	6.7	0.04	-2.64	-0.59	**31**	E
NM_000088	COL1A1	5.1	0.04	-9.18	1.71	**6**	E
NM_006274	CCL19	4.0	0.04		1.26	**18**	E
NM_000090	COL3A1	3.8	0.04	-12.44	2.12	**14**	E
NM_005940	MMP11	3.6	0.00	-7.03	2.66	**36**	E
NM_015886	PI15	3.5	0.03		0.23	**33**	E
NM_003013	SFRP2	3.4	0.04		1.14	**26**	E
NM_020127	TUFT1	3.3	0.03	-2.47	-0.55	0	E
NM_006072	CCL26	3.3	0.03		1.99	**20**	E
NM_002426	MMP12	3.1	0.02	-11.05	2.63	**19**	E
NM_006274	CCL19	3.0	0.03		1.26	**18**	E
NM_006419	CXCL13	2.8	0.02		1.41	**17**	E
NM_015973	GAL	2.8	0.01	-6.1	-0.7	**11**	E
NM_024746	HHIPL2	2.7	0.02	-6.1	3.69	0	E
NM_000493	COL10A1	2.6	0.04		4.97	**6**	E
NM_001145031	PLAU	2.5	0.03	-3.4	0.33	**38**	E
NM_002658	PLAU	2.5	0.04	-3.4	0.33	**38**	E
NM_016584	IL23A	2.3	0.01		-0.44	**12**	E
NM_201442	C1S	2.3	0.03	-6.6	0.17	**4**	E
NM_003811	TNFSF9	2.2	0.01	-6.1	0.23	**49**	E
NM_001734	C1S	2.1	0.03	-6.6	0.17	**4**	E
ENST00000390282		2.0	0.03				E
NM_153221	CILP2	1.9	0.03			**19**	E
NM_002674	PMCH	1.8	0.04	4.02	-1.26	**3**	E
NM_000435	NOTCH3	1.8	0.04	-8	0.35	**40**	E
NM_152890	COL24A1	1.8	0.04		3.69	**17**	E
NM_001733	C1R	1.7	0.04	-5.04	-0.32	**5**	E
NM_002415	MIF	1.5	0.01	-2.75	1.57	**27**	E
NM_144651	PXDNL	1.3	0.03		5.23	**11**	E
NM_000034	ALDOA	1.2	0.01	-4.46	-0.17	**4**	E
NM_005101	ISG15	1.2	0.04	-3.22	0.46	**13**	E
ENST00000402102	THNSL2	1.2	0.03	-7.59	-0.78	**13**	E
NM_144772	APOA1BP	1.1	0.03	-3.84	0.75	**17**	E
NM_144772	APOA1BP	1.0	0.03	-3.84	0.75	**17**	E
NM_000918	P4HB	0.9	0.03	-3.2	0.99	**9**	E
NM_021130	PPIA	0.7	0.04	-0.92	0.12	**9**	E
NM_012449	STEAP1	3.9	0.01	4.02	-2.67	**43**	M
NM_001216	CA9	3.4	0.01		2.04	**74**	M

NM_173076	ABCA12	3.3	0.04		-6.38	0	M
NM_012449	STEAP1	2.9	0.01	4.02	-2.67	**43**	M
NM_003039	SLC2A5	2.5	0.01	-7.59	-0.21	**16**	M
NM_024893	SYNDIG1	2.5	0.04		0.48	0	M
NM_173843	IL1RN	2.5	0.03	-2.64	-0.06	**12**	M
NM_144717	IL20RB	2.4	0.04	7.6	0.94	**13**	M
NM_024628	SLC12A8	2.3	0.03	-1.85	-0.7	0	M
NM_012484	HMMR	2.3	0.01	4.86	1.59	**36**	M
NM_152999	STEAP2	2.2	0.03	2.87	-1.53	**20**	M
NM_002402	MEST	2.2	0.03	-9.56	1.14	**18**	M
NM_172229	KREMEN2	2.2	0.03	0.82	-0.7	0	M
NM_004476	FOLH1	2.1	0.03		-1.18	**31**	M
NM_001042422	SLC16A3	2.0	0.03	-4.18	0.85	**9**	M
NM_013314	BLNK	2.0	0.04	-8	0.73	**2**	M
NM_178865	SERINC2	1.9	0.03	-8.57	-0.39	**14**	M
NM_018423	STYK1	1.9	0.04	-0.52	-2.26	**6**	M
NM_003360	UGT8	1.9	0.03	1.75	-3.1	0	M
NM_001033113	ENTPD8	1.9	0.04		0.79	0	M
NM_002589	PCDH7	1.9	0.03	1.75	-0.05	**8**	M
NM_001466	FZD2	1.8	0.01	-1.68	2.76	**5**	M
NM_004603	STX1A	1.7	0.03	-1.68	2.51	**1**	M
NM_017514	PLXNA3	1.7	0.02	-8.31	1.25	**21**	M
NM_012140	SLC25A10	1.7	0.04	-8	-0.82	0	M
NM_182690	EFNA4	1.7	0.01	-3.57	-0.85	**18**	M
NM_014353	RAB26	1.7	0.03		2.44	**17**	M
NM_033544	RCCD1	1.6	0.03	-5.11	1.85	0	M
NM_004442	EPHB2	1.5	0.01	0.69	1.76	**39**	M
NM_001193472	FOLH1	1.4	0.04		-1.18	**31**	M
NM_017821	RHBDL2	1.4	0.03	5.01	0.86	**8**	M
NM_012378	OR8B8	1.3	0.03			0	M
NM_017767	SLC39A4	1.3	0.02		-1.08	**7**	M
NM_001077263	TMPRSS13	1.3	0.03	-2.64	1.26	**13**	M
NM_003885	CDK5R1	1.2	0.02		2.08	**8**	M
NM_024298	MBOAT7	1.2	0.03	-1.33	-0.32	**8**	M
NM_173517	VKORC1L1	1.2	0.04		-0.7	0	M
NM_178148	SLC35B2	1.1	0.02	-0.35	0.85	**7**	M
NM_030650	KIAA1715	1.0	0.02	-0.34	-0.13	0	M
NM_001146082	MBOAT7	1.0	0.02	-1.33	-0.32	**8**	M
NM_004720	LPAR2	0.9	0.03		0.16	**29**	M
NM_032871	RELT	0.8	0.04	-3.92	1.41	**30**	M
NM_014184	CNIH4	0.8	0.04	3.95	1.43	0	M
NM_031905	ARMC10	0.8	0.03	-0.66	0.11	**7**	M

NM_031905	ARMC10	0.7	0.03	-0.66	0.11	**7**	**M**
NM_002653	PITX1	5.2	0.02	-1.85	-0.64	**13**	**I**
XM_001718065		4.3	0.02				**I**
NM_014736	KIAA0101	3.9	0.01	-3.89	0.99	**25**	**I**
NM_004701	CCNB2	3.9	0.01	-0.43	2.12	**10**	**I**
NM_014791	MELK	3.9	0.02	0.03	0.58	**12**	**I**
NM_016343	CENPF	3.8	0.02	0	3.69	**15**	**I**
NM_031942	CDCA7	3.8	0.02	2.31	3.09	**15**	**I**
NM_018136	ASPM	3.7	0.03	5.39	2.24	**11**	**I**
NM_181803	UBE2C	3.7	0.02	-1.2	3.62	**28**	**I**
NM_005515	MNX1	3.6	0.01		1.77	**12**	**I**
NM_018492	PBK	3.5	0.02	0.94	0.94	**28**	**I**
lincRNA:chr7:7928975-7996750_R		3.4	0.01				**I**
NM_001012271	BIRC5	3.4	0.01	0.09	2.24	**39**	**I**
NM_018410	HJURP	3.3	0.02	-0.88	1.09	0	**I**
NM_001067	TOP2A	3.3	0.03	0.3	3.78	**24**	**I**
NM_003283	TNNT1	3.3	0.03		-0.13	**2**	**I**
NM_013282	UHRF1	3.2	0.02	0.05	2.76	**16**	**I**
NM_001005464	HIST2H3A	3.2	0.02			**2**	**I**
ENST00000397186	LOC645249	3.2	0.02		5.23	0	**I**
NM_006907	PYCR1	3.0	0.03	-4.86	1.42	**13**	**I**
NM_014176	UBE2T	3.0	0.01	-1.29	1.12	**20**	**I**
NM_001827	CKS2	3.0	0.03	2.52	0.85	**35**	**I**
NM_004219	PTTG1	3.0	0.01	3.21	0.67	**69**	**I**
NM_014750	DLGAP5	3.0	0.01	0.98	5.81	**25**	**I**
NM_018131	CEP55	3.0	0.01	0.47	2.14	**23**	**I**
NM_021067	GINS1	2.9	0.04	-0.96	2.14	**14**	**I**
NM_012310	KIF4A	2.9	0.04	-0.26	-1.02	**11**	**I**
NM_018685	ANLN	2.9	0.01	0.53	-0.7	**12**	**I**
NM_003579	RAD54L	2.9	0.01	-3.57	2.18	**14**	**I**
NM_006902	PRRX1	2.9	0.04	-7.59	-0.12	**14**	**I**
NM_020675	SPC25	2.9	0.03	-1.66	0.06	0	**I**
NM_182507	KRT80	2.8	0.04	-8.57	0.09	0	**I**
NM_006845	KIF2C	2.8	0.01	-1.65	1.65	**12**	**I**
NM_003509	HIST1H2AI	2.8	0.01	-6.1		**20**	**I**
ENST00000377836	HES2	2.8	0.03	-0.24	-1.08	**25**	**I**
NM_005346	HSPA1B	2.8	0.04	-1.28	1.72	**13**	**I**
lincRNA:chr2:216585154-216585719_F		2.8	0.01				**I**
NM_016359	NUSAP1	2.8	0.01	1.78	3.14	**8**	**I**
NM_021127	PMAIP1	2.8	0.03	-1.66	-0.7	**35**	**I**
NM_004336	BUB1	2.7	0.02	-0.88	1.41	**24**	**I**

NM_003514	HIST1H2AM	2.7	0.03		2.76	**8**	I
NM_012112	TPX2	2.7	0.03	-0.86	2.31	**19**	I
NM_006101	NDC80	2.7	0.03	1.06	-0.53	**13**	I
NM_032117	MND1	2.7	0.03	-3.62	4.65	0	I
NM_023940	RASL11B	2.7	0.02	-7.03	0.61	**17**	I
A_33_P3223860		2.7	0.03				I
NM_005651	TDO2	2.6	0.03	-7.03	-2.36	**8**	I
NM_020436	SALL4	2.6	0.01		2.76	**20**	I
NM_178229	IQGAP3	2.6	0.01	3.03	2.87	0	I
NM_006461	SPAG5	2.6	0.02	-0.12	1.41	**13**	I
NM_005192	CDKN3	2.6	0.03	0.49	3.49	**28**	I
NM_032899	FAM83A	2.6	0.02	-1.85	0.51	**33**	I
NM_152562	CDCA2	2.6	0.01	0.54	1.24	0	I
NM_020242	KIF15	2.6	0.03	0.24	2.63	**11**	I
NM_080668	CDCA5	2.6	0.02	-9.56	1.79	**5**	I
NM_003537	HIST1H3B	2.5	0.03			**2**	I
NM_006607	PTTG2	2.5	0.01			**50**	I
NM_003504	CDC45	2.5	0.03	-3.2	2.76	**5**	I
NM_001114173	CTSC	2.5	0.02	-0.75	0	**2**	I
NM_024629	MLF1IP	2.5	0.03	-0.88	1.14	**4**	I
NM_002875	RAD51	2.5	0.03	-2.16	1.26	**20**	I
NM_032997	ZWINT	2.5	0.03	-0.96	0.44	**10**	I
NM_005480	TROAP	2.5	0.03	-2.86	6.87	0	I
NM_182687	PKMYT1	2.5	0.03	-4.19	-0.29	**8**	I
NM_198433	AURKA	2.5	0.01	-0.2	3.08	**40**	I
NM_005729	PPIF	2.5	0.03	-4.52	0.38	0	I
NM_004523	KIF11	2.4	0.03	2.13	-0.32	**11**	I
NM_016095	GINS2	2.4	0.02	-0.18	2.06	**12**	I
NM_207006	FAM83A	2.4	0.02	-1.85	0.51	**33**	I
NM_004260	RECQL4	2.4	0.03	-7.03	6.22	**15**	I
NM_001211	BUB1B	2.4	0.01	-2.08	6.33	**22**	I
NM_030928	CDT1	2.4	0.03	-9.3	7.02	**5**	I
ENST00000293743	SPC24	2.4	0.02		2.76	0	I
NM_001002876	CENPM	2.4	0.02	-8.57	1.14	**14**	I
NM_007280	OIP5	2.4	0.04	1.75	4.25	**20**	I
NM_138555	KIF23	2.3	0.03	2.87	0.13	**5**	I
NM_001071	TYMS	2.3	0.03	0.22	-0.08	**23**	I
NM_003686	EXO1	2.3	0.03	3.87	6.33	**6**	I
NM_001024599	HIST2H2BF	2.3	0.03	0.82	-1.11	0	I
A_24_P177553		2.3	0.03				I
NR_024377	FER1L4	2.3	0.03	3.03	-0.23	0	I
NM_018101	CDCA8	2.3	0.01	-0.42	1.52	**8**	I

NM_025108	C16orf59	2.3	0.03	-6.1	0.79	0	I
NM_002263	KIFC1	2.3	0.04	-7.03	6.22	**6**	I
NM_016426	GTSE1	2.2	0.01	-2.64	2.58	**7**	I
NM_001002033	HN1	2.2	0.03	-1.24	1.6	**14**	I
NM_005192	CDKN3	2.2	0.03	0.49	3.49	**28**	I
NM_001238	CCNE1	2.2	0.03	2.72	3.32	**36**	I
NM_002358	MAD2L1	2.2	0.03	-2.44	-1.16	**26**	I
NM_001002032	HN1	2.2	0.03	-1.24	1.6	**14**	I
NM_001114121	CHEK1	2.2	0.02	-0.98	-0.08	**26**	I
NM_003511	HIST1H2AL	2.1	0.02			**7**	I
NM_031966	CCNB1	2.1	0.01	0.33	1.45	**23**	I
NM_014501	UBE2S	2.1	0.04	-5.1	1.32	**14**	I
NM_145697	NUF2	2.1	0.04	-1.65	1.2	**15**	I
NM_138555	KIF23	2.1	0.03	2.87	0.13	**5**	I
NM_032997	ZWINT	2.1	0.03	-0.96	0.44	**10**	I
NM_153696	FOLH1B	2.1	0.03			**29**	I
ENST00000369159		2.1	0.02				I
NM_001136137	RPL28	2.1	0.02	-3.93	0.48	0	I
NM_002497	NEK2	2.0	0.02	-1.17	0.25	**11**	I
NM_170589	CASC5	2.0	0.02	-1.36	0.79	**23**	I
NM_004936	CDKN2B	2.0	0.02	-4.18	-0.29	**25**	I
NM_202002	FOXM1	2.0	0.03	0.19	1.6	**53**	I
NM_004237	TRIP13	2.0	0.03	-1.44	1.02	**5**	I
NM_001024599	HIST2H2BF	2.0	0.02	0.82	-1.11	0	I
NM_002417	MKI67	2.0	0.03	-0.08	3.31	**40**	I
NM_031299	CDCA3	1.9	0.03	-4.14	4.21	0	I
NM_003541	HIST1H4K	1.9	0.04			**8**	I
NM_001237	CCNA2	1.9	0.03	-0.76	7.21	**26**	I
NM_018131	CEP55	1.9	0.01	0.47	2.14	**23**	I
NM_016343	CENPF	1.9	0.04	0	3.69	**15**	I
lincRNA:chr6:86381806-86396056_R		1.9	0.03				I
NM_005225	E2F1	1.9	0.02	-4.54	1.82	**39**	I
lincRNA:chr1:173831852-173837202_F		1.9	0.02				I
NM_005914	MCM4	1.8	0.03	-1.1	1.08	**9**	I
NM_002417	MKI67	1.8	0.03	-0.08	3.31	**40**	I
NM_002863	PYGL	1.8	0.02	-0.23	0.08	0	I
ENST00000425412		1.8	0.03				I
NM_138432	SDSL	1.8	0.04	-4.54	0.07	0	I
NM_020726	NLN	1.8	0.03	-0.14	0.82	**6**	I
NM_021065	HIST1H2AD	1.8	0.01			0	I
NM_024745	SHCBP1	1.8	0.02	0.61	-0.42	0	I

NM_002573	PAFAH1B3	1.7	0.02	-2.23	2.38	**6**	**I**
NM_005052	RAC3	1.7	0.03	-8.98	2.63	**21**	**I**
NM_018012	KIF26B	1.7	0.01		0.6	0	**I**
lincRNA:chr4:15657202-15695627_R		1.7	0.01				**I**
A_33_P3417305		1.7	0.03				**I**
lincRNA:chr4:53579544-53580304_F		1.7	0.04				**I**
NM_001813	CENPE	1.7	0.03	6.27	2.48	**6**	**I**
NM_018645	HES6	1.7	0.02	-8.57	2.57	**20**	**I**
NM_001826	CKS1B	1.7	0.03	1.02	1.08	**18**	**I**
NM_005733	KIF20A	1.7	0.02	-0.08	6.22	**9**	**I**
lincRNA:chr6:86381806-86396056_R		1.7	0.03				**I**
NM_002357	MXD1	1.7	0.03	4.78	-0.43	**20**	**I**
NM_006455	LEPREL4	1.7	0.04	-1.3	0.54	**17**	**I**
NM_001012985	C1orf31	1.7	0.04	1.9	0.87	**8**	**I**
NR_003098	SNHG1	1.7	0.03	-4	0.69	0	**I**
NM_002357	MXD1	1.7	0.03	4.78	-0.43	**20**	**I**
NM_004083	DDIT3	1.7	0.04	-1.19	-0.89	**24**	**I**
XR_037614		1.6	0.01				**I**
NM_002105	H2AFX	1.6	0.03	-4.56	1.54	**19**	**I**
NM_198949	NUDT1	1.6	0.04	-8	1.84	**11**	**I**
lincRNA:chr20:32861439-32866564_R		1.6	0.03				**I**
NR_002734	PTTG3P	1.6	0.03			**25**	**I**
NM_005461	MAFB	1.6	0.01	1.75	0.49	**13**	**I**
lincRNA:chr9:37876125-37881925_F		1.6	0.04				**I**
NM_002497	NEK2	1.6	0.03	-1.17	0.25	**11**	**I**
NM_003655	CBX4	1.5	0.02	-1.68	0.16	**9**	**I**
NM_001079675	ETV4	1.5	0.03	-5.98	7.15	**52**	**I**
BC036909	LOC284889	1.5	0.02	0.82	1.28	**33**	**I**
NM_003513	HIST1H2AB	1.5	0.03			0	**I**
NM_003536	HIST1H3H	1.5	0.03	-2.64		**9**	**I**
NM_182470	PKM2	1.5	0.01	-3.83	1.62	**33**	**I**
NM_001040874	HIST2H2AA4	1.5	0.03			0	**I**
A_33_P3321923		1.5	0.01				**I**
NM_021018	HIST1H3F	1.5	0.03			**13**	**I**
NM_153824	PYCR1	1.5	0.02	-4.86	1.42	**13**	**I**
A_33_P3368636		1.5	0.01				**I**
NM_001039841	ARHGAP11B	1.5	0.01		1.2	0	**I**
NM_001195228	FAM64A	1.5	0.03	-0.88	6.44	0	**I**

NM_017779	DEPDC1	1.5	0.03	-0.82	2.18	0	I
NM_005310	GRB7	1.5	0.03	-6.1	0.66	**37**	I
NM_003536	HIST1H3H	1.5	0.03	-2.64		**9**	I
NM_020183	ARNTL2	1.5	0.04	4.7	0.39	**19**	I
NM_002627	PFKP	1.5	0.02	-1.7	0.9	**9**	I
NM_001009936	PHF19	1.4	0.03	-0.43	2.96	**29**	I
NM_002627	PFKP	1.4	0.03	-1.7	0.9	**9**	I
NM_148170	CTSC	1.4	0.02	-0.75	0	**2**	I
NM_203463	CERS6	1.4	0.03	-1.66	0.45	**21**	I
NM_002046	GAPDH	1.4	0.01	-2.44	1.38	**13**	I
NM_013312	HOOK2	1.4	0.04	-7.03	-0.33	0	I
NM_198175	NME1	1.4	0.03	-0.25	0.82	**47**	I
lincRNA:chr21:30566954-30600954_F		1.4	0.01				I
NM_138433	KLHDC7B	1.4	0.04		-6.06	0	I
NM_024680	E2F8	1.4	0.04		1.51	**15**	I
A_33_P3239184		1.3	0.03				I
NM_032246	MEX3B	1.3	0.03	0.82	1.56	0	I
NM_001039140	C20orf27	1.3	0.03	-8.57	0.76	**7**	I
NM_005208	CRYBA1	1.3	0.03			0	I
NM_005461	MAFB	1.3	0.03	1.75	0.49	**13**	I
lincRNA:chr4:15669186-15683175_R		1.3	0.03				I
NM_152463	EME1	1.3	0.04	-0.74	0.23	**4**	I
NM_024656	GLT25D1	1.3	0.02	2.15	2.12	**13**	I
NM_001039140	C20orf27	1.3	0.02	-8.57	0.76	**7**	I
NM_016545	IER5	1.3	0.03	-0.52	1.09	0	I
NM_003528	HIST2H2BE	1.3	0.03	-8	0.82	**6**	I
NM_001136136	RPL28	1.3	0.03	-3.93	0.48	0	I
NM_003107	SOX4	1.3	0.01	5.98	1.37	**29**	I
lincRNA:chr4:15669186-15683175_R		1.3	0.02				I
NM_018455	CENPN	1.3	0.03	0.75	0.03	0	I
NM_139160	DEPDC7	1.3	0.04	-7.03	-1.47	**33**	I
NM_003524	HIST1H2BH	1.2	0.04			0	I
NM_003521	HIST1H2BM	1.2	0.03			**5**	I
NM_001043229	METTL12	1.2	0.03	3.03	-0.14	0	I
NM_032340	C6orf125	1.2	0.03			0	I
NR_002187	TPI1P2	1.2	0.03	1.75	3.69	0	I
NM_004095	EIF4EBP1	1.2	0.03	-3.21	1.12	**21**	I
NM_007215	POLG2	1.2	0.01	4.5	0.1	**10**	I
NM_003527	HIST1H2BO	1.2	0.03			0	I
NM_003525	HIST1H2BI	1.2	0.03			**9**	I

NM_003530	HIST1H3D	1.2	0.03	-2.64	-2.92	**8**	**I**
A_33_P3241043		1.2	0.01				**I**
NM_006554	MTX2	1.2	0.02	-0.46	-1	0	**I**
NM_003512	HIST1H2AC	1.1	0.02	-4.54	-1.12	0	**I**
ENST00000454548	LOC340340	1.1	0.03			0	**I**
NM_000365	TPI1	1.1	0.03	0.05	-0.12	**4**	**I**
NM_178014	TUBB	1.1	0.03	-2.21	2.25	**6**	**I**
NM_021058	HIST1H2BJ	1.1	0.03			**6**	**I**
NM_004049	BCL2A1	1.1	0.03	-7.03	-0.7	**25**	**I**
XR_017567	LOC647086	1.1	0.03		3.69	0	**I**
NM_001255	CDC20	1.1	0.03	-3.63	1.64	**12**	**I**
NM_001040716	PC	1.1	0.03	-6.1	-1.05	0	**I**
NM_002512	NME2	1.1	0.01	-0.42	-0.26	**43**	**I**
NM_002949	MRPL12	1.1	0.03	-9.15	1.08	**6**	**I**
NM_003519	HIST1H2BL	1.1	0.03			**4**	**I**
A_24_P647682		1.1	0.03				**I**
NM_175856	CHSY3	1.1	0.03		0.23	0	**I**
NM_016535	ZNF581	1.1	0.04	-1.92	-0.7	0	**I**
NM_001003714	ATP5J2	1.1	0.03	2.14	0.69	**7**	**I**
NM_019116	UBFD1	1.1	0.03	0.19	-0.1	0	**I**
A_33_P3265205		1.1	0.03				**I**
NM_001039613	IAH1	1.1	0.02	-1.07	0.08	0	**I**
NM_080596	HIST1H2AH	1.1	0.03			**5**	**I**
NM_000291	PGK1	1.1	0.02	-0.69	0.01	**10**	**I**
NM_004037	AMPD2	1.0	0.03	-4.91	-0.54	0	**I**
lincRNA:chr2:217380530-217421230_F		1.0	0.02				**I**
NM_001069	TUBB2A	1.0	0.03	-3.48	3.41	**3**	**I**
NM_004911	PDIA4	1.0	0.03	-0.72	0.5	**6**	**I**
XR_109904		1.0	0.04				**I**
NM_199285	PRR19	1.0	0.03	-7.03	1.05	0	**I**
NM_003526	HIST1H2BC	1.0	0.03	1.75	-0.7	**13**	**I**
NM_021062	HIST1H2BB	1.0	0.04			0	**I**
NM_004889	ATP5J2	1.0	0.03	2.14	0.69	**7**	**I**
NM_001790	CDC25C	1.0	0.04	-0.88	3.49	**14**	**I**
NM_175065	HIST2H2AB	1.0	0.03			**4**	**I**
NM_003510	HIST1H2AK	1.0	0.02			**5**	**I**
NM_012321	LSM4	1.0	0.04	-3.89	1.44	0	**I**
NR_023920	WT1-AS	1.0	0.04		0.26	**50**	**I**
NM_002949	MRPL12	1.0	0.04	-9.15	1.08	**6**	**I**
NM_016071	MRPS33	1.0	0.03	1.03	0.33	0	**I**
NM_080593	HIST1H2BK	1.0	0.03	-7.59	0.16	**6**	**I**

NR_026878	MGC12982	1.0	0.04	-7.59	2.4	0	**I**
NM_003821	RIPK2	1.0	0.03	-1.06	0.61	**15**	**I**
ENST00000435388	LOC100652764	1.0	0.03	0.82	2.18	0	**I**
NM_021063	HIST1H2BD	1.0	0.03		-1.46	**4**	**I**
ENST00000362079	COX3	0.9	0.03			0	**I**
lincRNA:chr2:216582765-216584147_F	XLOC_001851	0.9	0.03			0	**I**
NM_007208	MRPL3	0.9	0.03	-0.99	0.53	0	**I**
NM_006527	SLBP	0.9	0.03	-0.21	2.08	**3**	**I**
NM_001014812	FAM96A	0.9	0.03	-1.65	-0.21	0	**I**
NM_001039182	BOLA2B	0.9	0.03	-2.96	-0.25	0	**I**
NM_016587	CBX3	0.8	0.03	-1.36	0.16	**6**	**I**
NM_001001436	C16orf87	0.8	0.03	-2.24	-0.62	0	**I**
NM_001039182	BOLA2B	0.8	0.04	-2.96	-0.25	0	**I**
NM_003514	HIST1H2AM	0.8	0.03		2.76	**8**	**I**
NM_153702	ELMOD2	0.8	0.04	-1.09	0.51	0	**I**
NM_013299	SAC3D1	0.8	0.04	0.82	0.18	0	**I**
NM_033118	MYLK2	0.7	0.03			**6**	**I**
lincRNA:chrX:100696094-100725244_R		0.7	0.04				**I**
NM_001008388	CISD2	0.7	0.04	-1.43	-0.3	0	**I**
NM_005984	SLC25A1	0.7	0.04	-9.97	0.1	0	**I**

Table 5.4.1; This table lists the 343 significantly up-regulated genes. The last column has the cellular component based on a combination of gene ontology cellular component, signal sequence and transmembrane domain predictions. The key of classifications is; E = an extracellular gene, M = a membrane bound gene and I = an intracellular gene. 36 genes were found to be extracellular, 45 membrane and the remaining 262 intracellular.

5.5 Combining RNA-seq and microarray data for candidate TEMs

In alternative screening, potential tumour endothelial markers were generated by combining lung array genes with the multi organ RNA-seq contrasts from chapter 4. This time statistical significance was ignored and lung microarray genes chosen with a log2(FC) ≥ 1 enrichment in tumour endothelium. This identified further candidate genes, which are listed in appendix table A.5.5.1. A total of 790 genes were found in this screen, 11 with three RNA-seq contrasts and 146 with two RNA-seq contrasts.

5.6 Summary

In this analysis, genes were identified that were statistically significantly over expressed in tumour endothelial cells from NSCLC samples. It attempted to account for tumour and sample variability to identify the most likely pan-lung tumour vascular targets. The lung screen was also combined with the multi-contrast RNA-seq data from chapter 4 as an extra source of possible vascular targets. Genes found were compared with the other data sets and are discussed in more detail in chapter 6.

5.7 Acknowledgement of contributions

All the bioinformatics analyses, combined RNA-seq analyses and exploratory plots were carried out by the author of this thesis. The author would like to acknowledge and thank Xiaodong Zhuang who performed the experimental work; fresh tumour endothelial cell isolation of the lung samples, microarray hybridizations, QPCR and staining. This work has led to a co-author publication planned for submission to the Oncogene journal, entitled "Molecular profiling of tumour endothelium for the identification novel vascular targets for lung cancer". This work was done in collaboration with Xiaodong Zhuang and Astra Zeneca as part of a KTP project (Darren Cross and James Bradford).

CHAPTER 6; DISCUSION; COMBINING DATA SETS TO FIND CANDIDATE VASCULAR TARGETS

6.1 Introduction

In order to gain a better understanding of the tumour endothelial transcriptome and find vascular targets, 5 analyses of different data sources were carried out in chapters 2, 3, 4 and 5. Vascular targets were sought using a variety genomic gene expression measure techniques, including proteomic, first generation sequenced cDNA libraries, 2nd generation sequenced RNA-seq libraries and microarray. In each of those chapters, the resulting gene lists were not investigated but are investigated here, first separately for the proteomic and RNA-seq data, and then all analyses combined with a comparison to those vascular targets and angiogenesis genes that have already been published. The aim of this chapter is to discuss a selection of genes with published data and to identify some that are possibly worth further experimental validation or possible targeting. Ultimately, it will be stated that this work serves as a starting point for laboratories looking to find novel tumour endothelial or just vascular genes by providing an array of possible candidates from a variety of data sets.

6.2 *In-vivo* biotinylation novelty

In this work *in-vivo* biotinylation and perfusion was combined with mass spectrometry and bioinformatic analyses to deduce the extra cellular matrix and vascular proteome from the chicken embryo. For an embryonic organism, this is the first time it has been performed, which facilitates a description of the ECM and vasculature proteome. To add more gene signature models to the results, this approach was also applied to U87 glioma cells implanted onto a CAM and to granulation tissue as a result of wounds inflicted on the CAM. Although many studies have previously taken place to investigate gene signatures of different chicken systems using proteomics, like studies of the chicken cardiovascular system [144] or analyses of chicken cerebro-spinal fluid [145], these studies did not use the high-resolution proteomic procedures or *in-vivo* capture of blood vessel system proteins or the comprehensive bioinformatics used here.

6.3 High quality proteomic data

High quality proteomic data was attained from the *in-vivo* biotinylation perfusion and proteomic mass-spectrometry analyses, which is evidenced by the high minimum correlation coefficient scores derived for all biological replicates (0.84). From the full proteomic data, all the proteins were searched against the Refseq human proteome to decipher the most likely ortholog of the chicken protein. 1,157 total non-redundant orthologs were assigned for all proteins and 51% of them were extracellular or membrane genes, which is in line with the Rybak et al. *in-vivo* biotinylation study [53].

Although the results show a significant number of intracellular proteins, possibly derived through tissue injury caused by reactive biotin or from cell apoptosis due to experimental procedures, these proteins were characteristic of derived tissues; for example, in the tumour sample there were glycolytic enzymes. PKM2 (pyruvate kinase muscle isozyme) is a glycolytic enzyme that is found to be expressed in cells that are dependent on a high metabolic rate. It has been shown that PKM2 contributes to the Warbug effect in tumour cells, evidenced by the swapping of PKM2 for PKM1 causing loss of the Warburg effect in cancer cells [146]. The Warburg effect in cancer details the different way cancer cells derive their energy. They derive energy through a high metabolic rate of glycolysis in the cytosol, mediated by enzymes like PKM2, which produces stockpiles of lactic acid due to extra fermentation. In contrast, normal cells have a low metabolic rate and perform glycolysis reactions and subsequent oxidation of pyruvate in the mitochondria [147].

6.4 Functional enrichment consistent with embryonic development

A functional enrichment analysis revealed biological processes and molecular pathways consistent with the biology of a developing embryo. It is evident that all tissues from the embryo are enriched for extracellular matrix and cellular adhesion biological processes. Wound healing, blood vessel development and response to wounding are categories also consistent with the granulation tissue and with a developing embryonic system. Focal adhesion pathways were switched on in these samples, suggesting the occurrences of

cellular construction and cytoskeleton building. The functional enrichment analysis was consistent with the biological systems being analysed.

6.5 Differentially expressed genes found between tumour and wound

Malignant tumours have been known to develop at sites of chronic injury and the environment of an injury can catalyse malignant disease [148]. In fact, it was once proposed that tumours are actually wounds that never heal because of the almost identical characteristics between the tumour stroma development and wound granulation tissue formation [149]. Tumours and wounds are known to have many molecular signature similarities and it has been shown that inflammation is a key step in both processes, as is angiogenesis. For instance, there was a study that found 77% of genes were concordant between renal regenerating cells (similar to the wound healing model) and renal cell carcinoma (a mimic of the tumour model). The other 23% of genes changed in the opposite directions [150]. In this study we performed CAM models of wound and tumours, and, thus, it was possible to elucidate the gene expression differences between these two models. The analyses revealed 16 genes enriched in the tumour model and 18 genes enriched in the wound. For a subset of these genes, an investigation of published literature was taken to assess their relevance to either model. It will attempt to address the question; Are these molecular differences characteristic of the differences between tumours and wounds?

6.6 Tumour enriched NAMPT; evidence for tumour and angiogenesis involvement

NAMPT is the HUGO gene symbol of the enzyme Nicotinamide phosphoribosyltransferase. This enzyme is critically important in the biosynthesis of nicotinamide-adenine-dinucleotide (NAD). This enzyme catalyzes 5-phosphoribosyl-pyrophosphate and nicotinamide condensation to produce the new 2 products of nicotinamide mononucleotide and inorganic pyrophosphate. This enzyme is rate limiting for the NAD synthesis salvage pathway. NAMPT was the most tumour/CAM enriched gene (64 fold up-regulated) versus the wound/CAM samples. In the literature, NAMPT

has previously been shown to be up-regulated ≥ 6 fold in colon cancer at both mRNA and protein levels [151, 152]. In addition, it has also been shown to be up-regulated in glioma and has been proposed as a serum prognostic marker of glioblastoma [153]. Additional evidence is that NAMPT could also have a role in angiogenesis. A Korean group has published several papers relating to the adipokine Visfatin (which is another name for NAMPT but this isoform is proposed to be extracellular) and its involvement in angiogenesis. An example is Kim et al. who showed in 2007 that NAMPT promotes angiogenesis and is a potent activator of *in-vivo* neovascularization of the chick chorioallantoic membrane [154]. There is now evidence that Notch1-dependent production of FGF2 is induced in NAMPT-primed endothelial angiogenesis and several groups have shown NAMPT/Visfatin can influence angiogenesis [154-158]. Lastly, NAMPT has been shown to induce proliferation and tube formation in Human Umbilical Vein Endothelial Cells (HUVECs) in a dose moderating manner [159]. Additionally a compound, FK866, has been reported to inhibit NAMPT and starve tumour cells that rely on NAD metabolism for energy. This is a potential therapy for cancer [159]. Combining the multiple publication sources of evidence for NAMPT induced angiogenesis and over-expression in the U87/CAM tumour model, gives compelling evidence that NAMPT is a marker of glioma and angiogenesis in the CAM tumour model.

6.7 S100P was found to be a tumour enriched gene

S100 calcium binding protein P (S100P) was 4 fold higher in its expression in glioblastoma versus granulation tissue. The P in S100P stands for placenta as this protein was first found in the placenta but it has also been found to be expressed in the fibroblasts, in the cell line 3T3, where it was shown to activate the inflammatory pathway of NFκB [160]. It has subsequently been found to be expressed in several tissues, including expression in different brain tumours [161]. There is other compelling evidence that this gene is involved in several cancers and cell proliferation. According to the Atlas of Genetics and Cytogenetics in Oncology and Haematology [162], the entry for S100P gene has many articles demonstrating S100P over expression in cancer cells and it is reported to be linked to the progression of a number of cancer types that include, breast, lung prostate, colon and pancreatic cancers [163-167]. From these studies, a relationship

between over expression of S100P and different tumour like phenotypes was noticed. These included elevated survival, resistance to treatment and incidence of metastasis. The Koltzscher group noted that an intracellular form of S100P could increase metastasis, cell growth and survival upon binding to and activating the gene EZR (Ezrin) [168]. EZR itself is implicated in cancer cells by affecting cell surface structure adhesions, cellular migration and cell order. An extracellular form of S100P has been shown to bind the receptor of advanced glycation end products (RAGE) [169], which can also effect cancer. S100P is also a target of glucocorticoid signalling in tumour cells [170]. In summary, there is a significant amount of published evidence linking S100P over expression to cancer and it appears that this is the first time the over expression of S100P has been reported in a CAM model of glioblastoma.

6.8 OAS3 tumour and angiogenesis gene?

In a study where epithelial cells were stimulated with Amyloid-B protein to mimic the effects of Age-related macular degeneration (AMD), which includes inflammation and pathological angiogenesis [171], the gene 2'-5'-oligoadenylate synthetase 3 (OAS3) was found to be 4 fold up-regulated on stimulation. β-amyloid peptide has also been used in an Alzheimer's model, where it was shown to induce angiogenesis several days after introduction [172]. This gene has also been shown to be highly expressed during early glioma tumour development prior to a fully vascularised tumour, prior to the angiogenic switch [173]. Another study showed up-regulation of OAS3 after multiple doses of radiation in glioma and breast cancer cells [174]. β-amyloid induced angiogenesis shows OAS3 over expression, which could be a link to over expression in the tumour cam model. However, angiogenesis is also a characteristic of wound healing, so it is not clear why OAS3 would be over expressed in glioma compared to the wound. It should be noted that OAS3 was shown to be up-regulated on day 2 in vascularisation of the tumour CAM model [174].

6.9 TNFAIP6 tumour endothelial gene

This gene is involved in angiogenesis/inflammation and has already been found to be up-regulated in a CAM tumour model using HIF1A primed Small Cell Lung Carcinoma

cells [175]. There is also a patent for this gene as a tumour endothelial marker of Ovarian cancer, where they are targeting the gene with an inhibitory compound [176]. TNFAIP6 has recently been shown to play a regulatory role in FGF2 mediated angiogenesis by up-regulating angiogenesis through its mopping up binding to PTX3; since this reduces the availability of PTX3 to bind to FGF2, which inhibits angiogenesis [175]. As TNFAIP6 can induce angiogenesis, then it is feasible this gene has a role in neovascularisation of gliomas in the CAM, which is exemplified by the fact that it has also been reported as a tumour endothelial marker [177, 178], albeit in ovarian cancer.

6.10 CD276, found over-expressed in the wound

CD276 was previously identified as a tumour endothelial marker in mouse using SAGE library analyses from endothelial cells that were isolated from tumour tissues using antibody coupled magnetic beads [71]. In humans, the CD276 protein has been identified to be expressed by trophoblasts in the placenta, to act as a decoy for maternal leukocytes, which ultimately assists in preventing foetal immune attack from the mothers immune system [179]. A recent publication shows that the endothelium of tumours have an elevated expression of CD276 and that cancer cells themselves also express this molecule. Its expression in cancer is associated with a lower survival rate [180]. Taking this evidence together, it makes some sense that this molecule could help a tumour deactivate the immune system (causing a more potent tumour) as it does in the placenta to keep a mother's immune system from attacking the foetus. However, with these results showing the over expression of CD276 in the wound versus the tumour, this suggests it is helping prevent the immune system from causing excessive scaring or attacking other molecules that are only expressed in developing tissues, such as in wound healing.

6.11 CALCRL was found to be a wound enriched gene

Calcitonin receptor-like protein (CALCRL) was found to be 64 fold up-regulated in wound model versus the tumour model. This gene is part of the Adrenomedullin receptor complex that is involved in vasodilation. This gene has also been shown to be up-regulated in hypoxia, which in conjunction with Adrenomedullin, can enhance endothelial cell survival and migration [181]. Another publication shows that CALCRL

has an effect of increasing vessel permeability and can facilitate the migration of white blood cells [182], which gives a feasible role for CALCRL in wound healing as white blood cells are delivered to a wound to substantiate an immune response and stimulate macrophages to perform phagocytosis of necrotic tissue.

6.12 H1F0 was found to be a wound enriched gene

Retinoic acid (Vitamin A) is an absolute prerequisite for normal wound healing (epithelialisation, immune system function and production of proteoglycans) [183-185]. The mouse H1 histone family, member 0 (H1F0) is up-regulated in mouse in response to retinoic acid [186]. Therefore, there is a possible link between the up-regulation of H1F0 in wound granulation tissue versus the tumour sample.

6.13 EPB41 was found to be a wound enriched gene

EPB41 has been shown to have a pivotal role in wound healing in mice. It was demonstrated that EPB41 knockout mice show a phenotype of wound healing impairment [187]. In the same study, it was also shown that keratinocytes with silenced EPB41 exhibited reduced adhesion, cell spreading, migration and motility. In addition, a gene called DLG1, has been reported to promote wound healing and it is proposed that the transglutiminase EPB41 binds to DLG1 [188].

6.14 Recent literature on discovering VT genes

To discuss vascular targets in the proteomic and other data sets, first a list of publications relating to the identification of endothelial, tumour endothelial, angiogenic and vascular target genes was collected. Although it is not possible to investigate all related previous publications, a selection was chosen that spanned the last 12 years of research into identifying endothelial genes and vascular targets, which, if totalled, is 883 non-redundant genes. Some genes are classed as both endothelial and tumour endothelial genes, depending on research type (tumour endothelial = 244, endothelial = 815). This includes 2 reviews that evaluate the field and the history, thus providing good overviews of what the state of play is today. The publications of previous vascular, endothelial and tumour endothelial enriched genes are listed in table 6.14.1.

Table 6.14.1 Publications on vasculature and tumour vasculature gene discoveries

Publication Title	Year	TEMs	Endothelial or general angiogenesis	Total
Stan, R. V., L. Ghitescu, et al. (1999). "Isolation, cloning, and localization of rat PV-1, a novel endothelial caveolar protein." J Cell Biol 145(6): 1189-1198.	1999	**0**	1	1
St Croix, B., C. Rago, et al. (2000). "Genes expressed in human tumor endothelium." Science 289(5482): 1197-1202.	2000	**25**	25	50
Huminiecki, L. and R. Bicknell (2000). "In silico cloning of novel endothelial-specific genes." Genome Res 10(11): 1796-1806.	2000	**2**	35	37
Ho, M., E. Yang, et al. (2003). "Identification of endothelial cell genes by combined database mining and microarray analysis." Physiol Genomics 13(3): 249-262.	2003	**0**	64	64
Chi, J. T., H. Y. Chang, et al. (2003). "Endothelial cell diversity revealed by global expression profiling." Proc Natl Acad Sci U S A 100(19): 10623-10628.	2003	**0**	45	45
Madden, S. L., B. P. Cook, et al. (2004). "Vascular gene expression in nonneoplastic and malignant brain." Am J Pathol 165(2): 601-608.	2004	**5**	16	21
Rybak, J. N., A. Ettorre, et al. (2005). "In vivo protein biotinylation for identification of organ-specific antigens accessible from the vasculature." Nat Methods 2(4): 291-298.	2005	**9**	0	9
Oh, P., Y. Li, et al. (2004). "Subtractive proteomic mapping of the endothelial surface in lung and solid tumours for tissue-specific therapy." Nature 429(6992): 629-635	2004	**2**	0	2
Neri, D. and R. Bicknell (2005). "Tumour vascular targeting." Nat Rev Cancer 5(6): 436-446.	2005	**41**	0	41
van Beijnum, J. R., R. P. Dings, et al. (2006). "Gene expression of tumor angiogenesis dissected: specific targeting of colon cancer angiogenic vasculature." Blood 108(7): 2339-2348.	2006	**17**	48	65
Seaman, S., J. Stevens, et al. (2007). "Genes that distinguish physiological and pathological angiogenesis." Cancer Cell 11(6): 539-554.	2007	**13**	51	64
Herbert, J. M., D. Stekel, et al. (2008). "A novel method of differential gene expression analysis using multiple cDNA libraries applied to the identification of tumour endothelial genes." BMC Genomics 9: 153.	2008	**9**	457	466

Bhati, R., C. Patterson, et al. (2008). "Molecular characterization of human breast tumor vascular cells." Am J Pathol 172(5): 1381-1390.	2008	**13**	15	28
Wallgard, E., E. Larsson, et al. (2008). "Identification of a core set of 58 gene transcripts with broad and specific expression in the microvasculature." Arterioscler Thromb Vasc Biol 28(8): 1469-1476.	2008	**0**	58	58
Pasqualini, R., B. J. Moeller, et al. (2010). "Leveraging molecular heterogeneity of the vascular endothelium for targeted drug delivery and imaging." Semin Thromb Hemost 36(3): 343-351.	2010	**15**	0	15
Sasaroli, D., P. A. Gimotty, et al. (2011). "Novel surface targets and serum biomarkers from the ovarian cancer vasculature." Cancer Biol Ther 12(3): 169-180. **Buckanovich, R. J. and an earlier paper by the same group; D. Sasaroli, et al. (2007). "Tumor vascular proteins as biomarkers in ovarian cancer." J Clin Oncol 25(7): 852-861.**	2011	**93**	0	93

Table 6.14.1; This tables lists a selection of publications that span the last 12 years of research involved in identifying tumour endothelial and vascular endothelial genes using an assortment of methods. There are over 800 genes in this list when sorted non-redundantly. They serve as a guide to validation of the putative targets found in this work. It should be noted that endothelial genes, though not necessarily tumour endothelial, will be used as evidence for expression in endothelial cells. Some studies were performed in mice (E.g. Seaman et al. and Wallgard et al.); nevertheless, this table used human orthologs for those mouse genes if they exist. Note, Sasaroli et al. and Buckanovich et al. are a merged data set as their works are from the same group and have the same purpose.

6.15 A sweet shop of vascular genes

The proteomic analyses from chapter 2 was preferentially to acquire membrane and extracellular proteins from a developing chicken embryo. Since angiogenesis is a process pivotal to both embryonic development and tumour growth, a comparison to angiogenic and endothelial library screens was performed to rank human orthologs as potential vascular targets. 205 genes were enriched in endothelium alone and 136 genes had a positive fold change in both endothelial and angiogenic screens. Several genes among these are ANGPT2, ANGPTL2, ANXA5, COL15A1, EPHB1, EPHB2, IGFBP7, IL31RA, JAM3, DKK3, MMP15, PLXNA2, PLXNA4, PTX3, SULT1E1, SPARC, PXDN, COL4A1, NID1 and VIM. These are very encouraging results since some are already published as tumour endothelial markers or vascular targets. For instance the genes COL4A1, SPARC, VIM and IGFBP7 were reported by van Beijnum et al. as tumour angiogenesis genes [68]. And the genes PXDN, PLXNA2, NID1 and DKK3 were reported as endothelial markers of brain or colon tumours [69, 73]. Therefore, in this list there could be other potential tumour endothelial genes.

Another notable fact is that 204 genes in the proteomic data have previously been published as endothelial or tumour endothelial genes, some of them without the cDNA library confirmation. This includes some established markers of endothelium: PECAM1, KDR, VWF, MCAM, CDH5, CD93, TIE1, THBS1, TEK, NRP1, LYVE1, PROCR and LAMA4. Similarly, there are also other published tumour endothelial markers. These include: COL4A2, THY1, MMP2, TEM5, TEM8, COL1A1, INSR and TNC [68, 69, 73, 77, 132, 177, 178]. In summary, there are a multitude of endothelial and tumour endothelial genes here that provide clear proof that this chicken embryonic model system is a great source of vascular targets, either in conjunction with other data sets, like cDNA libraries, or alone. Looking down the list of genes, ones without publications from table 6.14.1, another gene is quickly found. Plexin D1 (PLXND1) is a membrane gene that, when looked up, has also been identified as a marker of tumour endothelium and, in addition, has been used to mark tumour endothelium when targeted with antibodies for imaging [189]. This one data source is a great resource of vascular genes and it would be

very interesting to analyse data from more samples of CAM implanted tumour cells from different tumour types.

6.16 Deep sequencing for the first time on fresh tumour endothelium

One massive challenge in tumour endothelial biology is to acquire a pure population of endothelial cells. Endothelial cells are embedded amongst an ensemble of different cell types. Antibody coupled bead isolation has been performed with different positive and negative selection markers. St. Croix used MCAM as a positive selection marker following negative selection of non-endothelial cell types with anti EPCAM, CD45, CD64 and CD14 [69]. van Beijnum et al. used anti-PECAM1 and anti-CD34 to isolate endothelium and Madden et al. anti-MCAM and Ulex europaeus agglutinin 1 (UEA-1) [68, 73]. Bhati et al. used micro-dissection on breast samples to isolate endothelial cells [77]. In this work UEA-1 alone was used as a positive magnetic bead selection marker for fresh endothelial cell isolates from bladder, colon and lung tumour and normal tissues prior to deep sequencing (chapter 4). Human hepatic sinusoidal endothelial cells (HSEC) were immunomagnetic selected for using anti-PECAM1, which was used in the RPKM analyses (liver endothelial cells, 24hrs at 1% oxygen).

New technologies, like Illumina Hiseq sequencers, give this experiment the opportunity to discover new genes or alternative isoforms not yet available in public databases. Although the human genome is sequenced, it is not known whether all gene transcripts have been found. Annotation relies on gene prediction programs and the presence of cDNA sequences in databases like dbEST. Both microarrays and traditional transcriptomic techniques such as cDNA and SAGE libraries are limited. Microarrays are limited by the fact that you have to have knowledge of the sequence before it can appear on the chip and traditional transcriptomics by the problems of cost of sequencing and generation time. Now these limitations have been removed by the advent of new technologies such as Illumina or 454 that can generate millions of bases for several thousand pounds, in a short time frame [190]. In this project, we have utilised new

technologies, Roche 454, Illumina and SOLiD4 to decipher new vascular targets and construct an endothelial transcriptome.

RNA-seq and/or deep sequencing has already been used in many studies, including transcriptomes of novel species and various cell types. However, not many have been reported sequencing endothelial cells [191-193]. Here, deep sequencing has been performed for the first time on fresh isolated endothelium from bladder, colon and lung tumours, as well as on tumour conditioned HUVEC as a pure population of endothelial cells. These analyses provide genes that could be pan vascular targets across multiple organ tumours.

6.17 pan-VT genes, markers of multiple tumour endothelium types

The data in chapter 4 was presented in a series of 6 views, ranking genes with 2 or more contrast results with other sources of evidence. Here, some of the candidate genes from the different ranks will be investigated and discussed in terms of how likely the gene is related to angiogenesis and endothelial biology.

When looking at genes that were found in more than a single contrast, GPR143 is an interesting result. This gene is associated with being expressed in the eye and has a role in nystagmus and ocular albinism, which is surprising as this gene was found in 2 contrasts of bladder and colon tumour endothelium. It was also found expressed in endothelial cells but at a low level from the cDNA library results. There is a tenuous link between GPR143 and angiogenesis. The L-3,4-dihydroxyphenylalanine (L-Dopa) factor is able to induce angiogenesis. In the rat model of Parkinson disease, a disease that is characterised by an increase in number of endothelial cells around neurons [194], L-Dopa was found to induce endothelial cell proliferation and angiogenesis [195]. The tenuous link comes from the fact that L-Dopa has also been shown to be a ligand for the GPR143 receptor [196]. However, the aforementioned angiogenic effect in the rat model, was caused by L-Dopa binding to dopamine receptor D1 rather than GPR143.

Four jointed box 1 (Drosophila) (FJX1) was also found in multiple contrasts (lung and bladder tumour endothelium) and this gene has been reported to be a tumour endothelial marker of Ovarian cancer [177]. It has also been established that this gene is targeted downstream in the notch signalling pathway [197]. This makes some logical sense as delta-like 4 (Drosophila) (DLL4) was also found enriched in bladder and lung in this analysis, providing supporting evidence for Notch signalling during tumour endothelium activation. The Dll4 and Notch signalling pathway is important in sprouting angiogenesis as it tightly co-ordinates endothelial tip and stalk cell differentiation [198].

Sphingosine-1-phosphate receptor 3 (S1PR3) or EDG3 is a receptor that is activated through the lipid ligand molecule Sphingosine 1-phosphate (S1P), which itself is an initiator of angiogenesis. S1P stimulates EDG3 and activates endothelial cells into a proliferative and migratory state [199, 200]. In the RNA-seq data sets, this gene was found enriched in tumour endothelium from two organs (lung and colon).

Dystrobrevin, alpha (DTNA) was found to be up-regulated in both lung and bladder tumour endothelium. Research shows that A-catulin (CTNNAL1) is important in vascular remodelling in the vascular system and that DTNA has a role in transporting A-catulin to signalosomes, which enhances GPCR signalling of the α(1D)-Adrenergic receptors. α(1D)-Adrenergic receptor itself is reported to induce angiogenesis following binding of the phenylephrine ligand in low oxygen conditions [201]. In addition, A-catulin itself has been implicated to have a role in wound healing [202]. The combination of these genes and the published research provides a clue as to how DTNA could be involved in tumour angiogenesis.

Gaorav et al. showed that the multi-contrast genes: MMP1 (tumour HUVEC, lung and colon), MMP2 (tumour HUVEC and lung) and EREG (lung and colon) are mediators of tumour angiogenesis [203]. There is strong evidence that endothelial cells are expressing both MMP1 and MMP2 but this is not so for EREG as there are no ESTs in the EREG Unigene cluster from a vascular tissue source.

6.18 Endothelial Specific Molecule 1; pan vascular target of multiple tumours from deep sequencing

Endothelial Specific Molecule 1 (ESM1) was one of the best performing markers of tumour endothelium in this work. It has been published in 5 of the publications used as benchmarks in this chapter (table 6.14.1) and was found to be enriched in all 3 human tumour organ types as well as the tumour conditioned HUVEC sample, plus the RPKM and cDNA analyses. From these results, ESM1 appears to be a pan vascular target and it also has a high AngioScore of 68%. There is extensive literature linking ESM1 to angiogenesis and different cancers, so the result in this work is not a surprise but serves as a positive control that the analyses have found relevant genes [204-213].

6.19 BAMBI; possible pan vascular target of multiple tumours from deep sequencing

BMP and activin membrane-bound inhibitor homolog (BAMBI) was found in lung and bladder tumour endothelium, which makes it a possible pan-vascular target. BAMBI has been shown to interact with Frizzled5 (Wnt receptor), in conjunction with LRP6 (co-receptor) and DVL2, which culminates in the signalling and translocation of beta-catenin to the nucleus. The beta-catenin/Wnt pathway enhances cell proliferation and is important for embryonic growth and tumour progression. BAMBI has already been shown to inhibit the TGF-B signalling pathway by acting as a decoy pseudo-receptor but it has also been identified as a target for beta-catenin in liver and colon tumour cells [214]. BAMBI has an endothelial cell specific gene expression profile [215] and in BAMBI knockout mice, there is not an obvious phenotype, except that the endothelial cells of the kidney and heart are prominent and swollen. In response to TGFB, inflating the expression of BAMBI, reduced capillary growth and migration, whereas the opposite was seen if BAMBI was knocked down. BAMBI knockdown caused an increase of free TGFB, which stimulated phosphorylation of the genes SMA1/5 and ERK1/2 [215].

6.20 Combining multiple analyses

There have been several studies that have identified organ specific tumour endothelial markers. For instance, van Beijnum et al. and St. Croix et al. both found markers of colon cancer [68, 69], Bhati et al. markers of breast cancer [77] and Sasaroli et al. markers of ovarian cancer [178]. In this work, deep sequencing of multiple organ fresh isolated endothelium was performed for the first time and the results were combined with that of multiple experiments: proteomic, tumour conditioned HUVEC, public cDNA libraries and microarray data. There were 5 different data sets and analyses in the earlier chapters that aimed to find tumour endothelial genes. A short letter code, bold type face in the list, will be used to refer to each of the different analyses and the key to these is given below, along with the number of genes found.

- *in-vivo* biotinylation data (chapter 2) = full embryonic data = **PROT** = 1,157 genes
- cDNA library (chapter 3) endothelial and angiogenic screens (Extracellular and Membrane) = **cDNA** = 698 genes
- RPKM analyses (chapter 4) of tumour conditioned HUVEC + low oxygen liver endothelial cells versus combined normal tissues = **RPKM** = 753 genes
- RNA-seq multi-contrast (chapter 4) of fresh tumour endothelial cells versus normal endothelial cells = **NGS** = 3,922 genes
- Microarray of 4 biological replicates (chapter 5) of fresh lung tumour endothelium vs. normal lung endothelium data = **LUNG** = 284 genes

For all analyses, the combined non-redundant list = 5,892 genes. An AngioScore analyses of all genes produces 162 genes that have a score ≥ 50. The number of genes that are predicted to be membrane or containing a transmembrane domain is 1,240, those predicted to be extracellular is 706 and the remaining 3,946 are intracellular. A table with all genes combined with the number of publications, identity of analyses, AngioScores and product information can be found in appendix table 6.20.1. Figure 6.20.1 shows the overlap counts between the different analyses. Not a single gene was found in all five

analyses, with the following found in multiple analyses: 7 genes found in four analyses, 101 genes found in three analyses and 699 genes found in two analyses. The remaining 5,085 genes were found in a single data set. One possible criterion for selecting genes could be to only consider those genes found in more than one analysis, which would give a total 807 genes.

Figure 6.20.1 Overlap between the different data sets

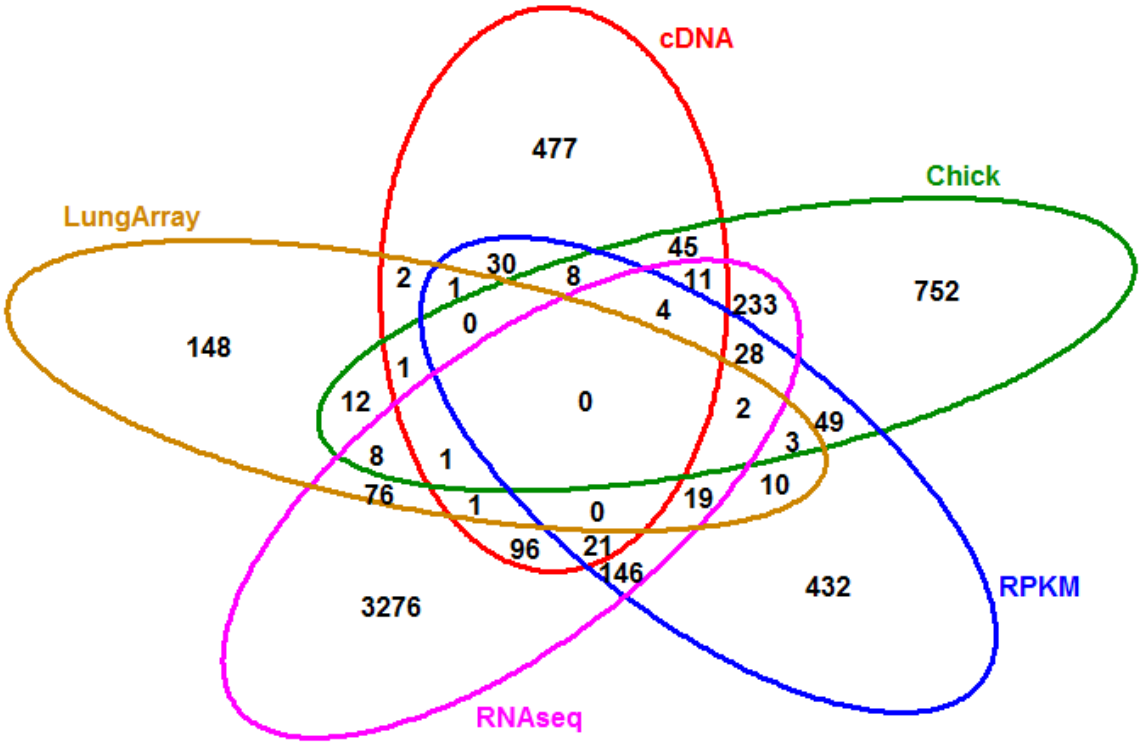


Figure 6.20.1; This figure shows overlaps between the five different data set analyses carried out in this work. In total there were 5,892 genes found to be enriched in tumour endothelial cells from all five analyses combined. Not a single gene was found in all five analyses, with the following found in multiple analyses: 7 genes found in four analyses, 101 genes found in three analyses and 699 genes found in two analyses. The remaining 5,085 genes were found in a single data set. The labels and corresponding data sets' and chapters are: chick = Proteomic data from chapter 2, cDNA = cDNA library results from chapter 3, RPKM = chapter 4, the RPKM analyses of endothelial cells versus bulk normal tissue results, RNAseq = the multi organ tumour endothelial results from chapter 4 (RNA-seq) and LungArray = the lung array results from chapter 5.

6.21 Combined data finds published tumour endothelial genes

When comparing all the genes found in this work, from all analyses, with published genes from table 6.14.1, 104 genes have been reported to be endothelial or tumour endothelial in 2 or more publications. Similarly, the total number of genes reported as endothelial or tumour endothelial in any number of publications is 495. The different data sets can be broken down into a percentage of total genes published from each data set: PROT = 18%, NGS = 7%, LUNG = 7%, RPKM = 18% and cDNA = 13%. A rough measure of efficiancy would suggest that the proteomics and RPKM analyses performed best for finding known target genes in this work. About 6 thousand genes were found in this work and this represents ~24% of the transcriptome (based on transcriptome estimates [216]). Although endothelial cells have been reported to be very transcriptionally rich [217], it is likely that some genes found in these data sets will be passengers, either expressed due to other cell contaminants or expressed due to individual differences. False positives are an important consideration as choosing the wrong genes can lead to unnecessary experimental work.

In contrast, it is surprising that 388 genes published to be endothelial and/or tumour endothelial were not found here. Serum deprivation response (SDPR), RAS interacting protein 1 (RASIP1) and RhoGEF and PH domain containing 5 (FGD5) were found in multiple publications [62, 63, 218, 219] but were not found in this study. This is possibly due to the focus on tumour endothelial gene discovery in this work, rather than the emphasis on endothelial gene discovery in these publications.

A list of the 7 genes that were found in 4 different analyses is as follows: angiopoietin 2 (ANGPT2), peptidylprolyl isomerase A (PPIA), Pentraxin-related protein 3, (PTX3), EGF containing fibulin-like extracellular matrix protein 1 (EFEMP1), endoglin (ENG), Ephrin type-B receptor 2 (EPHB2) and triosephosphate isomerase 1 (TPI1). Four of these genes have publications (from table 6.14.1) relating to endothelial cell expression (ANGPT2, EFEMP1, ENG and PPIA). Although there were no reports from the benchmark publications for TPI1, it has been reported as a marker of renal cell carcinoma and also to be up-regulated in response to sokotrasterol sulfate (an apparent angiogenesis

promoting steroid) [220, 221]. EPHB2 is reported to regulate angiogenesis and be sensitive to both VEGF and hypoxia. It is up-regulated at sites of neovascularisation [222-224].

Pentraxin-related protein 3 (PTX3) is an ECM gene found enriched in 4 different analyses. This is a somewhat surprising result as this gene is reported to sequester fibroblast growth factor 2 (FGF2) by interacting with and inhibiting the molecule; FGF2 is a potent pro-angiogenic growth factor. However, it is also reported that the effects of PTX3 can be quenched in the presence of another gene, Tumor necrosis factor-inducible gene 6 (TSG-6), which binds to PTX3 in the same position as FGF2. Leali et al. propose this competitive binding between FGF2-PTX3 and TSG-6-PTX3 offers a fine control mechanism for FGF2 mediated angiogenesis [225]. Tumour angiogenesis is a hallmark of tumour growth, and with PTX3 been over expressed in 4 analyses, it is possible that TSG-6 is also over expressed. However, TSG-6 was only found in the embryonic proteome data. It is possible that tumour angiogenesis is not using specifically the FGF2 angiogenic pathway or it could be that the body is attempting to quell the tumour pathological angiogenesis by elevating the expression of PTX3 over TSG-6 to commandeer active FGF2. Alternatively, it could be that FGF2 mediated angiogenesis is utilized more in the chicken embryo, hence the dual presence of PTX3 and TSG-6. Alternatively, PTX3 could share similar promoter machinery with FGF2 and therefore be always co-expressed with FGF2. It should be noted that TSG-6 was up-regulated in the tumour versus wound as reported in an earlier chapter.

ANGPT2 is a late stage angiogenesis gene, which has been shown to orchestrate angiogenesis using the integrin and TEK signalling pathways and is involved in vessel maturation. ANGPT2 acts in an autocrine way (produced mainly by ECs) to destabilize vessels. Depending on cell type, it is able to activate or inhibit the TEK receptor [226] and its purpose is to act as an Angiopoietin 1 factor and TEK receptor negative regulator, which ends up controlling the level of response endothelial cells have to cytokines [227].

Matrix metalloprotease-1 (MMP1) was identified as a general angiogenesis gene by van Beijnum et al. using subtractive hybridisation and microarray analyses [68] and was later found to be highly expressed in endothelial cells [62]. It has been shown that stimulation of G protein coupled receptor protease activated receptor 1 (PAR1) by MMP1 promotes angiogenesis and metastasis in mouse models of ovarian cancer, whereby it compels cancer cells to release pro-angiogenic molecules such as Interleukin 8. This in turn leads to increased endothelial cell proliferation and sprouting. This form of angiogenesis is resistant to Avastin treatment and has been proposed as an alternative pathway target for treating ovarian cancer [228]. In total, 14 MMP genes were found in this work, 7 of which have been reported in the literature as being markers of angiogenesis (MMP2, MMP1, MMP11, MMP9, MMP3, MMP14, MMP10 [62, 68, 77]

6.22 Lysyl oxidase-like 2; vascular target candidate

807 genes were found in more than one experiment from all data sets combined. Of these, 166 are published as endothelial or TEM genes. For instance the lysyl oxidase-like 2 (LOXL2) gene is not published as a TEM but is found in 3 deep sequencing contrasts as well as from the PROT and RPKM analyses (3 different analyses). This family of secreted genes are usually involved in the cross-linking of the extracellular matrix. Although there are published links to cancer [229-232], with an AngioScore of 6 there is little current evidence of LOXL2 as a tumour endothelial marker but it is encouraging that it is found to be involved in sprouting angiogenesis; Bignon et al. performed a proteomics analyses on the ECM of endothelial cells grown in hypoxia (angiogenic state) [233] and report that LOXL2 is expressed in neovessels and amasses in the ECM. They also state that Morpholino knockdown of the LOXL2 ortholog in zebrafish leads to a loss of intersegmental vessels during development, and, that it is a pivotal enzyme in the assembly of ECM. They conclude that LOXL2 regulates angiogenesis in conjunction with collagen IV during ECM construction. This one publication provides encouraging evidence of LOXL2 being a tumour endothelial marker required for tumour vessel development, and coupled with the multi-contrast deep sequencing and RPKM/PROT results, gives compelling evidence for LOXL2 to be a vascular target.

6.23 Fibronectin leucine rich transmembrane protein 2; vascular target candidate

Fibronectin leucine rich transmembrane protein 2 (FLRT2) was found in 3 analyses; it was up-regulated in the tumour conditioned HUVEC, the cDNA library analyses and in the chicken embryo. Haines et al. reported that this family of genes regulate FGF2 signalling by interacting with Fibroblast growth factor receptor 1 (FGFR1) [234]. Magnuson et al. used Embryoid body and teratomas model systems to show that FGFR1 has an inhibitory effect on endothelial cells, where inhibition of FGFR1 led to increased vascularisation [235]. FGFR1 deficient teratomas (originating from stem cells without FGFR1) also showed an increased rate of vessel growth and a distinct morphology. Magnuson et al. then used antibody targeting to show that the cytokines interleukin-4 and pleiotrophin have a negative and positive angiogenic effect on FGFR1 respectively and proposed this was the system that controls FGFR1 angiogenesis. FLRT2 and other family members have been shown to interact with FGFR1 and it has been proposed that the FLRT genes regulate FGF2 signalling through FGFR1 [234]. It is possible this gene plays a role in embryonic development as it was found in the chicken embryo but not found in any fresh tumour endothelial cell isolates. Nevertheless, this gene could be worth validating.

6.24 Angiopoietin-like 4; vascular target candidate?

Although differences have been noted between cultured endothelial cells versus fresh isolates [68, 69], for this gene we have both tumour HUVEC and bladder tumour endothelium, plus the chicken embryo and RPKM analyses. Angiopoietin-like 4 (ANGPTL4) appears to be an excellent candidate based on its expression profile here and most notably because it has been shown to be orchestrated by hypoxia inducible factor 1 (HIF1A), which is a gene that induces VEGF secretion and drives angiogenesis under low oxygen conditions [236]. This is a bit surprising though as this gene has also been classed as an adipokine, being expressed in liver and fatty tissues, and so not restricted to tumour endothelium. It is possible that this is related to the dual domain protein structure of ANGPTL4, which has an N-terminal coiled-coil domain and a C-terminal fibrinogen like domain. It has been shown that the domains can be cleaved and released to apply

distinct biological effects. In a recent review, spanning 10 years of ANGPTL4 research, this gene has been found to have a variety of functions involved in wound healing, metastasis, angiogenesis, energy homoeostasis and redox regulation [237]. One concern in this work is that ANGPTL4 has little public EST evidence of endothelial cell expression and, if it is controlled by HIF1A, it is likely be expressed in tumour or perivascular cells rather than endothelium.

6.25 MCT1 another positive control target found

Monocarboxylate transporter 1 (MCT1 or SLC16A1) was found in the tumour conditioned HUVEC, lung and the PROT data sets. Recent research has shown that lactate saturation in normoxic endothelial cells causes the up-regulation of HIF-1alpha, which in turn leads to an increase in KDR and bFGF expression in the endothelial cells. The lactate transporter gene, MCT1, was then knocked down in endothelial cells and this inhibited lactate intake into endothelial cells. This ultimately abolished lactate induced Hif1-alpha tumor angiogenesis and has now been proposed as a new anti-angiogenic therapy [237].

6.26 What was new in this work?

- A new *in-vivo* biotinylation procedure used in an embryonic system for the first time, coupled with high-resolution proteomics and bioinformatics analyses to provide a list of vascular target genes.
- 2nd generation sequencing (Roche 454) was combined with public cDNA libraries to find vascular targets.
- RPKM analyses of an *in-vitro* tumour conditioned endothelial cells and hypoxia liver endothelium versus normal bulk adult human tissues was performed to produce a third list of targets.
- Deep sequencing RNA-seq data of freshly isolated tumour endothelium from multiple organs was provided for the first time to illuminate the tumour endothelial transcriptome and reveal pan-vascular-targets.
- Finally, 4 biological replicate tumour lung and normal lung endothelium were used in a microarray analyses to find perpetual lung tumour endothelial markers.

- 5,892 vascular target candidates were identified, with 1,240 membrane and 706 extracellular genes.

6.27 Summary and outlook

Solid tumours remain one of the most deadly and most common diseases in man in the world today and there is an urgent need for new treatments. The classical treatments are limited by off-target delivery of therapeutics to a tumour and by the fact that the hypoxic environment of tumour cells can lead to resistance, through hyper-mutation. Surgical removal remains uncomfortable and there is a risk of relapse. Therefore, there are multiple reasons for finding more vascular targets: 1) the inefficiency of conventional treatments and 2) the expense of current anti-angiogenics, as emphasised in the introduction. There are many therapeutics currently being researched and some of them are approved by the FDA but ultimately, there is not one treatment that fits all tumour types. It is hoped that using multiple gene expression sources from a variety of model systems will lead to the identity of new molecules that can be used for vascular targeting of multiple tumours.

Bioinformatics is only as good as the biological data it is given. Judging by the number of endothelial and tumour endothelial genes found in the data that are already published, the data is of high quality. All data sets analysed found published vascular related genes to some degree. Ultimately, using the latest deep sequencing technologies on fresh tumour endothelial cells, leads to novel gene expression signatures and this is a source of new vascular targets. Likewise, the customised *in-vivo* biotinylation embryo technique of the Bikfalvi group combined with high-resolution mass-spectrometry has also supplied another high quality data set that contains many classic endothelial and tumour endothelial genes. The ultimate take home message is that researchers or companies from anywhere should take some markers and validate them, which will hopefully result in patient benefits in the future.

6.28 Serving the scientific community

The full list of 5,892 genes, together with annotation, will be available online as a downloading Excel file (if the thesis is published). The endothelial transcriptome assembled from multi deep sequencing experiments will also be available as a FASTA text file. **Please note, there is an accompanying Appendix volume containing all appendix tables (also downloaded at the following link; http://sara.molbiol.ox.ac.uk/userweb/jherbert/tissue_diffex/appendix_files.html).**

Some of this work was from "A new procedure for determining the genetic basis of a physiological process in a non-model species, illustrated by cold induced angiogenesis in the carp." [107]. The methods of AngioScore and endothelial cell expression signatures have been used by other researchers to identify genes most likely involved in angiogenesis and/or endothelial expressed from their data sets. The Pubmed identification numbers of these are listed: PMID: 20840761 [238], PMID: 19924294 [239], PMID: 21706054 [240], PMID: 20080097 [241] and PMID: 22103505 [242].

CHAPTER 7; METHODS

7.1 Principle component analyses of proteomic data

Protein abundance index values were imported as a table, using the read.delim function into R. The data was then logged to base 2 and then quantile normalized in R using the preprocessCore module [137] from Bioconductor [243]. The code for running a PCA in R was;

```
# import data matrix (either full data or most differentially expressed genes)
pca=prcomp(t(Data_matrix_normalized_data), scale=T)
mycolors <- c("red", "blue")
plot(main="Tumour and normal lung", pca$x, type="n");
text(pca$x, rownames(pca$x), col=mycolors[c(rep(1,4), rep(2,4))])
```

Images were copied into Microsoft PowerPoint to produce thicker borders and some colour customisations of symbols.

7.2 Correlation coefficient analyses

Correlation coefficient values were calculated using the cor function from R, with the default setting of Pearson coefficients. The returned coefficient values were copied into Microsoft Excel and a heatmap was generated.

7.3 Human ortholog analyses of chicken proteins

Protein sequences were downloaded from Uniprot [244] using the chicken accession numbers supplied from the proteomics group. The downloaded proteins were then BLAST [112] searched against the database of human proteins using a Refseq database of human proteins. If a best hit was significant, at a minimum of 1e-9, then that human protein was assigned the human ortholog of the chicken protein. The BLAST command line was run with the -m = 8 option and tabulated results were imported into Microsoft Excel for the making of tables.

7.4 Generating normal distributions of proteomic data and heatmaps

Quantile and logged Protein Abundance Index value normal distributions were generated using the R function hist. Images of the distributions generated by hist were copied into Microsoft PowerPoint. To generate a heat maps of tissue specific genes, first a Microsoft Excel file was manipulated by hand to select genes required in the heat map, using gene symbol as the gene name. This was then imported into R using the read.delim function. The heatmap.2 function was used to generate the heat map images using this data, which were copied directly in to the thesis.

7.5 Differential protein expression from the different samples

Linear Models for Microarrays (Limma) [108] from Bioconductor was used to find differentially expressed genes. A moderated T-test was used and it required the data to be normally distributed. The data followed an approximate normal distribution and Limma analyses was carried out as follows: First the data was logged to the base 2 and then quantile normalized using the preprocessCore module [137] from Bioconductor [243]. A data matrix with those normalized values was run as a single colour differential gene expression analysis as previously described, both as an individual differential gene expression experiment as well as using the decideTests function [108, 245]. The volcano plot was made using the in-built Limma function, volcanoplot.

7.6 Gene ontology analyses and literature searches

The online DAVID resource was used to analyse gene lists for functional enrichment. This bioinformatic method clusters genes with alike functions, which were defined using gene ontology biological process and KEGG pathways [114].

7.7 Identify human proteins from U87 model

Human and chicken proteins were downloaded from the Uniprot database. Perl regular expressions were used to match mass-spectrum peptides from the U87 implanted sample against both human and chicken Uniprot proteins. If a peptide matched a human protein exactly and not a chicken protein, this was counted as a specific human peptide.

7.8 cDNA library analyses

All methods were performed as published [62], with the exception of megablast instead of BLAST for the searches of EST to genes. A major part of the cDNA library analyses involved the collection and processing of data derived from the public domain. This required a relational database system that was made using MySQL. The database was pivotal to most analyses and the related programs can be downloaded from the article that was published during this work (please see the online additional files 30 to 36 from the manuscript [62]). The purpose of these programs was to manipulate the data and import data like genes and EST assignments into the database. Basic information about cDNA libraries was stored in the database, such as sources and accession numbers. Please see additional file 37 for a design of the database schema [62]. Genbank was searched and files were collected for relevant cDNA libraries [246]; these were imported into the database. mRNA sequences from the Refseq database project [247] were downloaded and related information imported into the database.

This screen combined 1st and 2nd generation sequenced mRNA libraries from proprietary and public data. Firstly, endothelial cells isolated from liver using immuno-magnetic beads were subject to Roche 454 titanium sequencing, which produced ~800,000 sequences between 200 and 600bp in length. This library was added to the 31,114 endothelial ESTs already collected from Genbank. This was the endothelial pool. A pool of non-endothelial libraries was collected from Genbank, those chosen by studying the CGAP library browser for non-endothelial cell characteristics. The total number of EST sequences for the non-endothelial was 136,336. ESTs from both pools were assigned to a gene symbol using megablast, searching each sequence against a Refseq of mRNA for human database. The best human gene hit from the mRNA database was assigned the hit. Each output was parsed with a Perl program in preparation for the differential gene expression experiments. Differential expression experiments were carried out online using the website that was created for this work (methods from [62] and site at [248]). The table of results returned from the website was manipulated in Microsoft Excel. The angiogenic screen was run in a similar manner.

7.9 Isolation of Human Umbilical Vein Endothelial Cells

- Fresh umbilical cords
- 70% Ethanol
- PBS (Phosphate buffered saline)
- HUVEC media with pen/strep and amphotericin B: Cells were grown in M199 medium (Cancer Research UK Central Services, Clare Hall) supplemented with bovine brain extracts [35], 10 % foetal calf serum (FCS, PAA Cell Culture Company, Ontario, USA), 4 mM glutamine (Cancer Research UK Central Services, Clare Hall), 90 μg/ml of heparin (Sigma, Poole, UK) and optional antibiotics (Gentamicin/AmphotericinB, TCS Cell Works, Buckingham, UK).
- Collagenase 1A (Sigma C2674-1g) 10x stock at 10 mg/ml in 1ml and filter sterilise if too turbid.
- Biohazard yellow bucket
- Metal dissecting tray (Fisher - DKB-350-030S)
- paper towels
- 20 ml syringes
- 2 canulae attached to tubing
- 2 plastic clamps (VWR - 229-0588)
- cable ties (RS components Ltd - 543-428)
- mixing needles (Henley Medical Supplies - MN5)
- scissors

The reagents and materials above were prepared and the air flow cabinet sterilized. The cord was cut in a bag to leave the placenta. Cords were placed in a tray and 6 inches of cord was enough for a 1 x T25 flask of cells. The cords were cleaned on the outside by wiping with tissue and checked for damage. Clean ends were cut at right angles to the cord. An olive canula was inserted into one end of the cord and sealed with a cable tie. PBS was then injected into the cord to remove any possible clots. Flushing with PBS was continued until it was clear. The other end of the cord was now canulated the same as the other end. Clamps were then attached to each end but left open. Collagenase in M199 but

with no serum was syringed into the cord until the canula at the other end of the cord contained the liquid. The clamps were tightened and the cords placed in a tissue culture incubator for 15 to 20 minutes at 37^{o}C. PBS was flushed through the cord and the contents collected into a 50 ml falcon tube. A further 20ml of air and then another10ml of PBS was flushed through. This step was repeated. During this procedure, the syringe was kept highest and the cord tilted lower, downwards towards the falcon tube. The cells in the falcon tube were centrifuged at 1200 rpm for 6 minutes. The supernatant was discarded and the cells re-suspended vigorously 20 times using a glass Pasteur. Cells were then plated onto a gelatine coated flask. Media was changed every two hours until 80% confluent.

7.10 creating tumour conditioned media at pH 6.8

Custom made Dulbecco's Modified Eagle Medium (Cancer Research UK Central Services, Clare Hall). No Sodium Pyruvate, no glucose and no Sodium Bicarbonate. 10% foetal calf serum (FCS, PAA Cell Culture Company, Ontario, USA), 5mM glutamine (Cancer Research UK Central Services, Clare Hall) and VEGF at 5ng/ml (PEPROTECH, cat: 100-20) were added to the media. To gain a pH of 6.8 10mM PIPES (piperazine-N,N′-bis(2-ethanesulfonic acid, cat no. 5625-37-6 Sigma) was added and the media adjusted using NaOH or HCL.

7.11 Normoxic media at pH 7.4

Dulbecco's Modified Eagle Medium (Cancer Research UK Central Services, Clare Hall) with 4.5g/L of glucose, Sodium Pyruvate, 10% fetal calf serum (FCS, PAA Cell Culture Company, Ontario, USA), 5mM glutamine (Cancer Research UK Central Services, Clare Hall) and no VEGF was made up. 10mM HEPES 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (cat no. 7365-45-9 Sigma) was added and the pH was adjusted to 7.4 using NaOH or HCL.

10cm tissue culture plates of early passaged HUVECs containing both media were incubated for 24hours at 1% and 4% oxygen respectively. The pH was tested to check the

media pH did not change overnight. After this was confirmed, the passage 0 HUVECs were grown for 24 hours in the conditions stated above.

7.12 Isolating RNA from the HUVEC and 2nd generation sequening

The Qiagen RNEasy kit was used in accordance with the manufacturers' instructions. Total RNA was sent to the Genepool sequencing service, Edinburgh, where 3 lanes of Illumina GII RNA-seq deep sequencing was performed. This was the tumour conditioned HUVEC sample.

7.13 Quantitative PCR DLL4 and control gene FLOT2

cDNA was amplified from 2 μg of RNA isolated from the HUVEC cells, using the High Capacity cDNA Archive Kit (Applied Biosystem, Foster City, USA). This was performed according to the manufacturers' instructions.

The Exiqon system (Vedbaek, Denmark) was employed for Quantitative PCR and performed on the Rotor-Gene RG-3000 (Corbett Research Limited, Sidney, Australia). Each reaction contained 10 ul of cDNA in a 25 μl final volume as follows: 12.5 μl qPCR Mastermix (Abgene, Epsom, UK), 1 μl of the forward and the reverse primer (10 μM), 0.25 μl of an Exiqon probe and 0.25 μl of deionised sterile water. Samples were run with the following procedure: 95°C for 10 minutes, followed by 40 cylces at 95°C for 15 seconds and 60°C for 45 seconds. Results of the QPCR were analysed with the Rotor-Gene 6 software (Corbett Research Limited).

QPCR primers and probes were defined at the online Roche design site.
https://www.roche-applied-science.com/sis/rtpcr/upl/adc.jsp
Overview of the design of QPCR primers for DLL4 with transcript NM_019074:

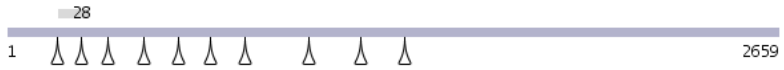
Left Primer	24	178 - 201	60	42	tgttgtggttccgactataaacag
Right Primer	20	266 - 285	60	50	gggctgcaacgtcataatct

Amplicon (108 nt)

tgttgtggttccgactataaacagtacgtgtttggcggctgggcctgggcctggtggtgt
atctccgacactcagaggatttccctagagattatgacgttgcagccc

Download pack insert | PDF report | Text report | Order probes or set

Transcript overview:



Detailed view:

Overview of the design of QPCR primers for FLOT2 with transcript NM_004475:

7.14 RNA-seq deep sequencing mapping reads

The Reads Per Kilobase of exon model per Million reads mapped (RPKM) analyses was carried out as follows. First sequences of normal adult human tissues were downloaded from the Sequence Read Archive database for accession numbers: Lung = SRR014265, heart = SRR014263, liver = SRR014264, brain = SRR036966, cerebral cortex = SRR014262 and muscle = SRR014266. This was the normal tissue pool. Sequences from the Roche 454 liver endothelium (as used in the cDNA libraries analyses) was combined with the tumour conditioned HUVEC to be the endothelial pool. To determine fragment sizes of sequenced fragments, bowtie 0.12.7 [102] was run in paired end mode and parsed to establish fragment lengths and length standard deviations of the different samples. Following this, all reads were mapped to the human genome (version HG19) using Tophat version 1.3.1, using the standard commands detailed in the manual [103, 127].

An example command is;

tophat --library-type fr-unstranded -G RefseqHG19_gtf_gene -g 20 -o 131modifOut -r 1 --mate-std-dev 80 -p 16 --closure-search --no-novel-indels --butterfly-search --coverage-search --solexa1.3-quals hg19_genome lane_5_f,lane_6_f,lane_7_f lane_5_r,lane_6_r,lane_7_r

This command mapped the tumour HUVEC reads and the normal human adult tissues were run the same way. The program htseq-count [249] was then used to assign raw counts to genes in conjunction with a Refseq GTF file that was downloaded from UCSC genome browser (as guided in the manual [249]). Raw gene counts were copied to Excel.

The normal adult human tissues were combined. For each individual sequencing library, a measure of RPKM was calculated which used the average transcript length for a gene. If a gene had 200 reads mapping to it, an average transcript length of 1Kb and a total of 2 million reads were mapped, this would give an RPKM value of 100 RPKM. This was worked out for all genes in the transcriptome and for all samples separately. The endothelial pool (tumour HUVEC and liver endothelium) was averaged for RPKM and likewise, the normal human adult tissues were calculated separately and averaged. Log to base 2 fold changes were calculated between endothelium and normal tissues to derive the final results.

Tophat mapping version 1.3.1 was also used to map the fresh isolated tumour and normal endothelial cell samples as detailed above. The output of Tophat is a BAM file, which is an indexed file used to store sequence read alignments. These BAM files for all samples were assembled into transcriptomes separately using Cufflinks (version 1.0.3 [104, 127, 130]). The file RefseqHG19_gtf is a gene transfer format file that stores exon positions in the genome, which programs like Tophat and Cufflinks understand and use.
A command line used was as follows;

```
cufflinks -o ./tumour_lung_20m_saet -p 16 -g RefseqHG19_gtf gene -M
hg19_rRNA_mask.gtf -b hg19.fa -u --library-type fr-secondstrand -N -L
Colon_30mill_normal -q colon.bam
```

The transcriptomes of all assemblies were merged together using Cuffmerge (part of the Cufflinks suite of tools) as shown in the published protocol [127]. The resulting merged transcriptome was used for the final transcriptome assembly in the results (this includes the novel transcripts and novel isoforms). The transcript sequences can be extracted using the "gffread" utility, which is part of the Cufflinks suite.

Finally, differential gene expression was carried using Cuffdiff on all the genome mapped files (BAM files). There were no biological replicates and the command that found differentially expressed genes was;

```
cuffdiff -o All_results -M hg19_rRNA_mask.gtf -b hg19.fa -u -N -L tum_lung,norm_lung,tum_colon,norm_colon,tblad,thuvec, -p 16 tlung.bam nlung.bam tcolon.bam ncolon.bam tbladd.bam thuvec.bam > all_results
```

The resulting Cuffdiff tab delimited "gene_exp.diff" output files were opened and manipulated in Microsoft Excel to produce the different tables of results.

The isoform switching candidates were identified from Cuffdiff output file " cds_exp.diff". Gene symbols were sorted in Excel and genes for multiple different protein ids (multiple coding proteins), a search was performed to find proteins of different lengths that appeared reciprocally expressed between tumour and normal.

7.15 Linear models for microarray analyses

Microarray analyses was carried out using the Linear Models for Microarrays exactly as listed in the Matticks guide [245]. Tables of differentially expressed genes were exported in table format to be viewed and manipulated in Microsoft Excel.

7.16 Principle component analyses of biological replicates

Principle component analyses was carried out on the full data set and subsequently the most differentially expressed genes from Limma using code as earlier;

```
# import data matrix (either full data or most differentially expressed genes)
pca=prcomp(t(Data_matrix_normalized_data), scale=T)
mycolors <- c("red", "blue")
plot(main="Tumour and normal lung", pca$x, type="n");
text(pca$x, rownames(pca$x), col=mycolors[c(rep(1,4), rep(2,4))])
```

The PCA results were copied into the thesis directly from R.

7.18 AngioScore

The programs to perform literature scanning were published during the time of this thesis (see Herbert et al. [107]). A list of gene symbols is provided to the program and a list of keywords relating to angiogenesis is searched against all abstracts assigned to that gene by the Entrez Gene database. This is done to generate an "AngioScore" for all differentially expressed genes. An "AngioScore" is calculated as the percentage of abstracts containing a given keyword divided by the total number of abstracts for that gene (only from publications that Entrez Gene at Genbank assign to a gene). A list of the keywords used is as follows "vascular", "tumour", "tumor", "angiogenic", "angiogenesis", "neovascularization", "neovascularisation", "vasculogenesis", "hypoxia", "endoth" and "VEGF". The total number of abstracts is returned, as well as the number of abstracts matching a keyword (abstract only counted once if multiple keywords present). In addition, the Pubmed ID numbers for the abstracts is also returned. The results of the literature scanning are imported into Microsoft Excel and an AngioScore percentage is calculated = number of abstracts containing a desired keyword divided by the total number of abstracts for that gene, multiplied by a hundred.

CHAPTER 8; PAPERS PUBLISHED

8.1 First author publications

3 first author publications were published during the time of this thesis and are listed below.

8.2 cDNA library analyses from chapter 2 based on;

Herbert JM, Stekel D, Sanderson S, Heath VL, Bicknell R. A novel method of differential gene expression analysis using multiple cDNA libraries applied to the identification of tumour endothelial genes. BMC Genomics. 2008 Apr 7;9:153. PubMed PMID: 18394197

BMC Genomics IMPACT FACTOR 4.07

Search BMC Genomics for Go

Advanced search

Home | Articles | Authors | Reviewers | About this journal | My BMC Genomics

Top
Abstract
Background
Results
Discussion
Conclusion
Methods
Abbreviations
Authors' contributions
Acknowledgements
References

Research article — Highly accessed — Open Access

A novel method of differential gene expression analysis using multiple cDNA libraries applied to the identification of tumour endothelial genes

John MJ Herbert[1], Dov Stekel[2], Sharon Sanderson[3], Victoria L Heath[1] and Roy Bicknell[1,3]*

* Corresponding author: Roy Bicknell r.bicknell@bham.ac.uk

Author Affiliations

1 Cancer Research UK Angiogenesis Group, Institute for Biomedical Research, University of Birmingham Medical School, Edgbaston, BIRMINGHAM, B15 2TT, UK
2 Centre for Systems Biology, School of Biosciences, University of Birmingham, Edgbaston, BIRMINGHAM, B15 2TT, UK
3 Cancer Research UK, Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, OXFORD, OX3 9DS, UK
For all author emails, please log on.

BMC Genomics 2008, **9**:153 doi:10.1186/1471-2164-9-153

The electronic version of this article is the complete one and can be found online at: http://www.biomedcentral.com/1471-2164/9/153

Received: 15 November 2007
Accepted: 7 April 2008
Published: 7 April 2008

BMC Genomics
Volume 9

Viewing options
Abstract
Full text
PDF (950KB)
Additional files

Associated material
PubMed record
About this article
Readers' comments

Related literature
Cited by
on Google blog search
Other articles by authors
on Google Scholar
on PubMed
Related articles/pages
on Google
on Google Scholar
on PubMed

8.3 The bioinformatics method of the AngioScore, ortholog searches and endothelial cell expression signatures are from this publication;

Herbert JM, Buffa FM, Vorschmitt H, Egginton S, Bicknell R. A new procedure for determining the genetic basis of a physiological process in a non-model species, illustrated by cold induced angiogenesis in the carp. BMC Genomics. 2009 Oct 23;10:490. PubMed PMID: 19852815

BMC Genomics IMPACT FACTOR 4.07

Search BMC Genomics for Go

Advanced search

Home | Articles | Authors | Reviewers | About this journal | My BMC Genomics

Top
Abstract
Background
Results and Discussion
Conclusion
Methods
Comparison wi...
Abbreviations
Authors' contributions
Acknowledgements
References

Methodology article

Open Access

A new procedure for determining the genetic basis of a physiological process in a non-model species, illustrated by cold induced angiogenesis in the carp

John MJ Herbert[1], **Francesca M Buffa**[2], **Henrik Vorschmitt**[1], **Stuart Egginton**[3] and **Roy Bicknell**[1]*

* Corresponding author: Roy Bicknell r.bicknell@bham.ac.uk

Author Affiliations

1 Cancer Research UK Angiogenesis Group, Institute for Biomedical Research, Schools of Immunity and Infection and Cancer studies, College of Medicine and Dentistry, University of Birmingham, Birmingham, B15 2TT, UK
2 Cancer Research UK, Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, OX3 9DS, UK
3 Division of Medical Sciences, Medical School, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK
For all author emails, please log on.

BMC Genomics 2009, **10**:490 doi:10.1186/1471-2164-10-490

The electronic version of this article is the complete one and can be found online at: http://www.biomedcentral.com/1471-2164/10/490

Received: 5 May 2009
Accepted: 23 October 2009
Published: 23 October 2009

BMC Genomics
Volume 10

Viewing options
Abstract
Full text
PDF (930KB)
Additional files

Associated material
PubMed record
About this article
Readers' comments

Related literature
Cited by
on Google blog search
Other articles by authors
on Google Scholar
on PubMed
Related articles/pages
on Google
on Google Scholar
on PubMed

8.4 A book chapter was produced that describe methods to performing analyses with public or proprietary cDNA library data through a web browser and with the help of a text guide;

Herbert JM, Stekel DJ, Mura M, Sychev M, Bicknell R. Bioinformatic methods for finding differentially expressed genes in cDNA libraries, applied to the identification of tumour vascular targets. Methods Mol Biol. 2011;729:99-119. PubMed PMID: 21365486

8.5 co-author publications

15 co-author publications are listed here, most of the methods applied in this work have assisted these other publications.

1: Williams TD, Turan N, Diab AM, Wu H, Mackenzie C, Bartie KL, Hrydziuszko O, Lyons BP, Stentiford GD, Herbert JM, Abraham JK, Katsiadaki I, Leaver MJ, Taggart JB, George SG, Viant MR, Chipman KJ, Falciani F. Towards a system level understanding of non-model organisms sampled from the environment: a network biology approach. PLoS Comput Biol. 2011 Aug;7(8):e1002126. Epub 2011 Aug 25. PubMed PMID: 21901081; PubMed Central PMCID: PMC3161900.

2: Bailey RL, Herbert JM, Khan K, Heath VL, Bicknell R, Tomlinson MG. The emerging role of tetraspanin microdomains on endothelial cells. Biochem Soc Trans. 2011 Dec;39(6):1667-73. Review. PubMed PMID: 22103505.

3: Mura M, Swain RK, Zhuang X, Vorschmitt H, Reynolds G, Durant S, Beesley JF, Herbert JM, Sheldon H, Andre M, Sanderson S, Glen K, Luu NT, McGettrick HM, Antczak P, Falciani F, Nash GB, Nagy ZS, Bicknell R. Identification and angiogenic role of the novel tumor endothelial marker CLEC14A. Oncogene. 2012 Jan 19;31(3):293-305. doi: 10.1038/onc.2011.233. Epub 2011 Jun 27. PubMed PMID: 21706054.

4: Soulet F, Kilarski WW, Antczak P, Herbert J, Bicknell R, Falciani F, Bikfalvi A. Gene signatures in wound tissue as evidenced by molecular profiling in the chick embryo model. BMC Genomics. 2010 Sep 14;11:495. PubMed PMID: 20840761; PubMed Central PMCID: PMC2996991.

5: Dumartin L, Quemener C, Laklai H, Herbert J, Bicknell R, Bousquet C, Pyronnet S, Castronovo V, Schilling MK, Bikfalvi A, Hagedorn M. Netrin-1 mediates earlyevents in pancreatic adenocarcinoma progression, acting on tumor and endothelial cells. Gastroenterology. 2010 Apr;138(4):1595-606, 1606.e1-8. Epub 2010 Jan 18. PubMed PMID: 20080097.

6: Javerzat S, Franco M, Herbert J, Platonova N, Peille AL, Pantesco V, De Vos J, Assou S, Bicknell R, Bikfalvi A, Hagedorn M. Correlating global gene regulation to angiogenesis in the developing chick extra-embryonic vascular system. PLoS One. 2009 Nov 17;4(11):e7856. Erratum in: PLoS One. 2010;5(3). PubMed PMID: 19924294; PubMed Central PMCID: PMC2774277.

7: Verissimo AR, Herbert JM, Heath VL, Legg JA, Sheldon H, Andre M, Swain RK, Bicknell R. Functionally defining the endothelial transcriptome, from Robo4 to ECSCR. Biochem Soc Trans. 2009 Dec;37(Pt 6):1214-7. PubMed PMID: 19909249.

8: Maloney SL, Sullivan DC, Suchting S, Herbert JM, Rabai EM, Nagy Z, Barker J, Sundar S, Bicknell R. Induction of thrombospondin-1 partially mediates the anti-angiogenic activity of dexrazoxane. Br J Cancer. 2009 Sep 15;101(6):957-66. PubMed PMID: 19738618; PubMed Central PMCID: PMC2743367.

9: Sheldon H, Andre M, Legg JA, Heal P, Herbert JM, Sainson R, Sharma AS, Kitajewski JK, Heath VL, Bicknell R. Active involvement of Robo1 and Robo4 in filopodia formation and endothelial cell motility mediated via WASP and other actin nucleation-promoting factors. FASEB J. 2009 Feb;23(2):513-22. Epub 2008 Oct 23. PubMed PMID: 18948384.

10: Protty MB, Watkins NA, Colombo D, Thomas SG, Heath VL, Herbert JM, Bicknell R, Senis YA, Ashman LK, Berditchevski F, Ouwehand WH, Watson SP, Tomlinson MG. Identification of Tspan9 as a novel platelet tetraspanin and the collagen receptor GPVI as a component of tetraspanin microdomains. Biochem J. 2009 Jan 1;417(1):391-400. PubMed PMID: 18795891; PubMed Central PMCID: PMC2652832.

11: Ahmed F, Steele JC, Herbert JM, Steven NM, Bicknell R. Tumor stroma as a target in cancer. Curr Cancer Drug Targets. 2008 Sep;8(6):447-53. Review. PubMed PMID: 18781891.

12: Armstrong LJ, Heath VL, Sanderson S, Kaur S, Beesley JF, Herbert JM, Legg JA, Poulsom R, Bicknell R. ECSM2, an endothelial specific filamin a binding protein that mediates chemotaxis. Arterioscler Thromb Vasc Biol. 2008 Sep;28(9):1640-6. Epub 2008 Jun 12. PubMed PMID: 18556573.

13: Legg JA, Herbert JM, Clissold P, Bicknell R. Slits and Roundabouts in cancer, tumour angiogenesis and endothelial cell migration. Angiogenesis. 2008;11(1):13-21. Epub 2008 Feb 9. Review. PubMed PMID: 18264786.

14: Senis YA, Tomlinson MG, García A, Dumon S, Heath VL, Herbert J, Cobbold SP, Spalton JC, Ayman S, Antrobus R, Zitzmann N, Bicknell R, Frampton J, Authi KS, Martin A, Wakelam MJ, Watson SP. A comprehensive proteomics and genomics analysis reveals novel transmembrane proteins in human platelets and mouse megakaryocytes including G6b-B, a novel immunoreceptor tyrosine-based inhibitory motif protein. Mol Cell Proteomics. 2007 Mar;6(3):548-64. Epub 2006 Dec 23. PubMed PMID: 17186946; PubMed Central PMCID: PMC1860054.

15: Watson SP, Herbert JM, Pollitt AY. GPVI and CLEC-2 in hemostasis and vascular integrity. J Thromb Haemost. 2010 Jul;8(7):1456-67. Epub 2010 Mar 25. Review. PubMed PMID: 20345705.

8.6 FORTHCOMING PUBLICATIONS

Mapping the extracellular and membrane proteome associated with the vasculature and the stroma in the embryo

Fabienne Soulet a,b*,Witold W Kilarski a,b*, Florence Roux-Dalvaic,d*, John M J Herberte*, Izabella Sacewicza,b, Emmanuelle Mouton-Barbosac,d, Roy Bicknelle, Patricia Lalore, Bernard Monsarrat c,d and Andreas Bikfalvi a,b

Molecular profiling of tumour endothelium for the identification novel vascular targets for lung cancer

Xiaodong Zhuang1, John Mathew Herbert1,4, James Bradford2, Alice Turner3, David Thickett5, Babu Naidu6, David Blakey2, Simon Barry2, Darren Cross2 and Roy Bicknell1*

8.7 APPENDIX

Please note, the appendix is in an accompanying volume, or as a download http://sara.molbiol.ox.ac.uk/userweb/jherbert/tissue_diffex/appendix_files.html; thank you.

BIBLIOGRAPHY

1. Hawkins RA, O'Kane RL, Simpson IA, Vina JR: **Structure of the blood-brain barrier and its role in the transport of amino acids.** *J Nutr* 2006, **136:**218S-226S.
2. Yamada E: **The fine structure of the renal glomerulus of the mouse.** *J Biophys Biochem Cytol* 1955, **1:**551-566.
3. **Inflammation, the adhesion cascade** [http://bme.virginia.edu/ley/]
4. Patan S: **Vasculogenesis and angiogenesis as mechanisms of vascular network formation, growth and remodeling.** *J Neurooncol* 2000, **50:**1-15.
5. Gerhardt H: **VEGF and endothelial guidance in angiogenic sprouting.** *Organogenesis* 2008, **4:**241-246.
6. Iruela-Arispe ML, Davis GE: **Cellular and molecular mechanisms of vascular lumen formation.** *Dev Cell* 2009, **16:**222-231.
7. Dejana E, Tournier-Lasserve E, Weinstein BM: **The control of vascular integrity by endothelial cell junctions: molecular basis and pathological implications.** *Dev Cell* 2009, **16:**209-221.
8. **Assay of angiogenesis** [http://www.amplab.de/angiogenesis.html]
9. **Angiogenesis** [http://en.wikipedia.org/wiki/Angiogenesis]
10. **With Angiogenesis, Tumor Growth Proceeds** [http://www.cancer.gov/cancertopics/understandingcancer/angiogenesis/page10]
11. Karamysheva AF: **Mechanisms of angiogenesis.** *Biochemistry (Mosc)* 2008, **73:**751-762.
12. Millauer B, Wizigmann-Voos S, Schnurch H, Martinez R, Moller NP, Risau W, Ullrich A: **High affinity VEGF binding and developmental expression suggest Flk-1 as a major regulator of vasculogenesis and angiogenesis.** *Cell* 1993, **72:**835-846.
13. Seetharam L, Gotoh N, Maru Y, Neufeld G, Yamaguchi S, Shibuya M: **A unique signal transduction from FLT tyrosine kinase, a receptor for vascular endothelial growth factor VEGF.** *Oncogene* 1995, **10:**135-147.
14. Hiratsuka S, Minowa O, Kuno J, Noda T, Shibuya M: **Flt-1 lacking the tyrosine kinase domain is sufficient for normal development and angiogenesis in mice.** *Proc Natl Acad Sci U S A* 1998, **95:**9349-9354.
15. **VEGF pathway: Target several targeted therapies against cancer Avastin Sutent Nexavar** [http://www.youtube.com/watch?v=SZpUi1EFWZU]
16. Cross MJ, Claesson-Welsh L: **FGF and VEGF function in angiogenesis: signalling pathways, biological responses and therapeutic inhibition.** *Trends Pharmacol Sci* 2001, **22:**201-207.
17. Tassi E, Wellstein A: **The angiogenic switch molecule, secreted FGF-binding protein, an indicator of early stages of pancreatic and colorectal adenocarcinoma.** *Semin Oncol* 2006, **33:**S50-56.

18. Duchesne L, Tissot B, Rudd TR, Dell A, Fernig DG: **N-glycosylation of fibroblast growth factor receptor 1 regulates ligand and heparan sulfate co-receptor binding.** *J Biol Chem* 2006, **281:**27178-27189.
19. Iozzo RV, San Antonio JD: **Heparan sulfate proteoglycans: heavy hitters in the angiogenesis arena.** *J Clin Invest* 2001, **108:**349-355.
20. Hanahan D, Folkman J: **Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis.** *Cell* 1996, **86:**353-364.
21. Folkman J: **Tumor angiogenesis: therapeutic implications.** *N Engl J Med* 1971, **285:**1182-1186.
22. Bergers G, Benjamin LE: **Tumorigenesis and the angiogenic switch.** *Nat Rev Cancer* 2003, **3:**401-410.
23. Bridges EM, Harris AL: **The angiogenic process as a therapeutic target in cancer.** *Biochem Pharmacol* 2011, **81:**1183-1191.
24. Dings RP, Loren M, Heun H, McNiel E, Griffioen AW, Mayo KH, Griffin RJ: **Scheduling of radiation with angiogenesis inhibitors anginex and Avastin improves therapeutic outcome via vessel normalization.** *Clin Cancer Res* 2007, **13:**3395-3402.
25. Heath VL, Bicknell R: **Anticancer strategies involving the vasculature.** *Nat Rev Clin Oncol* 2009, **6:**395-404.
26. Cundari S, Cavaletti G: **Thalidomide chemotherapy-induced peripheral neuropathy: actual status and new perspectives with thalidomide analogues derivatives.** *Mini Rev Med Chem* 2009, **9:**760-768.
27. Neal J, Wakelee H: **AMG-386, a selective angiopoietin-1/-2-neutralizing peptibody for the potential treatment of cancer.** *Curr Opin Mol Ther* 2010, **12:**487-495.
28. Bikfalvi A, Moenner M, Javerzat S, North S, Hagedorn M: **Inhibition of angiogenesis and the angiogenesis/invasion shift.** *Biochem Soc Trans* 2011, **39:**1560-1564.
29. Saidi A, Hagedorn M, Allain N, Verpelli C, Sala C, Bello L, Bikfalvi A, Javerzat S: **Combined targeting of interleukin-6 and vascular endothelial growth factor potently inhibits glioma growth and invasiveness.** *Int J Cancer* 2009, **125:**1054-1064.
30. Eberhard A, Kahlert S, Goede V, Hemmerlein B, Plate KH, Augustin HG: **Heterogeneity of angiogenesis and blood vessel maturation in human tumors: implications for antiangiogenic tumor therapies.** *Cancer Res* 2000, **60:**1388-1393.
31. Konerding MA, Fait E, Gaumann A: **3D microvascular architecture of pre-cancerous lesions and invasive carcinomas of the colon.** *Br J Cancer* 2001, **84:**1354-1362.
32. Loi M, Di Paolo D, Becherini P, Zorzoli A, Perri P, Carosio R, Cilli M, Ribatti D, Brignole C, Pagnan G, et al: **The use of the orthotopic model to validate antivascular therapies for cancer.** *Int J Dev Biol* 2011, **55:**547-555.
33. Dvorak HF, Nagy JA, Dvorak JT, Dvorak AM: **Identification and characterization of the blood vessels of solid tumors that are leaky to circulating macromolecules.** *Am J Pathol* 1988, **133:**95-109.

34. Hashizume H, Baluk P, Morikawa S, McLean JW, Thurston G, Roberge S, Jain RK, McDonald DM: **Openings between defective endothelial cells explain tumor vessel leakiness.** *Am J Pathol* 2000, **156:**1363-1380.
35. Vaupel P, Hockel M: **Blood supply, oxygenation status and metabolic micromilieu of breast cancers: characterization and therapeutic relevance.** *Int J Oncol* 2000, **17:**869-879.
36. Brown NS, Bicknell R: **Hypoxia and oxidative stress in breast cancer. Oxidative stress: its effects on the growth, metastatic potential and response to therapy of breast cancer.** *Breast Cancer Res* 2001, **3:**323-327.
37. Denekamp J, Hobson B: **Endothelial-cell proliferation in experimental tumours.** *Br J Cancer* 1982, **46:**711-720.
38. WOGLOM W: **A critique of tumour resistance.** *J Cancer Res* 1923, **7:**283-311.
39. Denekamp J: **Vascular attack as a therapeutic strategy for cancer.** *Cancer Metastasis Rev* 1990, **9:**267-282.
40. Burrows FJ, Thorpe PE: **Eradication of large solid tumors in mice with an immunotoxin directed against tumor vasculature.** *Proc Natl Acad Sci U S A* 1993, **90:**8996-9000.
41. Jennewein M, Lewis MA, Zhao D, Tsyganov E, Slavine N, He J, Watkins L, Kodibagkar VD, O'Kelly S, Kulkarni P, et al: **Vascular imaging of solid tumors in rats with a radioactive arsenic-labeled antibody that binds exposed phosphatidylserine.** *Clin Cancer Res* 2008, **14:**1377-1385.
42. Hill S, Williams KB, Denekamp J: **Vascular collapse after flavone acetic acid: a possible mechanism of its anti-tumour action.** *Eur J Cancer Clin Oncol* 1989, **25:**1419-1424.
43. Jeon SB, Kim G, Kim JI, Seok YM, Kim SH, Suk K, Shin HM, Lee YH, Kim IK: **Flavone inhibits vascular contraction by decreasing phosphorylation of the myosin phosphatase target subunit.** *Clin Exp Pharmacol Physiol* 2007, **34:**1116-1120.
44. Pastorino F, Marimpietri D, Brignole C, Di Paolo D, Pagnan G, Daga A, Piccardi F, Cilli M, Allen TM, Ponzoni M: **Ligand-targeted liposomal therapies of neuroblastoma.** *Curr Med Chem* 2007, **14:**3070-3078.
45. Nallamothu R, Wood GC, Pattillo CB, Scott RC, Kiani MF, Moore BM, Thoma LA: **A tumor vasculature targeted liposome delivery system for combretastatin A4: design, characterization, and in vitro evaluation.** *AAPS PharmSciTech* 2006, **7:**E32.
46. Arap W, Haedicke W, Bernasconi M, Kain R, Rajotte D, Krajewski S, Ellerby HM, Bredesen DE, Pasqualini R, Ruoslahti E: **Targeting the prostate for destruction through a vascular address.** *Proc Natl Acad Sci U S A* 2002, **99:**1527-1531.
47. Jacobson BS, Schnitzer JE, McCaffery M, Palade GE: **Isolation and partial characterization of the luminal plasmalemma of microvascular endothelium from rat lungs.** *Eur J Cell Biol* 1992, **58:**296-306.
48. **Patent; Method of recovering endothelial membrane from tissue and applications thereof** [http://www.patentgenius.com/patent/5610008.html]

49. Stan RV, Ghitescu L, Jacobson BS, Palade GE: **Isolation, cloning, and localization of rat PV-1, a novel endothelial caveolar protein.** *J Cell Biol* 1999, **145:**1189-1198.
50. Stan RV, Roberts WG, Predescu D, Ihida K, Saucan L, Ghitescu L, Palade GE: **Immunoisolation and partial characterization of endothelial plasmalemmal vesicles (caveolae).** *Mol Biol Cell* 1997, **8:**595-605.
51. Keuschnigg J, Henttinen T, Auvinen K, Karikoski M, Salmi M, Jalkanen S: **The prototype endothelial marker PAL-E is a leukocyte trafficking molecule.** *Blood* 2009, **114:**478-484.
52. Oh P, Li Y, Yu J, Durr E, Krasinska KM, Carver LA, Testa JE, Schnitzer JE: **Subtractive proteomic mapping of the endothelial surface in lung and solid tumours for tissue-specific therapy.** *Nature* 2004, **429:**629-635.
53. Rybak JN, Ettorre A, Kaissling B, Giavazzi R, Neri D, Elia G: **In vivo protein biotinylation for identification of organ-specific antigens accessible from the vasculature.** *Nat Methods* 2005, **2:**291-298.
54. Rybak JN, Scheurer SB, Neri D, Elia G: **Purification of biotinylated proteins on streptavidin resin: a protocol for quantitative elution.** *Proteomics* 2004, **4:**2296-2299.
55. Roesli C, Neri D, Rybak JN: **In vivo protein biotinylation and sample preparation for the proteomic identification of organ- and disease-specific antigens accessible from the vasculature.** *Nat Protoc* 2006, **1:**192-199.
56. Conrotto P, Roesli C, Rybak J, Kischel P, Waltregny D, Neri D, Castronovo V: **Identification of new accessible tumor antigens in human colon cancer by ex vivo protein biotinylation and comparative mass spectrometry analysis.** *Int J Cancer* 2008, **123:**2856-2864.
57. Hull MA, Hewett PW, Brough JL, Hawkey CJ: **Isolation and culture of human gastric endothelial cells.** *Gastroenterology* 1996, **111:**1230-1240.
58. Dong QG, Bernasconi S, Lostaglio S, De Calmanovici RW, Martin-Padura I, Breviario F, Garlanda C, Ramponi S, Mantovani A, Vecchi A: **A general strategy for isolation of endothelial cells from murine tissues. Characterization of two endothelial cell lines from the murine lung and subcutaneous sponge implants.** *Arterioscler Thromb Vasc Biol* 1997, **17:**1599-1604.
59. Adams MD, Dubnick M, Kerlavage AR, Moreno R, Kelley JM, Utterback TR, Nagle JW, Fields C, Venter JC: **Sequence identification of 2,375 human brain genes.** *Nature* 1992, **355:**632-634.
60. Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, et al.: **Complementary DNA sequencing: expressed sequence tags and human genome project.** *Science* 1991, **252:**1651-1656.
61. Bortoluzzi S, Danieli GA: **Towards an in silico analysis of transcription patterns.** *Trends Genet* 1999, **15:**118-119.
62. Herbert JM, Stekel D, Sanderson S, Heath VL, Bicknell R: **A novel method of differential gene expression analysis using multiple cDNA libraries applied to the identification of tumour endothelial genes.** *BMC Genomics* 2008, **9:**153.

63. Huminiecki L, Bicknell R: **In silico cloning of novel endothelial-specific genes.** *Genome Res* 2000, **10:**1796-1806.
64. Lash AE, Tolstoshev CM, Wagner L, Schuler GD, Strausberg RL, Riggins GJ, Altschul SF: **SAGEmap: a public gene expression resource.** *Genome Res* 2000, **10:**1051-1060.
65. Huminiecki L, Gorn M, Suchting S, Poulsom R, Bicknell R: **Magic roundabout is a new member of the roundabout receptor family that is endothelial specific and expressed at sites of active angiogenesis.** *Genomics* 2002, **79:**547-552.
66. Sullivan DC, Huminiecki L, Moore JW, Boyle JJ, Poulsom R, Creamer D, Barker J, Bicknell R: **EndoPDI, a novel protein-disulfide isomerase-like protein that is preferentially expressed in endothelial cells acts as a stress survival factor.** *J Biol Chem* 2003, **278:**47079-47088.
67. Kaur S, Leszczynska K, Abraham S, Scarcia M, Hiltbrunner S, Marshall CJ, Mavria G, Bicknell R, Heath VL: **RhoJ/TCL regulates endothelial motility and tube formation and modulates actomyosin contractility and focal adhesion numbers.** *Arterioscler Thromb Vasc Biol* 2011, **31:**657-664.
68. van Beijnum JR, Dings RP, van der Linden E, Zwaans BM, Ramaekers FC, Mayo KH, Griffioen AW: **Gene expression of tumor angiogenesis dissected: specific targeting of colon cancer angiogenic vasculature.** *Blood* 2006, **108:**2339-2348.
69. St Croix B, Rago C, Velculescu V, Traverso G, Romans KE, Montgomery E, Lal A, Riggins GJ, Lengauer C, Vogelstein B, Kinzler KW: **Genes expressed in human tumor endothelium.** *Science* 2000, **289:**1197-1202.
70. Burgio VL, Zupo S, Roncella S, Zocchi M, Ruco LP, Baroni CD: **Characterization of EN4 monoclonal antibody: a reagent with CD31 specificity.** *Clin Exp Immunol* 1994, **96:**170-176.
71. Seaman S, Stevens J, Yang MY, Logsdon D, Graff-Cherry C, St Croix B: **Genes that distinguish physiological and pathological angiogenesis.** *Cancer Cell* 2007, **11:**539-554.
72. Gerhardt H, Golding M, Fruttiger M, Ruhrberg C, Lundkvist A, Abramsson A, Jeltsch M, Mitchell C, Alitalo K, Shima D, Betsholtz C: **VEGF guides angiogenic sprouting utilizing endothelial tip cell filopodia.** *J Cell Biol* 2003, **161:**1163-1177.
73. Madden SL, Cook BP, Nacht M, Weber WD, Callahan MR, Jiang Y, Dufault MR, Zhang X, Zhang W, Walter-Yohrling J, et al: **Vascular gene expression in nonneoplastic and malignant brain.** *Am J Pathol* 2004, **165:**601-608.
74. Holthofer H, Virtanen I, Kariniemi AL, Hormia M, Linder E, Miettinen A: **Ulex europaeus I lectin as a marker for vascular endothelium in human tissues.** *Lab Invest* 1982, **47:**60-66.
75. Saha S, Sparks AB, Rago C, Akmaev V, Wang CJ, Vogelstein B, Kinzler KW, Velculescu VE: **Using the transcriptome to annotate the genome.** *Nat Biotechnol* 2002, **20:**508-512.
76. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW: **Serial analysis of gene expression.** *Science* 1995, **270:**484-487.

77. Bhati R, Patterson C, Livasy CA, Fan C, Ketelsen D, Hu Z, Reynolds E, Tanner C, Moore DT, Gabrielli F, et al: **Molecular characterization of human breast tumor vascular cells.** *Am J Pathol* 2008, **172:**1381-1390.
78. Pasqualini R, Moeller BJ, Arap W: **Leveraging molecular heterogeneity of the vascular endothelium for targeted drug delivery and imaging.** *Semin Thromb Hemost* 2010, **36:**343-351.
79. Papayannopoulou T, Priestley GV, Nakamoto B, Zafiropoulos V, Scott LM: **Molecular pathways in bone marrow homing: dominant role of alpha(4)beta(1) over beta(2)-integrins and selectins.** *Blood* 2001, **98:**2403-2411.
80. Smith GP: **Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface.** *Science* 1985, **228:**1315-1317.
81. Arap W, Kolonin MG, Trepel M, Lahdenranta J, Cardo-Vila M, Giordano RJ, Mintz PJ, Ardelt PU, Yao VJ, Vidal CI, et al: **Steps toward mapping the human vasculature by phage display.** *Nat Med* 2002, **8:**121-127.
82. Mahboubi K, Biedermann BC, Carroll JM, Pober JS: **IL-11 activates human endothelial cells to resist immune-mediated injury.** *J Immunol* 2000, **164:**3837-3846.
83. Campbell CL, Jiang Z, Savarese DM, Savarese TM: **Increased expression of the interleukin-11 receptor and evidence of STAT3 activation in prostate carcinoma.** *Am J Pathol* 2001, **158:**25-32.
84. Levene P: **The structure of yeast nucleic acid.** *Biol Chem* 1919, **40:**415–424.
85. Lorenz MG, Wackernagel W: **Bacterial gene transfer by natural genetic transformation in the environment.** *Microbiol Rev* 1994, **58:**563-602.
86. Avery OT, Macleod CM, McCarty M: **STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES : INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III.** *J Exp Med* 1944, **79:**137-158.
87. Astbury W: **Nucleic acid.** *Symp SOC Exp Biol* 1947, **1:**66.
88. Chargaff E, Vischer E, et al.: **The composition of the desoxypentose nucleic acids of thymus and spleen.** *J Biol Chem* 1949, **177:**405-416.
89. Chargaff E: **Structure and function of nucleic acids as cell constituents.** *Fed Proc* 1951, **10:**654-659.
90. Hershey AD, Chase M: **Independent functions of viral protein and nucleic acid in growth of bacteriophage.** *J Gen Physiol* 1952, **36:**39-56.
91. Franklin RE, Gosling RG: **Molecular configuration in sodium thymonucleate.** *Nature* 1953, **171:**740-741.
92. Watson JD, Crick FH: **Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid.** *Nature* 1953, **171:**737-738.
93. Wilkins MH, Stokes AR, Wilson HR: **Molecular structure of deoxypentose nucleic acids.** *Nature* 1953, **171:**738-740.
94. **Biography 19: Rosalind Elsie Franklin (1920-1958)** [http://www.dnalc.org/view/16439-Biography-19-Rosalind-Elsie-Franklin-1920-1958-.html]

95. **The history of DNA** [http://en.wikipedia.org/wiki/DNA#History_of_DNA_research]
96. Maxam AM, Gilbert W: **A new method for sequencing DNA.** *Proc Natl Acad Sci U S A* 1977, **74:**560-564.
97. Sanger F, Nicklen S, Coulson AR: **DNA sequencing with chain-terminating inhibitors.** *Proc Natl Acad Sci U S A* 1977, **74:**5463-5467.
98. Turcatti G, Romieu A, Fedurco M, Tairi AP: **A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis.** *Nucleic Acids Res* 2008, **36:**e25.
99. Ronaghi M, Uhlen M, Nyren P: **A sequencing method based on real-time pyrophosphate.** *Science* 1998, **281:**363, 365.
100. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM: **Accurate multiplex polony sequencing of an evolved bacterial genome.** *Science* 2005, **309:**1728-1732.
101. **RNA-seq protocol for sequencing preparation** [http://www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/Sequencing/Next-Generation-Sequencing/Sample-Prep-Library-Generation-for-Next-Generation-Sequencing/next-generation-sequencing-reagents-by-workflow/Library-Construction/rna-library/solid-total-rna-seq-kit.html]
102. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biol* 2009, **10:**R25.
103. Trapnell C, Pachter L, Salzberg SL: **TopHat: discovering splice junctions with RNA-Seq.** *Bioinformatics* 2009, **25:**1105-1111.
104. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L: **Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation.** *Nat Biotechnol* 2010, **28:**511-515.
105. Koenig T, Menze BH, Kirchner M, Monigatti F, Parker KC, Patterson T, Steen JJ, Hamprecht FA, Steen H: **Robust prediction of the MASCOT score for an improved quality assessment in mass spectrometric proteomics.** *J Proteome Res* 2008, **7:**3708-3717.
106. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS: **Probability-based protein identification by searching sequence databases using mass spectrometry data.** *Electrophoresis* 1999, **20:**3551-3567.
107. Herbert JM, Buffa FM, Vorschmitt H, Egginton S, Bicknell R: **A new procedure for determining the genetic basis of a physiological process in a non-model species, illustrated by cold induced angiogenesis in the carp.** *BMC Genomics* 2009, **10:**490.
108. Smyth GK: **Linear models and empirical bayes methods for assessing differential expression in microarray experiments.** *Stat Appl Genet Mol Biol* 2004, **3:**Article3.
109. **Reference sequence project sequence downloads** [ftp://ftp.ncbi.nih.gov/refseq]

110. Strausberg RL, Buetow KH, Emmert-Buck MR, Klausner RD: **The cancer genome anatomy project: building an annotated gene index.** *Trends Genet* 2000, **16:**103-106.
111. Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schaffer AA: **Database indexing for production MegaBLAST searches.** *Bioinformatics* 2008, **24:**1757-1764.
112. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403-410.
113. **Center for Biological Sequence Analysis Prediction Servers** [http://www.cbs.dtu.dk/index.shtml]
114. **The Database for Annotation, Visualization and Integrated Discovery** [http://david.abcc.ncifcrf.gov/home.jsp]
115. Herbert JM, Stekel DJ, Mura M, Sychev M, Bicknell R: **Bioinformatic methods for finding differentially expressed genes in cDNA libraries, applied to the identification of tumour vascular targets.** *Methods Mol Biol* 2011, **729:**99-119.
116. **Website; Bioinformatic methods for finding differentially expressed genes in cDNA libraries, applied to the identification of tumour vascular targets** [http://sara.molbiol.ox.ac.uk/userweb/jherbert/tissue_diffex/tissue_index.html]
117. Tsunoda T, Inada H, Kalembeyi I, Imanaka-Yoshida K, Sakakibara M, Okada R, Katsuta K, Sakakura T, Majima Y, Yoshida T: **Involvement of large tenascin-C splice variants in breast cancer progression.** *Am J Pathol* 2003, **162:**1857-1867.
118. Ventura E, Sassi F, Parodi A, Balza E, Borsi L, Castellani P, Carnemolla B, Zardi L: **Alternative splicing of the angiogenesis associated extra-domain B of fibronectin regulates the accessibility of the B-C loop of the type III repeat 8.** *PLoS One* 2010, **5:**e9145.
119. Harris AL: **Hypoxia--a key regulatory factor in tumour growth.** *Nat Rev Cancer* 2002, **2:**38-47.
120. Burbridge MF, West DC, Atassi G, Tucker GC: **The effect of extracellular pH on angiogenesis in vitro.** *Angiogenesis* 1999, **3:**281-288.
121. MacFadyen J, Savage K, Wienke D, Isacke CM: **Endosialin is expressed on stromal fibroblasts and CNS pericytes in mouse embryos and is downregulated during development.** *Gene Expr Patterns* 2007, **7:**363-369.
122. MacFadyen JR, Haworth O, Roberston D, Hardie D, Webster MT, Morris HR, Panico M, Sutton-Smith M, Dell A, van der Geer P, et al: **Endosialin (TEM1, CD248) is a marker of stromal fibroblasts and is not selectively expressed on tumour endothelium.** *FEBS Lett* 2005, **579:**2569-2575.
123. Patel NS, Li JL, Generali D, Poulsom R, Cranston DW, Harris AL: **Up-regulation of delta-like 4 ligand in human tumor vasculature and the role of basal expression in endothelial cell function.** *Cancer Res* 2005, **65:**8690-8697.
124. Mailhos C, Modlich U, Lewis J, Harris A, Bicknell R, Ish-Horowicz D: **Delta4, an endothelial specific notch ligand expressed at sites of physiological and tumor angiogenesis.** *Differentiation* 2001, **69:**135-144.
125. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ: **Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing.** *Nat Genet* 2008, **40:**1413-1415.

126. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq.** *Nat Methods* 2008, **5:**621-628.
127. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L: **Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.** *Nat Protoc* 2012, **7:**562-578.
128. **University of California Santa Cruz table browser** [http://genome.ucsc.edu/cgi-bin/hgTables]
129. DAVID: **The Database for Annotation, Visualization and Integrated Discovery.** 2009.
130. **Cufflinks User Online Manual** [http://cufflinks.cbcb.umd.edu/manual.html]
131. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R: **InterProScan: protein domains identifier.** *Nucleic Acids Res* 2005, **33:**W116-120.
132. Neri D, Bicknell R: **Tumour vascular targeting.** *Nat Rev Cancer* 2005, **5:**436-446.
133. Hancox RA, Allen MD, Holliday DL, Edwards DR, Pennington CJ, Guttery DS, Shaw JA, Walker RA, Pringle JH, Jones JL: **Tumour-associated tenascin-C isoforms promote breast cancer cell invasion and growth by matrix metalloproteinase-dependent and independent mechanisms.** *Breast Cancer Res* 2009, **11:**R24.
134. Sauer S, Erba PA, Petrini M, Menrad A, Giovannoni L, Grana C, Hirsch B, Zardi L, Paganelli G, Mariani G, et al: **Expression of the oncofetal ED-B-containing fibronectin isoform in hematologic tumors enables ED-B-targeted 131I-L19SIP radioimmunotherapy in Hodgkin lymphoma patients.** *Blood* 2009, **113:**2265-2274.
135. Kremmidiotis G, Baker E, Crawford J, Eyre HJ, Nahmias J, Callen DF: **Localization of human cadherin genes to chromosome regions exhibiting cancer-related loss of heterozygosity.** *Genomics* 1998, **49:**467-471.
136. **CTLA-4 (Cytotoxic T-Lymphocyte Antigen 4)** [http://en.wikipedia.org/wiki/CTLA-4]
137. **preprocessCore: A collection of pre-processing functions** [http://www.bioconductor.org/packages/release/bioc/html/preprocessCore.html]
138. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, et al: **Bioconductor: open software development for computational biology and bioinformatics.** *Genome Biol* 2004, **5:**R80.
139. Ringner M: **What is principal component analysis?** *Nat Biotechnol* 2008, **26:**303-304.
140. Clarke R, Ressom HW, Wang A, Xuan J, Liu MC, Gehan EA, Wang Y: **The properties of high-dimensional data spaces: implications for exploring gene and protein expression data.** *Nat Rev Cancer* 2008, **8:**37-49.
141. **E constant explained (used in exponential function)** [http://betterexplained.com/articles/an-intuitive-guide-to-exponential-functions-e/]
142. **Bioconductor limma for 2-color data** [http://bcf.arl.arizona.edu/documents]

143. Benjamini y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J R Statist Soc B* 1995, **57:**289-300.
144. Bon E, Steegers R, Steegers EA, Ursem N, Charif H, Burgers PC, Luider TM, Dekker LJ: **Proteomic analyses of the developing chicken cardiovascular system.** *J Proteome Res* 2010, **9:**268-274.
145. Parada C, Gato A, Aparicio M, Bueno D: **Proteome analysis of chick embryonic cerebrospinal fluid.** *Proteomics* 2006, **6:**312-320.
146. Christofk HR, Vander Heiden MG, Harris MH, Ramanathan A, Gerszten RE, Wei R, Fleming MD, Schreiber SL, Cantley LC: **The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth.** *Nature* 2008, **452:**230-233.
147. **The Warburg effect in cancer cells** [http://en.wikipedia.org/wiki/Warburg_effect]
148. Schafer M, Werner S: **Cancer as an overhealing wound: an old hypothesis revisited.** *Nat Rev Mol Cell Biol* 2008, **9:**628-638.
149. Dvorak HF: **Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing.** *N Engl J Med* 1986, **315:**1650-1659.
150. Riss J, Khanna C, Koo S, Chandramouli GV, Yang HH, Hu Y, Kleiner DE, Rosenwald A, Schaefer CF, Ben-Sasson SA, et al: **Cancers as wounds that do not heal: differences and similarities between renal regeneration/repair and renal cell carcinoma.** *Cancer Res* 2006, **66:**7216-7224.
151. Hufton SE, Moerkerk PT, Brandwijk R, de Bruine AP, Arends JW, Hoogenboom HR: **A profile of differentially expressed genes in primary colorectal cancer using suppression subtractive hybridization.** *FEBS Lett* 1999, **463:**77-82.
152. Van Beijnum JR, Moerkerk PT, Gerbers AJ, De Bruine AP, Arends JW, Hoogenboom HR, Hufton SE: **Target validation for genomics using peptide-specific phage antibodies: a study of five gene products overexpressed in colorectal cancer.** *Int J Cancer* 2002, **101:**118-127.
153. Reddy PS, Umesh S, Thota B, Tandon A, Pandey P, Hegde AS, Balasubramaniam A, Chandramouli BA, Santosh V, Rao MR, et al: **PBEF1/NAmPRTase/Visfatin: a potential malignant astrocytoma/glioblastoma serum marker with prognostic value.** *Cancer Biol Ther* 2008, **7:**663-668.
154. Kim SR, Bae SK, Choi KS, Park SY, Jun HO, Lee JY, Jang HO, Yun I, Yoon KH, Kim YJ, et al: **Visfatin promotes angiogenesis by activation of extracellular signal-regulated kinase 1/2.** *Biochem Biophys Res Commun* 2007, **357:**150-156.
155. Bae YH, Bae MK, Kim SR, Lee JH, Wee HJ, Bae SK: **Upregulation of fibroblast growth factor-2 by visfatin that promotes endothelial angiogenesis.** *Biochem Biophys Res Commun* 2009, **379:**206-211.
156. Bae YH, Park HJ, Kim SR, Kim JY, Kang Y, Kim JA, Wee HJ, Kageyama R, Jung JS, Bae MK, Bae SK: **Notch1 mediates visfatin-induced FGF-2 up-regulation and endothelial angiogenesis.** *Cardiovasc Res* 2011, **89:**436-445.
157. Kim JY, Bae YH, Bae MK, Kim SR, Park HJ, Wee HJ, Bae SK: **Visfatin through STAT3 activation enhances IL-6 expression that promotes endothelial angiogenesis.** *Biochim Biophys Acta* 2009, **1793:**1759-1767.

158. Kim SR, Bae YH, Bae SK, Choi KS, Yoon KH, Koo TH, Jang HO, Yun I, Kim KW, Kwon YG, et al: **Visfatin enhances ICAM-1 and VCAM-1 expression through ROS-dependent NF-kappaB activation in endothelial cells.** *Biochim Biophys Acta* 2008, **1783:**886-895.
159. Adya R, Tan BK, Punn A, Chen J, Randeva HS: **Visfatin induces human endothelial VEGF and MMP-2/9 production via MAPK and PI3K/Akt signalling pathways: novel insights into visfatin-induced angiogenesis.** *Cardiovasc Res* 2008, **78:**356-365.
160. Arumugam T, Simeone DM, Schmidt AM, Logsdon CD: **S100P stimulates cell proliferation and survival via receptor for activated glycation end products (RAGE).** *J Biol Chem* 2004, **279:**5059-5065.
161. Kimura T, Budka H, Soler-Federsppiel S: **An immunocytochemical comparison of the glia-associated proteins glial fibrillary acidic protein (GFAP) and S-100 protein (S100P) in human brain tumors.** *Clin Neuropathol* 1986, **5:**21-27.
162. **Atlas of Genetics and Cytogenetics in Oncology and Haematology** [http://atlasgeneticsoncology.org]
163. Beer DG, Kardia SL, Huang CC, Giordano TJ, Levin AM, Misek DE, Lin L, Chen G, Gharib TG, Thomas DG, et al: **Gene-expression profiles predict survival of patients with lung adenocarcinoma.** *Nat Med* 2002, **8:**816-824.
164. Crnogorac-Jurcevic T, Missiaglia E, Blaveri E, Gangeswaran R, Jones M, Terris B, Costello E, Neoptolemos JP, Lemoine NR: **Molecular alterations in pancreatic carcinoma: expression profiling shows that dysregulated expression of S100 genes is highly prevalent.** *J Pathol* 2003, **201:**63-74.
165. Fuentes MK, Nigavekar SS, Arumugam T, Logsdon CD, Schmidt AM, Park JC, Huang EH: **RAGE activation by S100P in colon cancer stimulates growth, migration, and cell signaling pathways.** *Dis Colon Rectum* 2007, **50:**1230-1240.
166. Guerreiro Da Silva ID, Hu YF, Russo IH, Ao X, Salicioni AM, Yang X, Russo J: **S100P calcium-binding protein overexpression is associated with immortalization of human breast epithelial cells in vitro and early stages of breast cancer development in vivo.** *Int J Oncol* 2000, **16:**231-240.
167. Mousses S, Bubendorf L, Wagner U, Hostetter G, Kononen J, Cornelison R, Goldberger N, Elkahloun AG, Willi N, Koivisto P, et al: **Clinical validation of candidate genes associated with prostate cancer progression in the CWR22 model system using tissue microarrays.** *Cancer Res* 2002, **62:**1256-1260.
168. Koltzscher M, Neumann C, Konig S, Gerke V: **Ca2+-dependent binding and activation of dormant ezrin by dimeric S100P.** *Mol Biol Cell* 2003, **14:**2372-2384.
169. Ebihara T, Endo R, Kikuta H, Ishiguro N, Ma X, Shimazu M, Otoguro T, Kobayashi K: **Differential gene expression of S100 protein family in leukocytes from patients with Kawasaki disease.** *Eur J Pediatr* 2005, **164:**427-431.
170. Tothova V, Isola J, Parkkila S, Kopacek J, Pastorek J, Pastorekova S, Gibadulinova A: **Glucocorticoid receptor-mediated transcriptional activation of S100P gene coding for cancer-related calcium-binding protein.** *J Cell Biochem* 2011, **112:**3373-3384.

171. Kurji KH, Cui JZ, Lin T, Harriman D, Prasad SS, Kojic L, Matsubara JA: **Microarray analysis identifies changes in inflammatory gene expression in response to amyloid-beta stimulation of cultured human retinal pigment epithelial cells.** *Invest Ophthalmol Vis Sci* 2010, **51:**1151-1163.
172. Zand L, Ryu JK, McLarnon JG: **Induction of angiogenesis in the beta-amyloid peptide-injected rat hippocampus.** *Neuroreport* 2005, **16:**129-132.
173. Hagedorn M, Javerzat S, Gilges D, Meyre A, de Lafarge B, Eichmann A, Bikfalvi A: **Accessing key steps of human tumor progression in vivo by using an avian embryo model.** *Proc Natl Acad Sci U S A* 2005, **102:**1643-1648.
174. Tsai MH, Cook JA, Chandramouli GV, DeGraff W, Yan H, Zhao S, Coleman CN, Mitchell JB, Chuang EY: **Gene expression profiling of breast, prostate, and glioma cells following single versus fractionated doses of radiation.** *Cancer Res* 2007, **67:**3845-3852.
175. Wan J, Chai H, Yu Z, Ge W, Kang N, Xia W, Che Y: **HIF-1alpha effects on angiogenic potential in human small cell lung carcinoma.** *J Exp Clin Cancer Res* 2011, **30:**77.
176. **Patent application; PRO-ANGIOGENIC GENES IN OVARIAN TUMOR ENDOTHELIAL CELL ISOLATES** [http://www.faqs.org/patents/app/20100286237#b]
177. Buckanovich RJ, Sasaroli D, O'Brien-Jenkins A, Botbyl J, Hammond R, Katsaros D, Sandaltzopoulos R, Liotta LA, Gimotty PA, Coukos G: **Tumor vascular proteins as biomarkers in ovarian cancer.** *J Clin Oncol* 2007, **25:**852-861.
178. Sasaroli D, Gimotty PA, Pathak HB, Hammond R, Kougioumtzidou E, Katsaros D, Buckanovich R, Devarajan K, Sandaltzopoulos R, Godwin AK, et al: **Novel surface targets and serum biomarkers from the ovarian cancer vasculature.** *Cancer Biol Ther* 2011, **12:**169-180.
179. Petroff MG, Kharatyan E, Torry DS, Holets L: **The immunomodulatory proteins B7-DC, B7-H2, and B7-H3 are differentially expressed across gestation in the human placenta.** *Am J Pathol* 2005, **167:**465-473.
180. Brunner A, Hinterholzer S, Riss P, Heinze G, Brustmann H: **Immunoexpression of B7-H3 in endometrial cancer: relation to tumor T-cell infiltration and prognosis.** *Gynecol Oncol* 2012, **124:**105-111.
181. Nikitenko LL, Blucher N, Fox SB, Bicknell R, Smith DM, Rees MC: **Adrenomedullin and CGRP interact with endogenous calcitonin-receptor-like receptor in endothelial cells and induce its desensitisation by different mechanisms.** *J Cell Sci* 2006, **119:**910-922.
182. Fonseca C, Abraham D, Ponticos M: **Neuronal regulators and vascular dysfunction in Raynaud's phenomenon and systemic sclerosis.** *Curr Vasc Pharmacol* 2009, **7:**34-39.
183. Cohen BE, Gill G, Cullen PR, Morris PJ: **Reversal of postoperative immunosuppression in man by vitamin A.** *Surg Gynecol Obstet* 1979, **149:**658-662.
184. Freiman M, Seifter E, Connerton C, Levenson SM: **Vitamin A deficiency and surgical stress.** *Surg Forum* 1970, **21:**81-82.
185. Shapiro SS, Mott DJ: **Modulation of glycosaminoglycan biosynthesis by retinoids.** *Ann N Y Acad Sci* 1981, **359:**306-321.

186. Balmer JE, Blomhoff R: **Gene expression regulation by retinoic acid.** *J Lipid Res* 2002, **43:**1773-1808.
187. Chen L, Hughes RA, Baines AJ, Conboy J, Mohandas N, An X: **Protein 4.1R regulates cell adhesion, spreading, migration and motility of mouse keratinocytes by modulating surface expression of beta1 integrin.** *J Cell Sci* 2011, **124:**2478-2487.
188. O'Neill AK, Gallegos LL, Justilien V, Garcia EL, Leitges M, Fields AP, Hall RA, Newton AC: **Protein kinase Calpha promotes cell migration through a PDZ-dependent interaction with its novel substrate discs large homolog 1 (DLG1).** *J Biol Chem* 2011, **286:**43559-43568.
189. Roodink I, Raats J, van der Zwaag B, Verrijp K, Kusters B, van Bokhoven H, Linkels M, de Waal RM, Leenders WP: **Plexin D1 expression is induced on tumor vasculature and tumor cells: a novel target for diagnosis and therapy?** *Cancer Res* 2005, **65:**8317-8323.
190. Morozova O, Marra MA: **Applications of next-generation sequencing technologies in functional genomics.** *Genomics* 2008, **92:**255-264.
191. Zhang LQ, Cheranova D, Gibson M, Ding S, Heruth DP, Fang D, Ye SQ: **RNA-seq reveals novel transcriptome of genes and their isoforms in human pulmonary microvascular endothelial cells treated with thrombin.** *PLoS One* 2012, **7:**e31229.
192. Voellenkle C, Rooij J, Guffanti A, Brini E, Fasanaro P, Isaia E, Croft L, David M, Capogrossi MC, Moles A, et al: **Deep-sequencing of endothelial cells exposed to hypoxia reveals the complexity of known and novel microRNAs.** *RNA* 2012, **18:**472-484.
193. Guduric-Fuchs J, O'Connor A, Cullen A, Harwood L, Medina RJ, O'Neill CL, Stitt AW, Curtis TM, Simpson DA: **Deep sequencing reveals predominant expression of miR-21 amongst the small non-coding RNAs in retinal microvascular endothelial cells.** *J Cell Biochem* 2012, **113:**2098-2111.
194. Faucheux BA, Bonnet AM, Agid Y, Hirsch EC: **Blood vessels change in the mesencephalon of patients with Parkinson's disease.** *Lancet* 1999, **353:**981-982.
195. Lindgren HS, Ohlin KE, Cenci MA: **Differential involvement of D1 and D2 dopamine receptors in L-DOPA-induced angiogenic activity in a rat model of Parkinson's disease.** *Neuropsychopharmacology* 2009, **34:**2477-2488.
196. Lopez VM, Decatur CL, Stamer WD, Lynch RM, McKay BS: **L-DOPA is an endogenous ligand for OA1.** *PLoS Biol* 2008, **6:**e236.
197. Rock R, Heinrich AC, Schumacher N, Gessler M: **Fjx1: a notch-inducible secreted ligand with specific binding sites in developing mouse embryos and adult brain.** *Dev Dyn* 2005, **234:**602-612.
198. Hellstrom M, Phng LK, Hofmann JJ, Wallgard E, Coultas L, Lindblom P, Alva J, Nilsson AK, Karlsson L, Gaiano N, et al: **Dll4 signalling through Notch1 regulates formation of tip cells during angiogenesis.** *Nature* 2007, **445:**776-780.
199. Kimura T, Watanabe T, Sato K, Kon J, Tomura H, Tamama K, Kuwabara A, Kanda T, Kobayashi I, Ohta H, et al: **Sphingosine 1-phosphate stimulates**

proliferation and migration of human endothelial cells possibly through the lipid receptors, Edg-1 and Edg-3. *Biochem J* 2000, **348 Pt 1:**71-76.

200. Hla T, Lee MJ, Ancellin N, Thangada S, Liu CH, Kluk M, Chae SS, Wu MT: **Sphingosine-1-phosphate signaling via the EDG-1 family of G-protein-coupled receptors.** *Ann N Y Acad Sci* 2000, **905:**16-24.
201. Vinci MC, Bellik L, Filippi S, Ledda F, Parenti A: **Trophic effects induced by alpha1D-adrenoceptors on endothelial cells are potentiated by hypoxia.** *Am J Physiol Heart Circ Physiol* 2007, **293:**H2140-2147.
202. Xiang Y, Tan YR, Zhang JS, Qin XQ, Hu BB, Wang Y, Qu F, Liu HJ: **Wound repair and proliferation of bronchial epithelial cells regulated by CTNNAL1.** *J Cell Biochem* 2008, **103:**920-930.
203. Gupta GP, Nguyen DX, Chiang AC, Bos PD, Kim JY, Nadal C, Gomis RR, Manova-Todorova K, Massague J: **Mediators of vascular remodelling co-opted for sequential steps in lung metastasis.** *Nature* 2007, **446:**765-770.
204. Chen LY, Liu X, Wang SL, Qin CY: **Over-expression of the Endocan gene in endothelial cells from hepatocellular carcinoma is associated with angiogenesis and tumour invasion.** *J Int Med Res* 2010, **38:**498-510.
205. Hatfield KJ, Lassalle P, Leiva RA, Lindas R, Wendelboe O, Bruserud O: **Serum levels of endothelium-derived endocan are increased in patients with untreated acute myeloid leukemia.** *Hematology* 2011, **16:**351-356.
206. Huang GW, Tao YM, Ding X: **Endocan expression correlated with poor survival in human hepatocellular carcinoma.** *Dig Dis Sci* 2009, **54:**389-394.
207. Kang YH, Ji NY, Lee CI, Lee HG, Kim JW, Yeom YI, Kim DG, Yoon SK, Kim JW, Park PJ, Song EY: **ESM-1 silencing decreased cell survival, migration, and invasion and modulated cell cycle progression in hepatocellular carcinoma.** *Amino Acids* 2011, **40:**1003-1013.
208. Leroy X, Aubert S, Zini L, Franquet H, Kervoaze G, Villers A, Delehedde M, Copin MC, Lassalle P: **Vascular endocan (ESM-1) is markedly overexpressed in clear cell renal cell carcinoma.** *Histopathology* 2010, **56:**180-187.
209. Liu N, Zhang LH, Du H, Hu Y, Zhang GG, Wang XH, Li JY, Ji JF: **Overexpression of endothelial cell specific molecule-1 (ESM-1) in gastric cancer.** *Ann Surg Oncol* 2010, **17:**2628-2639.
210. Maurage CA, Adam E, Mineo JF, Sarrazin S, Debunne M, Siminski RM, Baroncini M, Lassalle P, Blond S, Delehedde M: **Endocan expression and localization in human glioblastomas.** *J Neuropathol Exp Neurol* 2009, **68:**633-641.
211. Rennel E, Mellberg S, Dimberg A, Petersson L, Botling J, Ameur A, Westholm JO, Komorowski J, Lassalle P, Cross MJ, Gerwins P: **Endocan is a VEGF-A and PI3K regulated gene with increased expression in human renal cancer.** *Exp Cell Res* 2007, **313:**1285-1294.
212. Shin JW, Huggenberger R, Detmar M: **Transcriptional profiling of VEGF-A and VEGF-C target genes in lymphatic endothelium reveals endothelial-specific molecule-1 as a novel mediator of lymphangiogenesis.** *Blood* 2008, **112:**2318-2326.

213. Xiang X, Zhao WB, Wang X: **[Expression of ESM-1 in hepatocellular carcinoma is associated with angiogenesis and tumor invasion].** *Zhonghua Gan Zang Bing Za Zhi* 2009, **17:**661-664.
214. Lin Z, Gao C, Ning Y, He X, Wu W, Chen YG: **The pseudoreceptor BMP and activin membrane-bound inhibitor positively modulates Wnt/beta-catenin signaling.** *J Biol Chem* 2008, **283:**33053-33058.
215. Guillot N, Kollins D, Gilbert V, Xavier S, Chen J, Gentle M, Reddy A, Bottinger E, Jiang R, Rastaldi MP, et al: **BAMBI Regulates Angiogenesis and Endothelial Homeostasis through Modulation of Alternative TGFbeta Signaling.** *PLoS One* 2012, **7:**e39406.
216. **How Many Genes are in the Human Genome?** [http://www.wisegeek.com/how-many-genes-are-in-the-human-genome.htm]
217. Adams MD, Kerlavage AR, Fleischmann RD, Fuldner RA, Bult CJ, Lee NH, Kirkness EF, Weinstock KG, Gocayne JD, White O, et al.: **Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence.** *Nature* 1995, **377:**3-174.
218. Ho M, Yang E, Matcuk G, Deng D, Sampas N, Tsalenko A, Tabibiazar R, Zhang Y, Chen M, Talbi S, et al: **Identification of endothelial cell genes by combined database mining and microarray analysis.** *Physiol Genomics* 2003, **13:**249-262.
219. Wallgard E, Larsson E, He L, Hellstrom M, Armulik A, Nisancioglu MH, Genove G, Lindahl P, Betsholtz C: **Identification of a core set of 58 gene transcripts with broad and specific expression in the microvasculature.** *Arterioscler Thromb Vasc Biol* 2008, **28:**1469-1476.
220. Karsan A, Pollet I, Yu LR, Chan KC, Conrads TP, Lucas DA, Andersen R, Veenstra T: **Quantitative proteomic analysis of sokotrasterol sulfate-stimulated primary human endothelial cells.** *Mol Cell Proteomics* 2005, **4:**191-204.
221. Valera VA, Li-Ning TE, Walter BA, Roberts DD, Linehan WM, Merino MJ: **Protein expression profiling in the spectrum of renal cell carcinomas.** *J Cancer* 2010, **1:**184-196.
222. Gale NW, Baluk P, Pan L, Kwan M, Holash J, DeChiara TM, McDonald DM, Yancopoulos GD: **Ephrin-B2 selectively marks arterial vessels and neovascularization sites in the adult, with expression in both endothelial and smooth-muscle cells.** *Dev Biol* 2001, **230:**151-160.
223. Vihanto MM, Plock J, Erni D, Frey BM, Frey FJ, Huynh-Do U: **Hypoxia up-regulates expression of Eph receptors and ephrins in mouse skin.** *Faseb J* 2005, **19:**1689-1691.
224. Bochenek ML, Dickinson S, Astin JW, Adams RH, Nobes CD: **Ephrin-B2 regulates endothelial cell morphology and motility independently of Eph-receptor binding.** *J Cell Sci* 2010, **123:**1235-1246.
225. Leali D, Inforzato A, Ronca R, Bianchi R, Belleri M, Coltrini D, Di Salle E, Sironi M, Norata GD, Bottazzi B, et al: **Long pentraxin 3/tumor necrosis factor-stimulated gene-6 interaction: a biological rheostat for fibroblast growth factor 2-mediated angiogenesis.** *Arterioscler Thromb Vasc Biol* 2012, **32:**696-703.

226. Gale NW, Thurston G, Hackett SF, Renard R, Wang Q, McClain J, Martin C, Witte C, Witte MH, Jackson D, et al: **Angiopoietin-2 is required for postnatal angiogenesis and lymphatic patterning, and only the latter role is rescued by Angiopoietin-1.** *Dev Cell* 2002, **3:**411-423.
227. Felcht M, Luck R, Schering A, Seidel P, Srivastava K, Hu J, Bartol A, Kienast Y, Vettel C, Loos EK, et al: **Angiopoietin-2 differentially regulates angiogenesis through TIE2 and integrin signaling.** *J Clin Invest* 2012, **122:**1991-2005.
228. Agarwal A, Tressel SL, Kaimal R, Balla M, Lam FH, Covic L, Kuliopulos A: **Identification of a metalloprotease-chemokine signaling system in the ovarian cancer microenvironment: implications for antiangiogenic therapy.** *Cancer Res* 2010, **70:**5880-5890.
229. Macartney-Coxson DP, Hood KA, Shi HJ, Ward T, Wiles A, O'Connor R, Hall DA, Lea RA, Royds JA, Stubbs RS, Rooker S: **Metastatic susceptibility locus, an 8p hot-spot for tumour progression disrupted in colorectal liver metastases: 13 candidate genes examined at the DNA, mRNA and protein level.** *BMC Cancer* 2008, **8:**187.
230. Peinado H, Del Carmen Iglesias-de la Cruz M, Olmeda D, Csiszar K, Fong KS, Vega S, Nieto MA, Cano A, Portillo F: **A molecular role for lysyl oxidase-like 2 enzyme in snail regulation and tumor progression.** *EMBO J* 2005, **24:**3446-3458.
231. Peinado H, Moreno-Bueno G, Hardisson D, Perez-Gomez E, Santos V, Mendiola M, de Diego JI, Nistal M, Quintanilla M, Portillo F, Cano A: **Lysyl oxidase-like 2 as a new poor prognosis marker of squamous cell carcinomas.** *Cancer Res* 2008, **68:**4541-4550.
232. Sebban S, Davidson B, Reich R: **Lysyl oxidase-like 4 is alternatively spliced in an anatomic site-specific manner in tumors involving the serosal cavities.** *Virchows Arch* 2009, **454:**71-79.
233. Bignon M, Pichol-Thievend C, Hardouin J, Malbouyres M, Brechot N, Nasciutti L, Barret A, Teillon J, Guillon E, Etienne E, et al: **Lysyl oxidase-like protein-2 regulates sprouting angiogenesis and type IV collagen assembly in the endothelial basement membrane.** *Blood* 2011, **118:**3979-3989.
234. Haines BP, Wheldon LM, Summerbell D, Heath JK, Rigby PW: **Regulated expression of FLRT genes implies a functional role in the regulation of FGF signalling during mouse development.** *Dev Biol* 2006, **297:**14-25.
235. Magnusson PU, Dimberg A, Mellberg S, Lukinius A, Claesson-Welsh L: **FGFR-1 regulates angiogenesis through cytokines interleukin-4 and pleiotrophin.** *Blood* 2007, **110:**4214-4222.
236. Li H, Ge C, Zhao F, Yan M, Hu C, Jia D, Tian H, Zhu M, Chen T, Jiang G, et al: **Hypoxia-inducible factor 1 alpha-activated angiopoietin-like protein 4 contributes to tumor metastasis via vascular cell adhesion molecule-1/integrin beta1 signaling in human hepatocellular carcinoma.** *Hepatology* 2011, **54:**910-919.
237. Zhu P, Goh YY, Chin HF, Kersten S, Tan NS: **Angiopoietin-like 4: a decade of research.** *Biosci Rep* 2012, **32:**211-219.

238. Soulet F, Kilarski WW, Antczak P, Herbert J, Bicknell R, Falciani F, Bikfalvi A: **Gene signatures in wound tissue as evidenced by molecular profiling in the chick embryo model.** *BMC Genomics* 2010, **11:**495.
239. Javerzat S, Franco M, Herbert J, Platonova N, Peille AL, Pantesco V, De Vos J, Assou S, Bicknell R, Bikfalvi A, Hagedorn M: **Correlating global gene regulation to angiogenesis in the developing chick extra-embryonic vascular system.** *PLoS One* 2009, **4:**e7856.
240. Mura M, Swain RK, Zhuang X, Vorschmitt H, Reynolds G, Durant S, Beesley JF, Herbert JM, Sheldon H, Andre M, et al: **Identification and angiogenic role of the novel tumor endothelial marker CLEC14A.** *Oncogene* 2012, **31:**293-305.
241. Dumartin L, Quemener C, Laklai H, Herbert J, Bicknell R, Bousquet C, Pyronnet S, Castronovo V, Schilling MK, Bikfalvi A, Hagedorn M: **Netrin-1 mediates early events in pancreatic adenocarcinoma progression, acting on tumor and endothelial cells.** *Gastroenterology* 2010, **138:**1595-1606, 1606 e1591-1598.
242. Bailey RL, Herbert JM, Khan K, Heath VL, Bicknell R, Tomlinson MG: **The emerging role of tetraspanin microdomains on endothelial cells.** *Biochem Soc Trans* 2011, **39:**1667-1673.
243. **Bioconductor Organisation** [http://www.bioconductor.org/]
244. UniProt C: **Reorganizing the protein space at the Universal Protein Resource (UniProt).** *Nucleic Acids Res* 2012, **40:**D71-75.
245. **Single channel analysis of Agilent microarray data with Limma** [http://matticklab.com/index.php?title=Single_channel_analysis_of_Agilent_microarray_data_with_Limma]
246. **NCBI public download repository for Genbank flat files.** [ftp://ftp.ncbi.nih.gov/genbank/]
247. **Reference sequence project sequence downloads.** [ftp://ftp.ncbi.nih.gov/refseq]
248. **Differential gene expression experiment using the online website** [http://sara.molbiol.ox.ac.uk/userweb/jherbert/cgi-bin/cdna_lib_stats.cgi]
249. **HTSeq: Analysing high-throughput sequencing data with Python** [http://www-huber.embl.de/users/anders/HTSeq/doc/count.html#count]