Investigation into Survival Mechanisms of Malignant B Cells in the Central Nervous System

A Cousins
MB ChB

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Institute of Cancer Sciences,

College of Medical, Veterinary and Life Sciences,

University of Glasgow

December 2018

Abstract

Introduction

Acute lymphoblastic leukaemia (ALL) is the commonest childhood cancer. In the early trials of ALL treatment, central nervous system (CNS) relapse was a common occurrence - the introduction of CNS-directed therapy in the 1970s was associated with the largest single improvement in outcome for childhood ALL. Today, despite universal intensive CNS-directed therapy - with significant associated toxicity - the CNS is involved in around 50% of ALL relapses, with approximately 50% of these being isolated CNS (iCNS) relapse. Whilst many factors increase risk of CNS relapse, few are specific for CNS relapse. Discovery of specific risk factors for CNS relapse would allow increased therapy for children at high risk, and potentially less CNS-directed therapy for those at low risk.

Relatively little is known about the biological differences between systemic and CNS ALL. In the CNS, leukaemic cells form plaques adherent to the leptomeninges, bathed in low-nutrient, low-oxygen cerebrospinal fluid (CSF). It was hypothesised that leukaemic cells adapt metabolically to this nutritionally poor CNS microenvironment, and these metabolic adaptations may be targets for specific therapy and/or specific biomarkers for CNS relapse.

Findings

Transcriptional analysis of ALL cell lines from CNS and spleen in a mouse xenograft model, and of ALL cells retrieved from the CSF at CNS relapse of ALL, have shown the upregulation of cholesterol biosynthesis as a key adaptation to the CNS niche. Analysis of transcriptomic data from the bone marrow or peripheral blood from children with ALL at diagnosis have shown the potential for upregulated cholesterol biosynthesis (and, independently, upregulated IL7R) as a significant risk factor for CNS relapse of ALL. To support this finding, metabolomic analysis found evidence of changes in CSF cholesterol in the presence of CNS leukaemia, and of increased mevalonate (a cholesterol precursor) and cholesterol in ALL cells retrieved from the CNS in a xenograft

model. Therapeutic targeting of CNS ALL *in vivo* with statins resulted in a CNS-specific increase ALL disease burden.

Untargeted metabolomic analysis of CSF shows differences between children with ALL (either at diagnosis or on maintenance therapy) and non-ALL controls, and between children with ALL at diagnosis and the same children on maintenance therapy. Creatine abundance was significantly different in children with ALL at diagnosis compared with both other groups (1/3 lower at diagnosis than either on maintenance or non-ALL controls). This change in creatine and persisted on analysis of CSF from mice with and without leukaemia. On analysis of CSF from children at CNS relapse with ALL there is evidence of increased reduced creatine at time of CNS relapse in 3 of 4 patients.

Conclusions

There is evidence to confirm the hypothesis that ALL cell adapt metabolically to the CNS niche. Cholesterol biosynthesis was identified as a key pathway upregulated in CNS ALL, and upregulated cholesterol biosynthesis in ALL cells at diagnosis was found to be a key risk factor for CNS relapse of ALL. In addition, clear changes in the CSF metabolome related to both ALL and ALL therapy were shown, and a new potential marker for the presence of CNS ALL identified. Prospective analyses in independent cohorts are required to determine the clinical utility of these novel strategies for prediction of CNS relapse risk.

Table of Contents

List of tables

List of figures

List of publications arising from this work

Interleukin 7 receptor is associated with central nervous system infiltration in pediatric B-cell precursor acute lymphoblastic leukemia

Alsadeq A, Lenk L, Vadakumchery A, Cousins AF, Vokuhl C, Khadour A, Vogiatzi F, Seyfried F, Meyer LH, Cario G, Hobeika E, Debatin KM, Halsey C, Schrappe M, Schewe D, Jumaa H

Blood 2018 132(15):1614-1617

Abstract publications:

Upregulated Cholesterol Biosynthesis is a Biomarker for Isolated Central Nervous System Relapse in Acute Lymphoblastic Leukaemia

Cousins AF, Markert E, Degn M, Olivares O, Yousafzai Y, Bhatti S, Herzyk P, Michie A, Schmiegelow K, Gottlieb E, Halsey C

Presented at the American Society of Hematology Annual Meeting and Exposition 2017, granted Abstract Achievement Award

Metabolic Analysis Reveals Cerebrospinal Fluid Creatine And Xanthine Concentrations as Potential Biomarkers For Central Nervous System Involvement In Acute Lymphoblastic Leukaemia

Cousins AF, Sumpton D, MacKay G, Däbritz HM, Michie A, Gottlieb E, Halsey C

Presented at the European Haematology Association 23rd Annual Congress 2018, awarded travel grant

Association of CNS involvement in childhood acute lymphoblastic leukaemia with cholesterol biosynthesis upregulation

Cousins AF, Olivares O, Michie A, Gottlieb E, Halsey C

The Lancet, Vol. 389, S35, Feb 2017

Presented at the Academy of Medical Sciences Spring Meeting 2017

Acknowledgements

I am extremely grateful to the William and Elizabeth Davies Foundation for their generous funding of this project.

In addition I would like to gratefully acknowledge the significant contributions from my supervisors Chris Halsey, Eyal Gottlieb and Alison Michie, and thank them for their regular input and advice, and also to thank my assessors Vignir Helgason and Mhairi Copland for their constructive feedback.

I would like to thank members of the Wolfson Wohl CRC, the Institute of Infection, Immunity and Inflammation, and Paul O'Gorman Leukaemia Research Centre (university of Glasgow) and the Beatson Institute for their help and support through this project. In particular I would like to thank Gillian McKay, David Sumpton and Niels Van Den Broek at the Beatson Institute Metabolomics Unit, the Gottlieb and Kamphorst groups at the Beatson Institute (particularly Henry Däbritz, Sergey Tumanov and Grace MacGregor), The Wolfson Wohl CRC Flow Cytometry lab, the Vetrie and Helgason groups (particularly Mary Scott, Rebecca Mitchell and Pablo Baquero), and the Graham group at the Institute of Infection, Immunity and Inflammation (particularly Laura Medina-Ruiz), and my lab member Orianne Olivares. I would like to acknowledge the help of collaborators in the IL7Rα manuscript.

I am very grateful to Karen Keeshan for her help with mouse work on her project licence, and Helen Wheadon for her help with Fluidigm multiplex PCR experiments. I gratefully acknowledge the help of the staff at the Beatson Institute Biological Services unit, and the help of Glasgow Polyomics, particularly Pavel Herzyk.

I would like to thank to Bloodwise (formerly Leukaemia and Lymphoma Research) for maintaining and allowing access to the Childhood Leukaemia Cell Bank, I thank Yasar Yousafzai, Saeeda Bhatti, Nicholas Heaney, Brenda Gibson, Wendy Taylor and Alison Spence for their work in establishing the CSF biobank for children with ALL. I am extremely grateful to the participating patients and families.

Finally I would like to acknowledge the help and support from Professor Tessa Holyoake who sadly passed away last year.

Author's Declaration

I declare that, except where explicit reference is made to the contribution of others, that this thesis is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Signature:………………………………………….

Printed name: Antony Francis Cousins

List of Abbreviations

6-MP	6-Mercaptopurine
ß2M	ß2-microglobulin
ADK	Adenylate kinase
AFP	α-fetoprotein
ALL	Acute lymphoblastic leukaemia
ASP	L-asparaginase
AST	Aspartate transaminase
BCP-ALL	B-cell precursor ALL
BM	Bone Marrow
CEA	Carcinoembryonic antigen
CK	Creatine kinase
CNS	Central nervous system
COG	Children's Oncology Group
CSF	Cerebrospinal fluid
ES-FCS	Embryonic Stem Cells tested-FCS
FCS	Fetal Calf Serum
GC-MS	Gas chromatography-mass spectrometry
GPI	Glucose phosphate isomerase

HCG	Human chorionic gonadotrophin
HIF	Hypoxia inducible factor
HMDB	Human Metabolome Database
Hmens	Primary Human Meningeal cells
HMGCR	HMG CoA reductase
HMGCS	HMG CoA synthase
HPLC	High-performance liquid chromatography
ICAM1	Intercellular adhesion molecule 1
IDH	Isocitrate dehydrogenase
IL-15	Interleukin 15
IL7R	Interleukin 7 receptor gene (gene coding for IL7Rα protein)
IL7Rα	Interleukin 7 receptor subunit alpha (combines with the common gamma chain to form the interleukin 7 receptor)
IRF4	Interferon regulatory factor 4
LC-MS	Liquid chromatography-mass spectrometry
LDH	Lactate dehydrogenase
MCM	Meningeal Cell Media
NSG	NOD.Cg-Prkdcscid Il2rgtm1WjI/SzJ
MEK	Mitogen-activated protein kinase/Extracellular signal regulated kinase Kinase

MERTK	Mer tyrosine kinase
MRD	Minimal residual disease
MTX	Methotrexate
NBF	10% Neutral buffered formalin
NCBI	(US) National Center for Biotechnology Information
NHL	Non-Hodgkin lymphoma
PBX1	Pre-B cell leukaemia homebox 1
PCR	Polymerase Chain Reaction
PMVK	Phosphomevalonate kinase
qPCR	Quantitative polymerase chain reaction
SCD	Stearyl CoA Desaturase
SCAP	SREBP cleavage-activating protein
SPP1	Osteopontin
SREBF	Sterol-regulatory element-binding transcription factor (note isoforms 1a, 1c and 2)
SREBP	Sterol-regulatory element-binding proteins
TBP	TATA-box Binding Protein
TLP	Traumatic Lumbar Puncture
VEGFα	Vascular endothelial growth factor α

Chapter 1: Introduction

1.1 Background

Acute lymphoblastic leukaemia (ALL) is the commonest cancer in childhood, and among the leading causes of cancer death in childhood. It is a cancer of immature lymphoid cells which have acquired genetic mutations leading to maturation arrest, proliferation and loss of apoptosis. This leads to the accumulation of ALL cancer cells (commonly referred to as blast cells due to their morphological features when examined under a microscope) many, if not all, of which are capable of acting as cancer stem cells. If untreated, ALL blasts progressively infiltrate the organ systems of the body – most markedly in the bone marrow which is the presumed origin site for most leukaemias - displacing normal cells and disrupting organ function.

The focus of this thesis is on childhood B-cell precursor ALL (all references to "ALL" in the thesis are to childhood precursor-B ALL unless specifically stated). This is the most common subtype of ALL, accounting for approximately 85% of cases. ALL usually presents with symptoms and signs of bone marrow failure – anaemia, bleeding or infection. It is treated with intensive multi-agent chemotherapy over 6-7 months, followed by 2 years (girls) or 3 years (boys) of less intensive "maintenance" chemotherapy.

Childhood ALL treatment is an excellent example of the success of evidence-based medicine. Through the use of large-scale randomised trials, long-term survival of children with ALL has increased from below 10% in the 1960s to over 90% today (Figure 1-1).

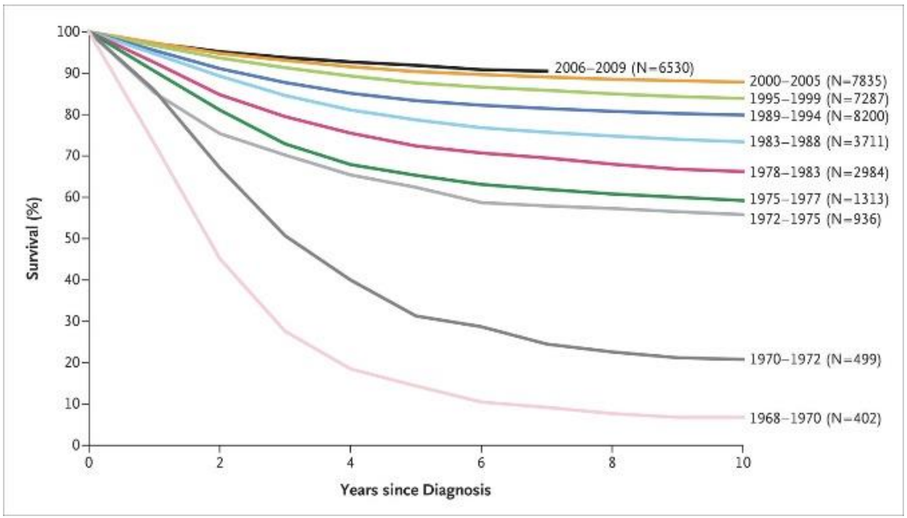


Figure 1-1 Overall survival among children with acute lymphoblastic leukemia (ALL) who were enrolled in Children's Cancer Group and Children's Oncology Group clinical trials, 1968–2009. (Reproduced with permission from *Hunger et at. N Engl J Med 2015; 373:1541-155*, © Massachusetts Medical Society)

Despite this success in improving outcome in ALL, there are still significant challenges that need to be addressed in ongoing research.

- Firstly, as noted above, despite this vast improvement in outcomes, ALL remains among the top causes of childhood death in the developed world and more research is needed as a priority to further improve disease outcome (0.1-0.5 per 100,000 deaths per patient-years at risk in 2016 (ISD Scotland), with an overall childhood mortality rate of around 10 per 100,000 children aged 1-15 in England 2016 (Office for National Statistics)).

- Secondly, the gains in survival which have been achieved have come mainly as a result of incremental intensification of treatment, which currently consists of up to 7 different chemotherapeutic agents, often given at maximum tolerated doses, for up to 3½ years. This intensification comes at the cost of increased treatment toxicity.

It is in this context that central nervous system (CNS) involvement with ALL is an extremely promising area of study for reasons explored below.

1.2 CNS involvement with ALL

1.2.1 History of CNS disease in ALL

Since the early 1950s it has been possible to induce morphological remission in children with ALL using chemotherapy including anti-metabolic chemotherapy (initially aminopterin(Farber et al. 1948), then 6-Mercaptopurine (6-MP)(Burchanal et al. 1953), and later methotrexate (MTX) and L-asparaginase (ASP)(Sutow et al. 1971)) – some early chemotherapy protocols and outcomes in the Southwest Cancer Chemotherapy Study Group (a predecessor to the current Children's Oncology Group) in the US have previously been summarised(George et al. 1973). These children would almost inevitably relapse, commonly with CNS involvement(Sullivan 1957) and had a poor overall survival.

Relapse with CNS involvement is known as CNS relapse. This can occur without evidence of relapsed ALL in the bone marrow, which is known as "isolated CNS relapse", or can occur with concurrent evidence of relapse in the bone marrow, which is known as "combined CNS relapse". As control of systemic ALL with chemotherapy improved it was noted that there was an increase in CNS relapse(Hardisty & Norman 1967).

It was postulated from the early descriptions of meningeal ALL that the therapy used at the time had only limited penetration into the CNS, allowing a "reservoir" of cancer cells to survive(Sullivan 1957). This led to the idea that the CNS is a special compartment of the body which acted as a sanctuary for leukaemic cells. In one 1960 post-mortem study of children who died with ALL, up to 70% had evidence of infiltration of the CNS with ALL blasts(Moore et al. 1960).

The largest single improvement in the outcome of children with ALL was achieved with the introduction of treatment directed at ALL blasts in the CNS to all children as part of routine schedules (indicated in (Figure 1-2)) by the Southwest Oncology Group "Sanctuary Therapy" trial in 1972-1974(Nesbit et al. 1982), and slightly earlier by the St Jude's Children's Research Hospital (Tennessee, US)-based "Total Therapy VII" trial in 1971(Simone et al. 1972).

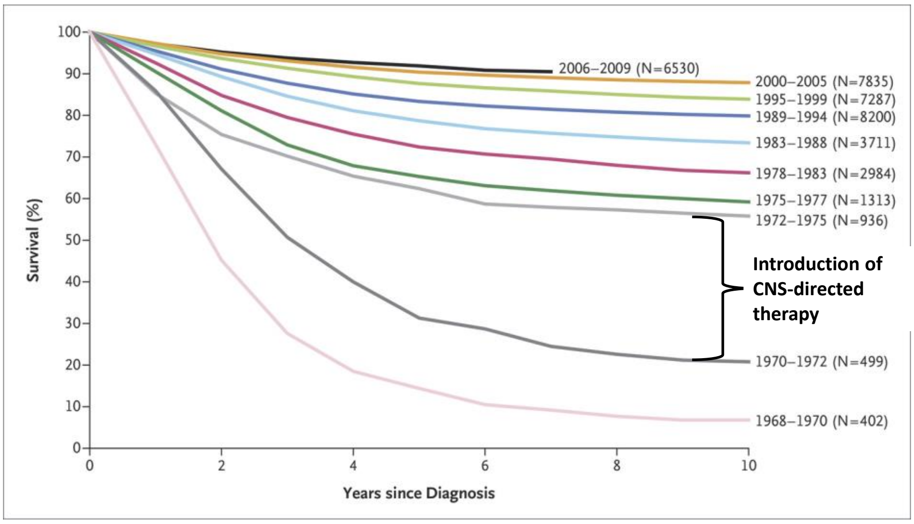


Figure 1-2 Overall survival among children with acute lymphoblastic leukemia (ALL) who were enrolled in Children's Cancer Group and Children's Oncology Group clinical trials, 1968–2009; introduction of universal CNS-directed therapy highlighted (adapted from *Hunger et at. N Engl J Med 2015; 373:1541-155*, © Massachusetts Medical Society)

This was initially done with cranial irradiation which is extremely effective, reducing overt CNS relapse rates from around 65% to less than 10% of children in the "Total Therapy VI" trial(Simone et al. 1972). Unfortunately this treatment is also extremely toxic. The side effects include cognitive impairment(Eiser 1978; Cousens et al. 1988; Jankovic et al. 1994; Halsey et al. 2011), hormonal dysfunction (secondary to pituitary gland damage), most prominently manifesting as reduced growth(Oliff et al. 1979; Clayton et al. 1988; Uruena et al. 1991), and an up to 20-fold increase in risk of secondary malignancy over 20 years follow up(Nygaard et al. 1991; Neglia et al. 1991; Pui et al. 2003). This risk is mainly due to an increase in secondary CNS cancers(Nygaard et al. 1991; Neglia et al. 1991; Rimm et al. 1987). These adverse effects appear to be particularly marked in children under 5 years of age who underwent cranial irradiation(Eiser & Lansdown 1977; Jankovic et al. 1994; Neglia et al. 1991).

One of the most important findings from the UK-based UKALL XI clinical trial in the 1990s was the success of replacing routine cranial irradiation with high-dose intravenous methotrexate and regular intrathecal administration of methotrexate even in high-risk patients(Hill et al. 2004) (and the introduction of high-dose dexamethasone in the subsequent ALL 97/99 trial(Mitchell et al. 2005)). While patients receiving cranial irradiation had a lower rate of CNS relapse, this was offset by an increased rate of systemic relapse and overall there was no benefit to event-free survival. This built on previous work from the Children's Cancer Study Group showing that cranial irradiation could be safely omitted in lower risk patients(Bleyer et al. 1991). It should be noted that methotrexate and dexamethasone, while not as toxic as cranial radiotherapy, do have significant side effect profiles.

Some of the common side effects of cranial irradiations, methotrexate and dexamethasone are listed below (Table 1-1). In addition, ALL survivors managed without cranial irradiation (i.e. with chemotherapy alone) have been shown to suffer long-term neurocognitive impairment(Moleski 2000).

Table 1-1 Table showing comparison of adverse events associated with cranial irradiation, dexamethasone and methotrexate therapy

Treatment	**Associated Adverse Events**
Cranial Irradiation (reviewed by Stone et al.(Stone & DeAngelis 2016))	Acute or subacute encephalopathy; Transient or progressive myelopathy; Cerebral necrosis; Late diffuse brain injury; Neuroendocrine disorders; Neuropathy; Vasculopathy; Secondary malignancy; Cognitive deficit;
Dexamethasone (manufacturer's information available at medicines.org.uk)	Electrolyte disturbance; Myopathy, osteoporosis and osteonecrosis; GI reflux, peptic ulceration, perforation, pancreatitis; Thinning of skin, reduced wound healing, acne, oedema; Increased intracranial pressure, psychiatric disturbance; Cushingoid, suppression of growth, secondary diabetes, pituitary-adrenal suppression, menstrual disturbance; Cataracts, glaucoma, exophthalmos, chorioretinopathy; Immunosuppression; Weight gain; Thromboembolism;
Methotrexate (manufacturer's	Severe (potentially fatal) skin reactions; Bone marrow suppression;

information available at medicines.org.uk)	GI mucositis; Hepatic toxicity; Renal toxicity; Pnuemonitis; CNS toxicity: stroke-like syndrome, seizures, impaired cognition, raised intracranial pressure, leukoencephalopathy; Pericarditis; Immunosuppression; Thromboembolism;

Despite universal intensive CNS-directed treatment – either historically with cranial irradiation or more typically now with combined high-dose systemic therapy and intrathecal therapy, CNS involvement is detected in around 50% of ALL relapses (i.e. 4% of children with ALL will have a CNS relapse), and isolated CNS relapse accounts for around 25% of relapses (2% of children with ALL)(Krishnan et al. 2010)..Improving our understanding of CNS ALL has potential therefore to both improve disease control by allowing design of better treatments for the children that relapse with CNS disease, and potential to reduce treatment toxicity by reducing the amount and intensity of CNS-directed therapy by identifying those at low risk of relapse. This principle was demonstrated in the UKALL 2003 trial, which successfully targeted patients with high-risk ALL for more intensive therapy while sparing low-risk patients by looking for very small amounts of residual cancer in the bone marrow (known as minimal residual disease (MRD))(Vora et al. 2013). It has been shown in this and other studies that the detection of disease at a low level at the end of the first month of intensive chemotherapy correlates with a higher risk of bone marrow relapse – and additionally that risk stratification based on this result can safely allow the escalation and de-escalation of treatment. There is currently no equivalent test for CNS ALL.

1.2.2 Biology of CNS ALL

The pattern of disease in CNS ALL remarkably consistent, with blast cells seen in the leptomeninges - between the arachnoid and pia layers of the meninges (membranes that surround the brain and spinal cord inside the skull/spinal canal), and adherent to these meningeal layers (explored in more detail below chapter 1.3.2). Involvement of the brain parenchyma is unusual and only seen as a very late sequelae of CNS ALL(Price & Johnson 1973).

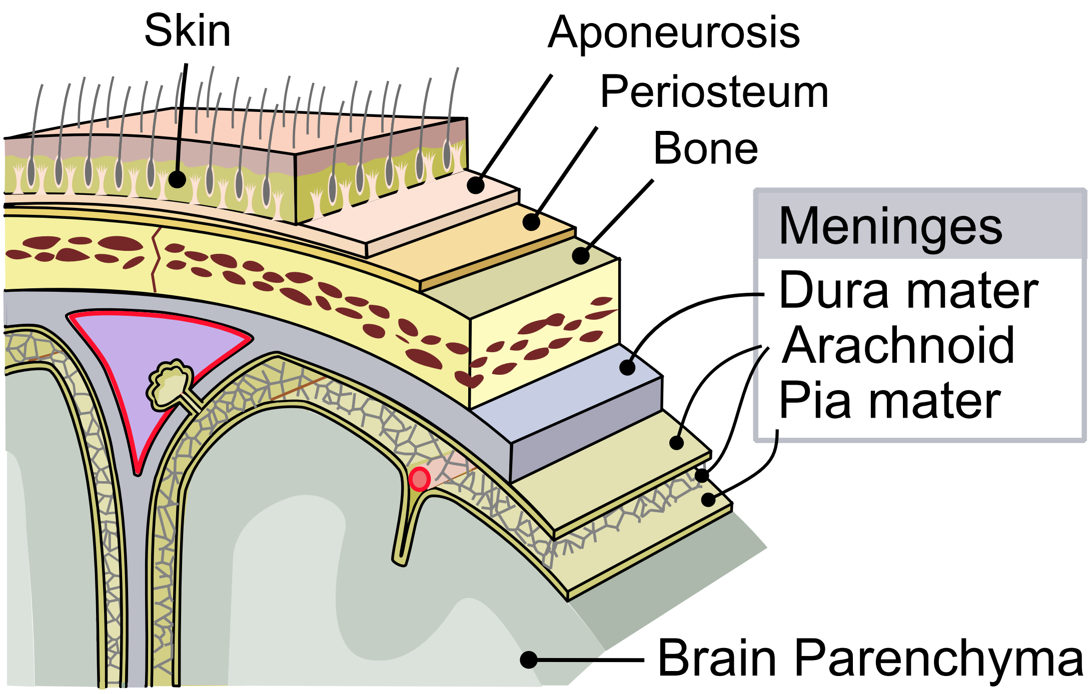


Figure 1-3 Schematic diagram of the layers of the skull and meninges adapted from "Meninges of the central nervous parts" by user Mysid based on work by the SEER development team, used under creative commons licence (https://commons.wikimedia.org/wiki/File:Meninges-en.svg)

Whilst there are some aspects that are well-studied, there are several questions about the biology of CNS ALL that are only part-answered.

1.2.2.1 Origin

It is likely that ALL cells in the CNS originate from the bone marrow (BM). This is supported by the fact the subclonal diversity of ALL in the CNS very closely mirrors that of the bone marrow in mouse xengraft models(Williams et al. 2016; Elder et al. 2017), and in primary samples(Bashford-Rogers et al. 2016; Bartram et al. 2018), that the bone marrow is the site of normal B-cell development, and that bone marrow is the predominant site of almost all cases of ALL at presentation. However, there are rare cases where ALL is detected in the CNS before bone marrow(Levine et al. 1973; Piovezani Ramos et al. 2016) though it is unclear if true "isolated" CNS ALL exists, or if in all cases of CNS ALL there is some bone marrow involvement albeit on or below the limits of detection with current technology as suggested by very sensitive next-generation-sequencing techniques(Bashford-Rogers et al. 2016; Hagedorn et al. 2007; Bartram et al. 2018). Additionally there is evidence of extramedullary haematopoeisis in the choroid plexus(Tabesh et al. 2011; Eskazan et al. 2012; Janssen et al. 2013), which makes a CNS-origin in some cases of ALL more feasible. On the balance of evidence it seems likely that CNS ALL arises from lymphocytes originating from the BM in the vast majority of cases.

1.2.2.2 Transit

It remains unproven how ALL blasts transit to the CNS. There is a significant body of literature exploring movement of blasts across models of the blood-brain barrier(Mielcarek et al. 1997; Holland et al. 2011; Akers et al. 2010; Feng et al. 2011; Kondoh et al. 2014), though there is little evidence to support whether cells enter the CNS via this route *in vivo*. There is, however, evidence of the transit of cells across the blood-cerebrospinal fluid (CSF) barrier through cranial veins, and of cells entering the subarachnoid space via the CSF-producing choroid plexus(Ransohoff et al. 2003; Kivisäkk et al. 2003). In addition there is evidence that cells are able to infiltrate the CNS via the dura layer of meninges in humans (adherent to the cranial bones), giving the possibility of cells transiting from cranial bone marrow to the leptomeninges(Bleyer 1989) - this is clearly seen in modern mouse models of CNS ALL. Finally there has recently been some evidence of lymphatic channels in the meninges which may provide another method of ALL cell transit(Louveau et al. 2015).

It is also unclear why cells transit to the CNS. There is no evidence in precursor-B ALL for a distinct chemokine signature allowing blasts to infiltrate the CNS (as opposed to T-cell ALL where CCR7 expression is required for CNS infiltration)(Williams et al. 2016; Buonamici et al. 2009). In addition there is evidence from murine models that the ability to infiltrate the CNS is a property common to almost every precursor-B ALL clone(Williams et al. 2016). This fits with the high degree of CNS infiltration seen in children who died with ALL in the 1960s(Price & Johnson 1973). It has traditionally been thought that specific receptor-ligand interactions would be required to permit ALL cells to enter the

CNS, but it may be that transiting to the CNS is a common feature of ALL blasts (and perhaps all B cells) requiring no special cell signals.

1.2.2.3 Behaviour of ALL in the CNS

There is also conflicting evidence of the behaviour of ALL cells once they reach the CNS. ALL cells in the CNS lie adherent to the meninges bathed in CSF. This is a potentially hostile environment (discussed in more depth below), and there is evidence that ALL cells in the CNS have lower proliferation (measured by ^{3}H-thymidine and ^{3}H-uridine incorporation) and are smaller than in the blood or bone marrow, though still maintaining viability(Huei-Mei Kuo et al. 1975; Tsuchiya et al. 1978; Bleyer 1989).In some patients, CNS relapse can be a very late event raising the possibility the cells causing relapse have been proliferating extremely slowly for a long time in the CNS. This, together with the evidence above, all suggests that the cells in the CNS may acquire a more quiescent phenotype.

The evidence overall is however not entirely consistent – CNS ALL is frequently widespread with large amounts of disease present before it can be detected clinically(Price & Johnson 1973; Glass et al. 1979). The presence of large amounts of CNS could be explained by continual seeding of the CNS from systemic disease leading to progressively increasing disease bulk, or the disease being present in the CNS for a long time slowly proliferating, however CNS relapse of ALL is frequently a relatively early event (less than 1 year) after the start of chemotherapy and induction of remission. This suggests that despite the absence of systemic disease capable of seeding the CNS, and despite ongoing CNS-directed therapy, and despite a relatively short amount of time(Roy et al. 2005), ALL cells in the CNS are in some cases able to proliferate rapidly enough to cause overt disease relapse.

It may be that there are distinct phenotypes of ALL cells in the CNS – some which are able to proliferate relatively rapidly and are responsible for early relapse, and others with very low levels of proliferation which are responsible for late relapse, but there is no evidence to support this hypothesis so far.

1.2.3 Molecular adaptations

There are some data available regarding transcriptional adaptations of ALL cells to the CNS environment. These often involve cell adhesion and signalling changes. The most convincing findings are described below.

Vascular endothelial growth factor α (VEGFα): this forms part of the hypoxia response - as discussed below (chapter 1.3.2), the CSF is a hypoxic environment. VEGFα gene expression was shown in two recent papers to be elevated in mouse xenograft models of CNS ALL (along with other metabolic adaptations). In addition, treatment of mice with anti-VEGFα antibody bevacizumab reduces CNS disease burden in ALL murine xenografts(Kato et al. 2017; Münch et al. 2017).

Osteopontin (SPP1): A secreted protein involved in modulating cell migration and adhesion with anti-apoptotic properties. Elevated osteopontin levels in the CSF have been correlated with CNS involvement in ALL, and ALL cells retrieved from the CSF of children have been shown to express high levels of osteopontin(Incesoy-Özdemir et al. 2013.; van der Velden et al. 2015). Interestingly, plasma osteopontin levels have been correlated with cholesterol metabolism(Isoda et al. 2003; Luomala et al. 2007; Yang et al. 2012).

Interleukin 15 (IL-15): A cytokine mainly recognised as a regulator of T and NK cell activation. High expression of IL-15 in the bone marrow at diagnosis was found to be associated with CNS ALL at diagnosis (determined using cytospin – see below chapter 1.2.4.1), with a possible link between higher IL-15 expression in bone marrow at diagnosis and subsequent CNS relapse in patients initially CNS-negative. In addition, exposure to IL-15 *in vitro* promoted ALL cell line growth under low-serum conditions, and may induce expression of cell trafficking genes associated with CNS disease(Cario et al. 2007a; Williams et al. 2014).

Intercellular adhesion molecule 1 (ICAM1): A protein involved in immune cell adhesion and transmembrane migration. Increased expression in CNS ALL cells has been associated with CNS involvement in ALL cell line xenograft models(Mielcarek et al. 1997; Holland et al. 2011).

Mer tyrosine kinase (MERTK): MERTK expression has been found to be elevated in t(1;19) ALL patients who have a slightly increased risk of CNS relapse. Within this subset of ALL, Mer signalling induced a quiescent phenotype *in vitro*, and ALL cells with higher MERTK expression showed increased CNS disease burden in murine xenograft models(Krause et al. 2015).

Pre-B cell leukaemia homebox 1(PBX1): A transcription factor that is a translocation target in t(1;19) ALL. Patients with this translocation are at higher risk of CNS (but not systemic) relapse(Jeha et al. 2009), and increased expression of PBX1 has been shown to be related to chemotherapy resistance, and moderately increased CNS disease in a cell line murine xenograft model(Gaynes et al. 2017). Interestingly the CNS relapses seen with t(1;19) are often late CNS relapses, which may have implications for the effects of PBX1 on cell biology.

Stearyl CoA Desaturase (SCD): An enzyme involved in lipid desaturation that was found to have increased expression in cells retrieved from the CNS of children with ALL at CNS relapse, and retrospectively SCD detection by flow cytometry at diagnosis was found to be a risk factor for CNS relapse(van der Velden et al. 2015).

Finally, ALL cells with RAS pathway activation appear to be enriched at relapse, particularly CNS relapse, and disruption of this pathway with mitogen-activated protein kinase (MEK) inhibitor Selemetinib is effective in reducing disease burden systemically and in the CNS(Irving et al. 2014).

1.2.4 Clinical aspects of CNS ALL

1.2.4.1 Detection

The standard method for detection of CNS ALL is CSF cytocentrifugation – a technique in which CSF is spun in a centrifuge onto glass slides, then excess fluid removed in order to concentrate cells, and the slides stained and analysed by microscopy.

Despite excellent specificity of this technique for the presence of CNS ALL (i.e. the presence of leukaemic cells in the CSF indicates a very high probability of

CNS infiltration(Glass et al. 1979)), it has been known for many years that this technique has very poor sensitivity – in a 1979 study, for example, of 9 people with known CNS leukaemia and CSF cytology, only 4 were found to have leukaemic cells in the CSF(Glass et al. 1979). In another study of leptomeningeal malignancy with solid tumours, 3 sequential CSF samples would have a sensitivity to detect meningeal cancer of only 90% - i.e. in children with CNS leukaemia 10% would be considered negative by cytospin even after 3 sequential CSF samples were analysed(Wasserstrom et al. 1982). This fits with data from assessment of CSF with more modern techniques (discussed below). Given the histological findings in leptomeningeal ALL of cells adherent to the meningeal layers, and the relative paucity of ALL cells in the CSF, it is reasonable to assume that only a small proportion of CNS cells will detach into the flow of CSF at any time, and that these cells may represent the less viable portion of CNS ALL. It is interesting to note that in the initial description of cytospin for detection of CNS disease in 1974 (i.e. the era before universal CNS-directed therapy), 65% of samples were positive for leukaemic blasts, perhaps indicating that the CNS involvement at that time was more widespread, or that the time from disease onset to diagnosis was longer(Evans et al. 1974).

CSF is examined by cytospin in every child at diagnosis with ALL and classified into 5 groups (Table 1-2).

Table 1-2 Table showing classification of cytospin findings in ALL

CNS status	Cytospin finding
CNS 1	No evidence of blast cells in the CSF, <10 red blood cells/μL
TLP –ve (traumatic lumbar puncture; no blasts)	No evidence of blast cells in the CSF, >10 red blood cells/μL
CNS 2	Blast cells in the CSF, <5 blast cells/μL, <10 red blood cells/μL
TLP +ve	Evidence of blast cells in the CSF, >10 red blood cells/μL
CNS 3	Blast cells in the CSF, ≥5 blast cells/μL, <10 red blood cells/μL

There is good evidence that the presence of leukaemic blasts in the CNS (particularly CNS 3 disease) is associated with increased risk of relapse overall, with CNS 2 patients particularly having a relatively increased risk of CNS relapse(Mahmoud et al. 1993; Burger 2003), summarised by Pui et al.(Pui 2006) . This was not found in all studies, perhaps reflecting differences in treatment protocols(Gilchrist et al. 1994; Dutch Childhood Oncology Group et al. 2006), and there is evidence that this risk can be abrogated by the introduction of additional CNS-directed therapy in the form of additional intrathecal doses of methotrexate(Burger 2003). These children therefore usually receive additional CNS-directed therapy which is effective at reducing risk of CNS relapses. In the latest UKALL trail this takes the form of weekly intrathecal chemotherapy until two sequential CSF samples are negative by cytospin (although as noted above, cytospin has a low sensitivity for detecting CNS ALL).

More modern approaches appear to hold some improvements on traditional CSF cytospin. Flow cytometry is able to improve the sensitivity of cytospin - e.g. in a 2007 study of 219 patients with haematological malignancy treated in The Netherlands, flow cytometry detected malignant cells in 44/219 patients at diagnosis (20.1%) vs 19/219 (8.7%) for cytospin(Bromberg et al. 2007). A slightly concerning result in this study (and similarly in other studies) was that 4 children had CSF considered positive for CNS disease by cytospin but not flow cytometry. As noted above, it is known that CSF positivity by cytospin confers increased risk of CNS relapse, but there is no good evidence as yet that CSF positive by flow cytometry confers this increase in risk. A recent study by Levinsen et al. in the NOPHO group showed similar sensitivity in CSF flow cytometry, and showed 5 of 52 patients with CSF blasts at diagnosis had not cleared their CNS disease by day 15, but there is no prognostic data as yet for these patients(Levinsen et al. 2014).

Another technique which may improve sensitivity further is detection of submicroscopic levels of disease in the CSF using polymerase chain reaction on ALL cell DNA. In one study of 30 children in New York, 6/30 (20%) children at diagnosis had evidence of CNS disease by CSF PCR compared with 2/30 (6.7%) by conventional cytology(Pine et al. 2005). In another short series carried out in Glasgow, 15/38 (39.5%) patients were positive by PCR and 4/38 (7%) positive (TLP+) by standard cytospin(Yousafzai 2015). In this series, one patient was positive by cytospin but not PCR. Again there is no evidence as yet to determine if PCR positivity in the CSF confers an increased risk of relapse.

Finally, there is limited evidence that in difficult cases imaging of the CNS may help in the detection of occult CNS infiltration with ALL(Chamberlain et al. 2009).

A recent paper has been published suggesting that the ability to colonise the CNS is a property common to all B-cell precursor ALL (BCP-ALL) cells, and that many if not all children with BCP-ALL may have CNS leukaemia at diagnosis, but this is not detected due to the limitations of the techniques discussed above(Williams et al. 2016).

1.2.4.2 Indirect biomarkers

There have been several candidates for indirect biomarkers of CNS infiltration with ALL. These markers are theoretically attractive – as they do not depend on capturing a small number of cells circulating in the CSF they may be more sensitive, and more representative of the bulk leptomeningeal disease burden which is tightly adherent to the meninges. Unfortunately, no indirect markers assessed so far have been sufficiently sensitive or specific for routine clinical use.

Testing of CSF glucose concentration does not provide any predictive information for CNS relapse. Measuring CSF protein does provide some information – CSF protein is consistently increased in the presence of a high CSF white cell count. Unfortunately this is less sensitive and specific than standard cytospin examination of CSF(Evans et al. 1974).

β_2-microglobulin (β_2M) is a biomarker which has shown promise in CNS ALL. B_2M has been found to be increased in the CSF in the presence of leukaemia, and on serial monitoring of CSF β_2M has been shown to increase prior to development of frank cytospin-positive CNS relapse(Warrier et al. 1981; Vicente et al. 1982), though this was not found in all studies(Clausen & Ibsen 1984; Pudek et al. 1985). Additionally, β_2M is not specific to ALL, and elevated β_2M has only a 50% positive predictive value for CNS relapse in ALL even in the most promising studies(Hansen et al. 1991).

Other biochemical tumour markers including carcinoembryonic antigen (CEA)(Clausen & Ibsen 1984), human chorionic gonadotrophin (HCG), and α-fetoprotein (AFP)(Domaniewski et al. 1987) have been assessed for suitability as markers of CNS disease but not found to be useful.

Several metabolic enzymes and metabolites have been explored as potential markers of CNS malignancy. Lactate dehydrogenase (LDH) is often elevated in blood in the presence of cancer, particularly high-grade haematological malignancy like ALL(Rao et al. 2012; Kovesi & Hsu; Van Zanten et al. 1986). There is some evidence that CSF LDH increases in the presence of CNS malignancy, but overall the evidence is mixed, likely due to changing normal ranges with age, different isoforms of LDH and different measurement techniques. Overall, raised LDH has a low sensitivity (66%) and specificity (62.5%) for the presence of CNS ALL, and therefore is not suitable as a marker of CNS disease in ALL(Seidenfeld & Marton 1979).

Creatine kinase (CK) (which catalyses the transfer of phosphate between ADP and creatine phosphate) has been found to be elevated in the CSF of patients with malignant brain tumours, usually involving destruction of the brain parenchyma, but is also raised in other neurological disorders so is non-specific(Herschkowitz & Cumings 1964).

Aspartate transaminase (AST), glucose phosphate isomerase (GPI), isocitrate dehydrogenase (IDH), and adenylate kinase (ADK) have also been investigated as potential markers of CNS cancer, though none have been found to be clinically useful(Seidenfeld & Marton 1979).

1.2.4.3 Cholesterol

One metabolic marker which has been clinically useful in the diagnosis of brain tumours was desmosterol (a precursor for cholesterol). Desmosterol levels were measured in a “sterol test” after 5 days of triparenol – a drug which prevents conversion of desmosterol to cholesterol - leading to its accumulation. If desmosterol levels were increased the diagnosis of a malignant brain tumour was more likely, and the test could be used to monitor for recurrence(Paoletti et al. 1969; Seidenfeld & Marton 1979). Some groups found simply measuring cholesterol to be as effective as a marker of CNS cancer(Fleisherl et al. 1979). Improvements in imaging have made this test redundant for solid brain tumours, and there is no evidence of the “sterol test” being useful to detect metastatic CNS tumours or CNS ALL. It is known that systemic cholesterol metabolism is disrupted in many cancers including ALL(Moschovi et al. 2004).

1.2.4.4 Cholesterol biosynthesis

Cholesterol biosynthesis is regulated by sterol-regulatory element-binding proteins (SREBP), which stimulates transcription of genes involved in cholesterol biosynthesis in response to reduced intracellular cholesterol. SREBP is normally located in the endoplasmic reticulum (ER) with chaperone protein SREBP cleavage-activating protein (SCAP). In conditions where there is reduced intracellular cholesterol, SCAP is able to transport SREBP from the ER to the golgi apparatus, where it is activated by proteolysis. Activated SERBP binds to sterol-responsive elements (SREs) in the DNA and activate transcription of genes involved in cholesterol and fatty acid synthesis. An additional level of control of SREBP function is at the transcriptional level, where SREBF (the gene coding for SREBP) is regulated via extracellular signalling via insulin and oxysterol LXR/RXR receptors. There are 3 forms of SREBP – SREBP1a predominantly activates genes involved in fatty acid synthesis, SREBP2 predominantly activates genes involved in cholesterol biosynthesis and SREBP1c activates both pathways similarly (Goldstein & Brown 1990; Hua et al. 1993; Repa et al. 2000; Azzout-Marniche et al. 2000; Horton et al. 2002).

Cholesterol is synthesised from acetate through the mevalonate and sterol pathways (Figure 1-4), and the rate of cholesterol synthesis in humans can be estimated by the abundance of plasma mevalonate(Parker et al. 1984). . Cholesterol biosynthesis intermediates geranyl and fernasyl and involved in protein prenylation and cell signalling. Cholesterol itself is a lipid molecule which forms an integral part of cell and organelle membranes and has a key role in cell signalling via lipid raft formation and also has a role in storage of other lipids as cholesterol esters. In addition cholesterol metabolism can be adapted for the production of steroid hormones(Ikonen 2008).

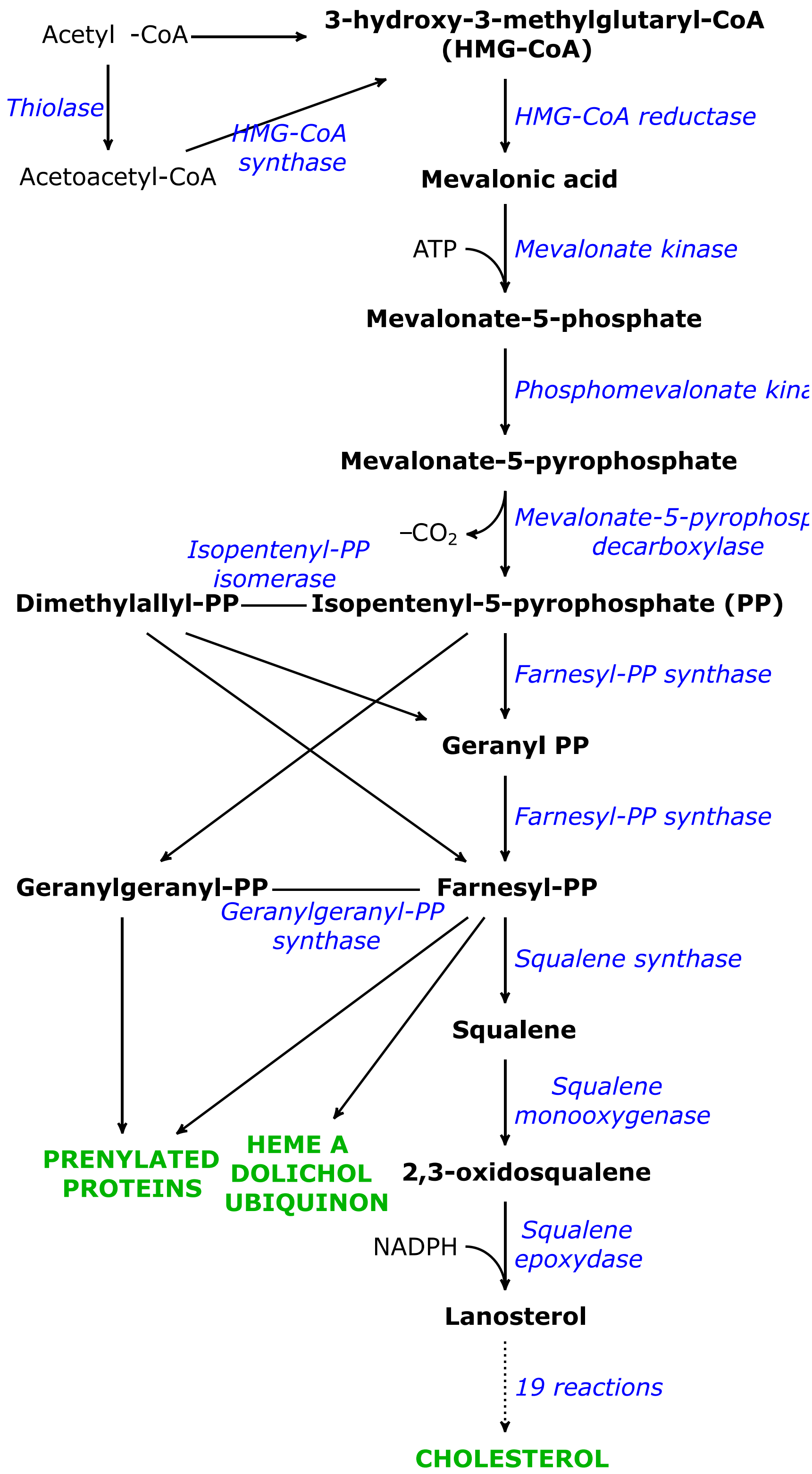


Figure 1-4 Figure showing cholesterol biosynthesis pathway adapted from "HMG-CoA reductase pathway" by user Krishnavedala used under a creative commons licence (https://commons.wikimedia.org).

1.3 ALL and metabolism

1.3.1 Overview of metabolism

Metabolism is the sum of the cellular processes that occur in living cells and organisms to maintain life. This includes hundreds of processes from generation of energy stores, to control of oxygen radicals, to synthesis of structural building blocks, to breakdown of obsolete structures and molecules and many more. Many metabolic processes can be measured either directly or indirectly, and by investigating the metabolism of cells in health and disease it is possible to understand pathology and allow rational interventions.

1.3.2 Normal leptomeningeal physiology

As discussed above (chapter 1.2.2), CNS ALL is found in the leptomeninges surrounding the brain and spinal cord. There are 3 meningeal layers: the pia mater is a thin layer of cells tightly adherent to the surface of the brain and spinal cord, the arachnoid mater a thin layer of cells and connective tissue beyond the pia and adherent to the third layer, the dura mater which is a thick fibrous sheath adherent to the skull and spine. The pia and arachnoid together are referred to as the leptomeninges. Between these two layers is an acellular space bathed in cerebrospinal fluid – it is here that leukaemic blasts are found in CNS ALL. There are no blood vessels within the leptomeninges; vessels carrying blood from the brain pass through small trabeculae which span the space between the pia and arachnoid mater, surrounded by the intact meningeal layer (Figure 1-3).

CSF is produced mainly in the choroid epithelial cells which form an interface between the systemic blood circulation and the CSF. In adult humans there is around 150ml of CSF at any one time, with around 25ml produced and 25ml drained into the systemic circulation per hour. CSF drains into the venous circulation via arachnoid granulations which act as valves to allow CSF to flow into the venous sinuses of the cranium, without reflux of venous blood into the leptomeninges. The role of CSF in the body includes providing a shock-absorber for the brain, helping to regulate intracranial pressure, and draining waste products from the brain(Sakka et al. 2011).

As would be expected for a fluid in a normally very hypocellular compartment, despite a comparable sodium concentration and osmolarity to plasma, normal CSF has a low abundance of most metabolites compared to plasma (i.e. the non-cellular liquid of the blood). For example: glucose concentration(Di Terlizzi & Platt 2006) and oxygen pressure are typically around two-thirds of that found in venous plasma(Venkatesh et al. 1999; Zaharchuk et al. 2005) (corresponding to oxygen pressure around half that of arterial plasma), protein levels are around 1% of plasma(Di Terlizzi & Platt 2006) and lipid/cholesterol levels are very low at around 0.1% of plasma(Illingworth & Glover 1971). It is worth noting that CSF is a 150ml fluid without haemoglobin (to which up to 95% of oxygen in the blood is bound within red blood cells), and these measures detect unbound oxygen. The true abundance of oxygen in the CSF is therefore much lower than that of blood.

A particular metabolite with interesting kinetics in the CNS is cholesterol. Cholesterol is relatively abundant in the CNS as a whole – mainly in the myelin of neural sheaths in the brain – but is extremely scarce in the CSF(Illingworth & Glover 1971). The cholesterol in the CSF consists of a specific metabolite (24-hydroxycholesterol) which is produced only in the CNS, mainly by glial neuron-supporting cells, with no systemic cholesterol detected(Vance et al. 2005). In addition the lipoproteins which normally transport cholesterol to the systemic circulation are not found in the CNS, with CNS-specific lipoproteins detected in the CSF(Pitas et al. 1987; Koch et al. 2001).

1.3.3 Oncometabolism

CNS ALL cells therefore grow in an environment with limited nutrient availability. They appear tightly adherent to the meningeal layers and it is possible that some nutrients are obtained from surrounding meningeal stromal cells, but despite this it is likely that CNS ALL cells are in a more nutritionally poor environment than systemic ALL cells which are able to draw nutrients from the circulating blood.

There is a large body of literature regarding metabolic adaptations of cancer cells reviewed by Vazquez et al.(Vazquez et al. 2016) Many of these metabolic changes are determined at a genome level – i.e. natural selection of cancer cell subclones best adapted to survive in a particular environment – some common genetic adaptations were recently reviewed by De Berardinis and Chandel(DeBerardinis & Chandel 2016). The most famous oncometabolic adaptation is aerobic glycolysis first described by Warburg(Warburg 1956), where cancer cells increase glycolysis and reduce TCA metabolism despite adequate oxygen levels. This results in less efficient ATP production from glucose but allows a much more rapid glycolytic flux and results in increased production of metabolic building blocks and reduced generation of free radicals.

In ALL the evidence suggests that the genomic landscape of ALL cells in the CNS is very similar to that in the systemic circulation, particularly when it comes to looking at which subclones of ALL are present in each compartment. This suggests that there is no genomic adaptation or natural selection of "CNS-avid" ALL subclones(Bashford-Rogers et al. 2016; Elder et al. 2017; Bartram et al. 2018).

There is, however, evidence that normal lymphoid cells are able to markedly adapt their metabolism transcriptionally between activated and quiescent states, sometimes in ways reminiscent of cancer cell metabolism with aerobic glycolysis and glutaminolysis, and changes in the metabolism of lipids. It may be that this transcriptional adaptation is sufficient for ALL cells to survive in the CNS(Ganeshan & Chawla 2014). Indeed as noted previously the expression of VEGFα is increased by ALL cells in the CNS, likely as a response to hypoxia(Kato et al. 2017; Munch et al. 2017).

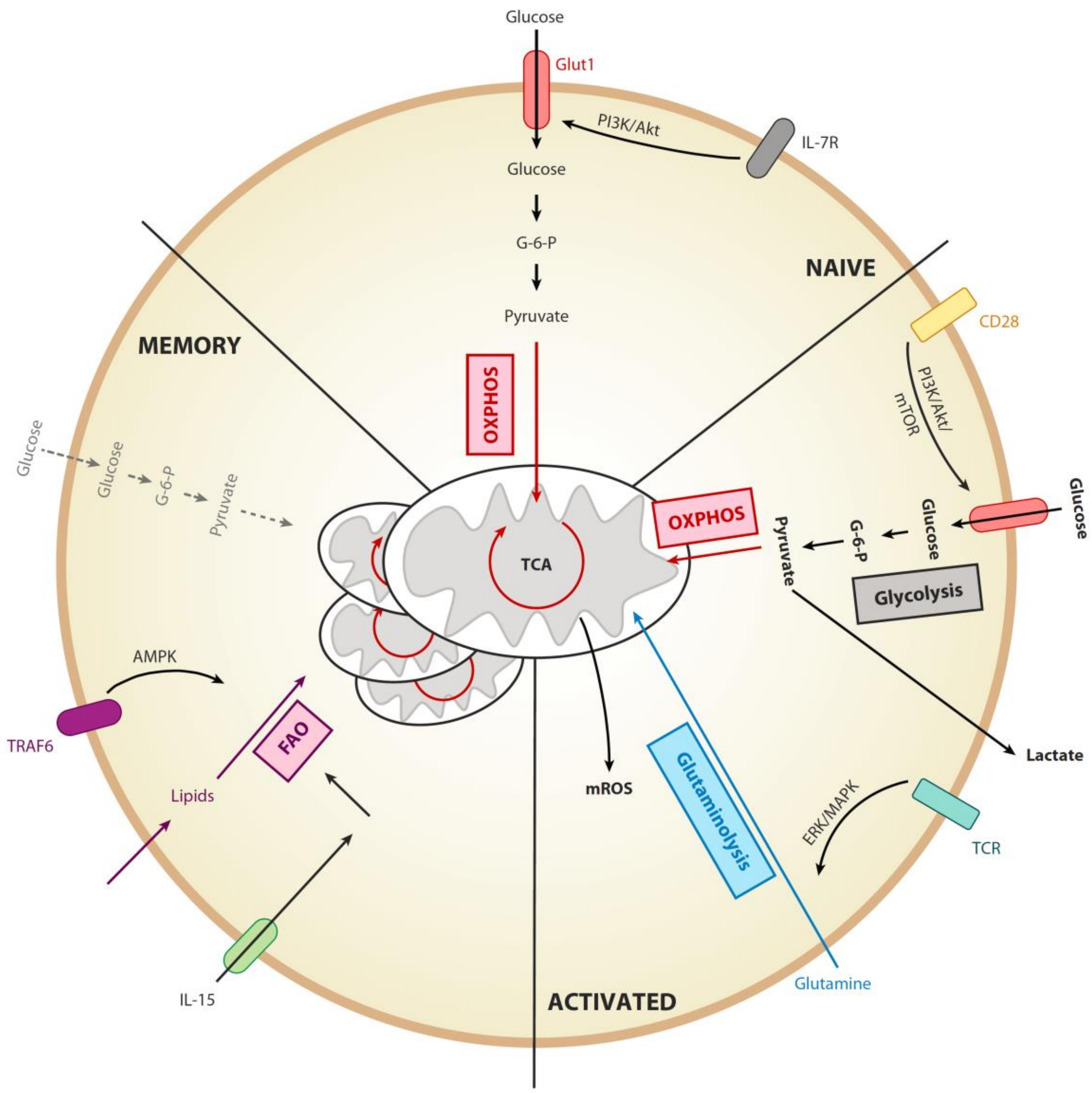


Figure 1-5 Schematic diagram of naïve, activated and memory T-cell metabolism(reproduced with permission from *Ganeshan + Chawla, Annu. Rev. Immunol. 2014. 32:609–34*, © Annual Reviews)

Targeting ALL cell metabolism has been a mainstay of ALL therapy since very early days in treatment. Some of the main anti-ALL therapies which form the backbone of current treatment regimens are anti-metabolite drugs: methotrexate disrupts folate/one-carbon metabolism, 6-mercaptopurine disrupts nucleotide metabolism, and asparaginase disrupts metabolism involving the amino acids asparagine and glutamine. Methotrexate and asparaginase in particular are among the most effective drugs for preventing CNS relapse in ALL. It may be that in the nutritionally poor CNS microenvironment, ALL cells are more sensitive to therapeutic metabolic disruption.

1.4 Hypothesis and aims

To summarise: ALL is the most common childhood cancer in the developed world, and is among the leading causes of cancer death in childhood. The major areas that can be improved in ALL treatment are, as noted above, improving disease control and reducing the toxicity of treatment.

Improving the understanding of CNS disease in ALL offers an opportunity to address both of these areas – CNS ALL is found to be involved (by cytospin) in 50% of ALL relapses. It is possible that the true incidence of CNS relapse may be higher given the known poor sensitivity of cytospin. In addition, CNS-directed therapy carries a significant toxicity burden.

The crux of these issues is knowing which children should receive more CNS-directed therapy to prevent relapse, and which could receive less. In addition understanding CNS leukaemia better may allow identification of critical weaknesses of ALL in the CNS to allow more targeted (and hopefully less toxic) treatments.

CNS ALL metabolism may be a key to improving both of these issues. CNS ALL blasts are found in a nutrient-poor environment, raising the possibility the cells are under metabolic stress which could be targeted with therapy. Additionally the presence of ALL cells in a normally acellular environment may provide a metabolic signature of ALL in the CSF – which can fairly easily be collected and analysed.

My hypothesis is:

ALL cells adapt metabolically to survive in the CNS microenvironment.

My research aims are:

1) To identify metabolic adaptations of ALL cells to the CNS microenvironment that could be targets for CNS-directed therapy

2) To identify new metabolic markers of CNS involvement in the CSF

3) To investigate directly metabolic changes of ALL cells in the CNS identified in the course of this investigation

Chapter 2: Materials and Methods

2.1 Materials and Supplies:

2.1.1.1 List of Suppliers and associated materials/supplies

Table 2-1 List of suppliers and associated materials/supplies

Supplier	Base location	Material Supplied	Catalogue number
ATCC	Virginia, USA	HS5 - Normal adult human bone marrow stroma	ATC-CRL-11882
Biolegend UK Ltd.	London, UK	PE/Cy7 anti-human CD19 antibody APC-Annexin V Apoptosis Detection Kit	302216 640932
BioRad Laboratories Ltd.	Hemel Hempstead, UK	Ssofast EvaGreen Low Rox Supermix Kit	1725210
Cambridge Isotope Laboratories Inc. (via CK isotopes Ltd, Ibstock UK)	Massachusetts, USA	D-Glucose ($U^{13}C6$ 99%) Sodium Acetate ($1,2^{13}C$ 99%)	CLM-1396 CLM-440
Corning Inc.	New York, USA	All plastics unless otherwise stated	
DSMZ	Leibniz, Germany	SEM - Immortalised childhood ALL cells with t(4;11) REH - Immortalised human childhood ALL cells with t(12;21)	AC-546 AC-22
E&O Laboratories, Ltd.	Bonnybridge, UK	Luria Bertoni Broth	BM5300
Elkay Laboratory Products (UK) Ltd.	Basingstoke UK	Flow Cytometry (5ml) tubes	
Eppendorf	Hamburg, Germany	Microcentrifuge tubes	
Fluorochem Ltd.	Hadfield, UK	Simvastatin	M03997
Labtech International Ltd.	Heathfield, UK	Embryonic Stem Cells tested-FCS (ES-FCS)	1001S/500
Lonza	Basel, Switzerland	MycoAlert Detection Kit	LT07-218

Miltenyi Biotec Ltd.	Bisley, UK	anti-CD45-FITC, mouse Anti-Ter119-Vioblue, mouse Red Blood Cell Lysis Solution	130-102-491 130-102-208 130-094-183
New England Biolabs	Massachusetts, USA	Exonuclease I (E. coli) Phusion High-Fidelity DNA Polymerase	M0293S M0530S
Qiagen	Hilden, Germany	RNEasey kits (micro, mini, midi) Multiplex PCR kit RNAse-free DNAse kit QIAquick Gel Extraction Kit QIAshredder	74004, 75142, 74104) 206143 79254 28704 79654
Santa Cruz Biotechnology, Inc.	Dallas, Texas	Pseudouridine (chemical standard) HMGCR Double Nickase Plasmid (with Control Nickase Plasmid, UltraCruz® transfection Reagent, Plasmid Transfection Medium) Puromycin	sc-291984 sc-400560 (sc-437281, sc-395739, sc-108062) sc-108071,
Sciencell Research Laboratories Inc. (via Caltag Medsystems Ltd. Buckingham, UK)	California, USA	Meningeal Cell Media Trypsin/EDTA Human meningeal cells	1401 0103 SC-1400
Sigma- Aldrich	Missouri, US	Oligonucleotides/PCR Primers HMGCR shRNA pLKO-plasmid bacterial glycerol stocks Methylcellulose MTFA Chemical standards (NG-NG Dimethyl-L-Arginine, Xanthine, N4-Acetylcytosine)	 SHCLNG-NM_000859 M7140 M7891 D0390, X0626, M4254, 377910,

		Chloroform (HPLC Plus)	650498
		(RS)-mevalonic acid lithium salt	90469
		6-Mercaptoethanol	M6250
		Demecolcine	D1925
		Cholesterol (Water Soluble Bioreagent)	C-4951
		Methanol	494291
		2-Propanol (Isopropanol)	I9516
		Calcium Chloride	C1016
		Trypan Blue	T8154
Starlab	Milton Keynes, UK	A pipette tips	
StemCell Technologies	Grenoble, France	Lymphoprep	07811
ThermoFisher Scientific (Includes subsidiaries: Invitrogen, Gibco, Life Technologies, eBioscience)	Massachusetts, USA	Nunc plasticware	
		Syringes	
		PBS (sterile)	10010023
		PBS (10x)	14200059
		DMEM "Low glucose, Pyruvate"	31885023
		RPMI 1640	21875158
		Penicillin/Streptomycin	15140122
		L-Glutamine	25030032
		Fetal Calf Serum (FCS)	10270106
		Trizol	15596026
		Amplex Red Cholesterol Assay Kit	A12216
		High-Capacity RNA-to-cDNA kit	4387406
		Cell proliferation Dye eFluor 450	(eBioscience) 65-0842-85
		Purelink HiPure Plasmid Filter Maxiprep Kit	K210016
		FAST SYBR Green Master Mix	4385612

VWR International Ltd	Lutterworth, UK	Ethanol	PROL20821.330

2.1.1.2 Other Materials/Supplies

Miscellaneous liquid chromatography-mass spectrometry (LC-MS) supplies (e.g. HPLC-grade acetylnitrate and some chemical standards) were kindly provided by the Metabolomics Unit in the Beatson Institute, Glasgow. Hexane, potassium hydroxide and pyridine were kindly provided by the Kamphorst group, Beatson Institute, Glasgow. All plastics and chemicals used for murine work (e.g. isofluorane, depilatory creams, 10% Neutral buffered formalin (NBF)) was provided by the Biological Service Unit in the Beatson Institute, Glasgow. Chloroquine and HEK293 cells, scramble-GFP, VSV-G and PsPax2 plasmids were kindly provided by the Helgason Group, Institute of Cancer Sciences, University of Glasgow. Agarose, TAE buffer and ethilium bromide were kindly provided by the Graham group, Institute of Infection, Immunity and Inflammation, University of Glasgow. Reagents and histology stains were provided by the Glasgow University School of Veterinary Medicine Pathology Department.

2.1.2 Human Tissues

All human tissues obtained had written consent in accordance with Declaration of Helsinki, and were used in protocols approved by the West of Scotland Research Ethics Committee (WoSREC: 09/S0703/77)

2.1.2.1 Primary cells/Primagraft cells

Primary bone marrow-derived human acute lymphoblastic leukaemia cells were obtained from the Bloodwise (formerly Leukaemia and Lymphoma Research) Childhood Leukaemia Cell Bank (York, UK).

2.1.2.2 CSF biobank

Leftover CSF taken from children at the Royal Hospital for Children, Glasgow (formerly Royal Hospital for Sick Children, Yorkhill Glasgow), approximately 1ml, was centrifuged at 1,000g for 10 minutes at 4°C, and carefully transferred into fresh microcentrifuge tubes and frozen at -80°C. CSF processing was largely carried out by Dr Saeeda Bhatti and Dr Yasar Yousafzai.

2.1.2.3 NSG Mice

3 Male and 3 female NOD.Cg-Prkdcscid Il2rgtm1WjI/SzJ (NSG) mice were obtained from Charles River Europe, and a breeding colony maintained in sterile, air-filtered, barrier cages at the Biological Services Unit, Beatson Institute, Glasgow. All *in vivo* experiments were carried out using UK Home Office approved protocols and in accordance with local procedures (project licence number 60-4512).

2.2 Techniques:

2.2.1 Patient Characteristics

2.2.1.1 Patient characteristics of cell donors for primary/primagraft xenotransplantion

Table 2-2 Patient characteristics for donors of primary and primagraft cells engrafted into NSG mice

Cell ID	Translocation	Sex	Age at diagnosis	White Cell Count at diagnosis	Outcome
Patient 1 (primagraft cell)	TEL:AML1 t(12;21)	F	2.5	84.6	Long-term survival; no relapse
Patient 2 (primagraft cell)	TEL:AML1 t(12;21)	F	3.6	120	Long-term survival; no relapse
Patient 3 (primagraft cell)	TEL:AML1 t(12;21)	F	3.5	113	Died on treatment; no relapse
Patient 4 (primary cell)	MLL t(4;11)	F	11.1	324	Long-term survival; no relapse
Patient 5 (primary cell)	MLL t(4;11)	F	3.2	489	Long-term survival; no relapse
Patient 6 (primary cell)	MLL t(4;11)	5	8.1	585	Long-term survival; no relapse

2.2.1.2 Patient characteristics for untargeted CSF analysis

Table 2-3 Patient characteristics for samples used in untargeted CSF analysis

ID	Age (diagnosis)	Sex	CNS status	WCC (x10⁹/L)	Genetics	Day 28 MRD	Medications at time of first lumbar puncture
1	3.3	M					Allopurinol, Amoxicillin, Azithromycin
2	10.3	M	1	1.8		High risk	Allopurinol, Piperacillin-Tazobactam, Gentamicin, Flecainide, Atenolol
3	3.8	M	1	10.6	(12,21)	Low	Allopurinol
4	4.9	M	1	52	(12,21)	Low	Allopurinol
5	4.7	M	1	6.1	(12,21)	Low	Allopurinol, Piperacillin-Tazobactam, Gentamicin
6	3.5	M	2	46	(12,21)	Low	Allopurinol, Piperacillin-Tazobactam, Gentamicin
7	4.7	M					
8	2.9	F	1	7.8	High hyprediploid	Risk	Allopurinol, Piperacillin-Tazobactam, Gentamicin
9	5.4	F	1	28	Normal	Low	Allopurinol, Piperacillin-Tazobactam, Gentamicin
10	6	F	1	1.1		Low	Allopurinol, Piperacillin-Tazobactam, Gentamicin
11	0.4	M	1	1.1		Risk	Allopurinol, Piperacillin-Tazobactam, Gentamicin
12	11.5	M	1	19.2		Risk	Allopurinol, Gentamicin
13	2.6	F	1	6.2		Low	Allopurinol, Piperacillin-Tazobactam, Gentamicin
14	15.8	F					
15	3.3	M	1	1.7	High hyperdiploid	Risk	Allopurinol, Piperacillin-Tazobactam, Gentamicin
16	3.8	F					
17	3.5	M	1	8.1		Low risk	Allopurinol, Piperacillin-Tazobactam, Gentamicin
18	8.3	M	1	16.3		Risk	
19	10.8	F	TLP	13.6	High hyperdiploid		
20	15.2	M	1			Risk	

2.2.1.3 Patient characteristics for CSF analysis at CNS relapse

Table 2-4 Patient characteristics for samples used in CNS relapse CSF analysis

ID	Age (diagnosis)	Sex	Genetics	WCC ($x10^9/L$)	CNS status	Induction Regimen	Day 28 MRD result	Time to 1st CNS relapse	Other details
1437	4y	F	t(1;19)	142.9	CNS 1	B	low risk	1 year	
6539	23m	M	unknown	2.9	CNS 3	C	unknown	18m	Bone marrow relapse 2 years after diagnosis
4234	8y	M	T14 + T21	7.8	CNS 1	A	low risk	4y	SCT (with cranial irradiation) 4.5 years post diagnosis; CSF analysed for CNS relapse 2, at 2 years post SCT and CNS relapse 3 at 4 years post-SCT
6941	3.5y	M	t(12;21)	10.4	CNS 1	A	low risk	3y	CNS relapse 2 at 5yrs post-diagnosis
046	2.4	M	Unknown	11.1	CNS 1	A	low risk		
0171	10.4	M	High hyperdiploid	1.1	CNS 1	B	Risk		
186	8.3	M	t(12;21)	16.3	CNS 1	A	Risk		
0129	3.3	M	High hyperdiploid	1.7	CNS 1	A	Risk		
195	3.5	M	Normal	8.1	CNS 1	A	Low risk		
135	11.1	F	t(7;9)	23.8	CNS 1	C	Risk		
0233	4.6	F	t(12;21)	62.5	CNS 1	B	Risk		
0081	2.5	F	t(12;21)	19.1	CNS 1	A	Risk		
0594	9.7	F	Normal	131	CNS 1	B	Risk		

2.2.2 Xenografting

Human ALL cell lines, primary cells, and primagraft cells (i.e. primary cells that had been xenografted, then human ALL cells retrieved from the spleens of mice post-engraftment) were introduced into immunocompromised NOD.Cg-Prkdcscid Il2rgtm1WjI/SzJ (NSG) mice.

2.2.2.1 Tail vein injection

SEM and REH ALL cell lines, in passage 5-10 from purchase from ATCC, were centrifuged at 300g for 5 minutes, supernatant discarded and cells resuspended in “complete” culture media (i.e. RPMI or DMEM with 10% FCS, 5% Penicillin, 5% Glutamine) at a density of $2x10^7$ cells/ml. This cell suspension was stored at 4°C and transferred to the animal facility where 0.1ml ($2x10^6$ cells) were introduced intravenously into the tail vein of 6-12 week old NSG mice with a 29-gauge needle under aseptic technique. These cells reliably engrafted in CNS and spleen by 28 days (SEM) or 35 days (REH).

At the end of experiments – either at 28 days (SEM) or 35 days (REH) or at the development of symptoms (e.g. weight loss), mice were culled by intraperitoneal injection of lethal dose of pentobarbital for experiments investigating the transcriptome or metabolome of CNS ALL cells or CSF. For any other experiments, mice were culled using increasing $FiCO_2$.

2.2.2.2 Intrafemoral injection

Primary and primagraft cells were thawed, resuspended in 10ml complete RPMI, centrifuged at 300g for 5 minutes and resuspended in complete RPMI to a cell density between $1x10^6$ and $1x10^7$ cells/ml at 4°C.

Primary and primagraft cells engraft less reliably than cell lines, likely due to reduced cell numbers and reduced cell viability. To combat this, these cells were introduced directly into the femurs of mice. Mice hindlimbs were treated with depilatory cream to remove hairs. 24 hours later, mice were anaesthetised with isofluorane, their femurs identified and immobilised manually, and an aperture made in the distal epiphysis with a 25-gauge needle. This needle was removed and 0.1ml of the ALL cell suspension introduced into the marrow space with a 29-gauge needle under sterile technique.

Mice received carprofen analgesia intraoperatively dosed according to their weight following local protocols, and were closely monitored for 48 hours post-procedure for signs of discomfort. Any mice with discomfort at 24 hours received a further dose of carprofen. No mice required carprofen beyond 24 hours.

Mice were culled either at the development of symptoms (e.g. weight loss or hindlimb paralysis), or at 6-9 months if no symptoms developed.

2.2.2.3 Tail vein injection of 13C-labelled glucose and acetate

1mg/g bodyweight in 100µL sterile water of fully 13C-labelled glucose, or 100µL of 300mM fully 13C-labelled acetate, was injected via the tail vein of mice, and mice culled by pentobarbital and CSF/cells retrieved as described below after 20 minutes or 40 minutes.

2.2.2.4 Retrieval of human ALL cells for RNA analysis

To retrieve cells for RNA analysis, mice were culled with intraperitoneal pentobarbital. Brains were carefully and rapidly removed and placed in a 2% FCS/PBS solution at 4°C, the meninges adherent to the skull were carefully and gently scraped with the rubber plunger of a 1ml syringe and the scrapings suspended in the same FCS/PBS solution as the brain. This solution was gently vortexed for 1 minute to dislodge cells from the meningeal layers adherent to the brain, and a meningeal cell suspension created by passing the solution through a 40µm filter. The brain and any connective tissues in the filter were discarded and the cell suspension carefully layered onto lymphoprep™ (StemCell Tech.) in 15ml centrifuge tube as per the manufacturer's instructions.

Spleens were rapidly removed, weighed and placed in a different 2% FCS/PBS solution. Spleens were then processed through a 40µm filter to create a cell suspension. This cell suspension was carefully layered onto lymphoprep™ in a 50ml centrifuge tube as per the manufacturer's instructions. ALL cells from the spleen were used as there is good evidence that the transcriptional profile of xenograft ALL cells from the spleen is very similar to that of the bone marrow(Samuels et al. 2010), and leukaemic cells in the spleen are more abundant and technically easier to harvest and isolate

Specimens on lymphoprep™ were centrifuged at 800g for 20 minutes to separate cell populations by density (i.e. to remove any erythrocytes, glia or granulocytes from cell suspensions). Purified lymphocytes were resuspended in 10ml 2%FCS/PBS at 4°C, supernatants discarded and cell pellets snap-frozen in dry ice and stored at -80°C. Time from cull to freezing was approximately 45 mins to 1 hour, with cells mainly kept at 4°C during this time. (Figure 2-1)

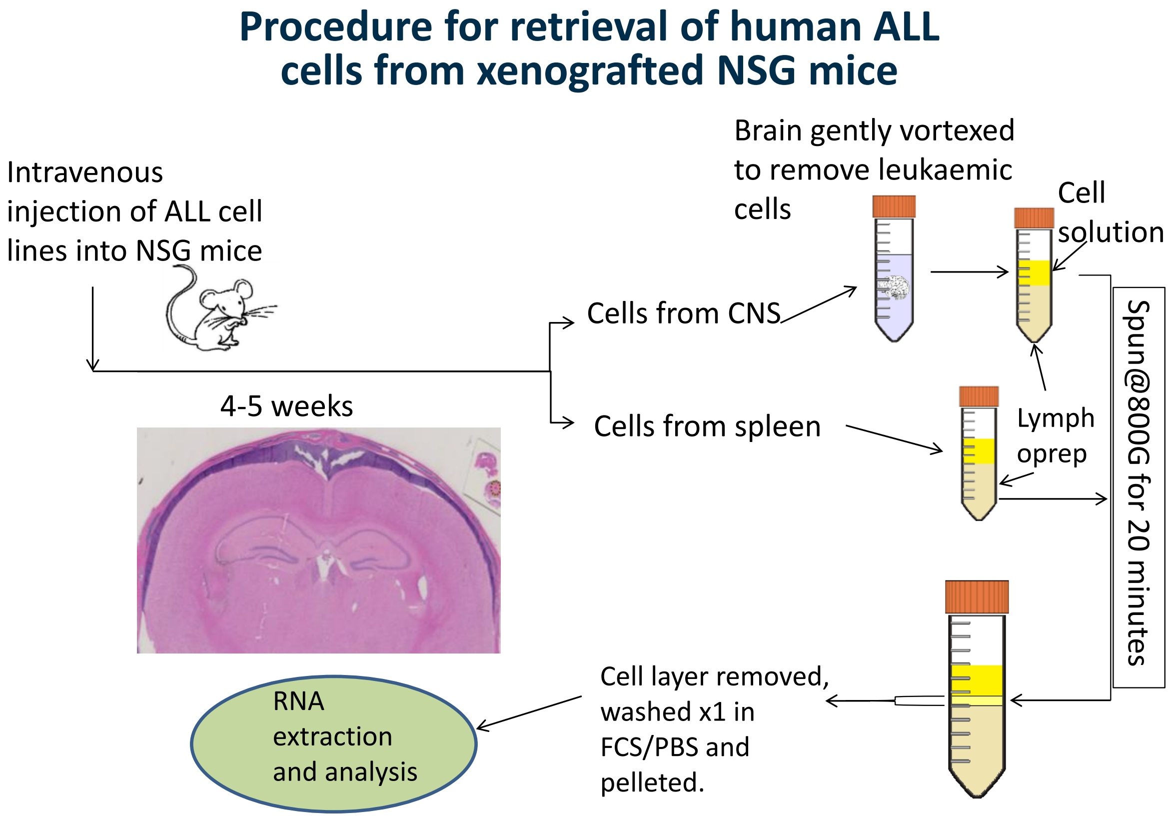


Figure 2-1 Schematic diagram of procedure for extraction of RNA from CNS and spleen ALL in mice

2.2.2.5 Retrieval of human ALL cells for metabolomic analysis

For metabolomic analysis time from cull to freezing is crucial as the metabolome can be disrupted very rapidly. In addition, plasma and CSF samples can provide extremely interesting data either on their own or as a supplement to cellular data. The technique for retrieving cells was therefore adapted to allow for maximum speed and accuracy of collection.

The mice were culled by lethal dose of phenobarbital and CSF was obtained when the mice were unconscious under terminal anaesthesia. Immediately as death was confirmed, the mice were placed on ice, a laparotomy performed, and blood drawn from the IVC or heart. Blood and CSF were both immediately stored at 4°C, and later centrifuged at 2000g for 15 minutes to make blood plasma, and supernatants removed and stored at -80°C. The mice were decapitated, the base of skull and brain removed, and meningeal cells scraped gently with a syringe plunger through a 40μm filter into 2% FCS/PBS at 4°C. The spleen was removed and processed through a 40μm filter into 2% FCS/PBS.

As discussed below, there were concerns that red cell contamination from the spleen would compromise any results, and an experiment to examine the utility of red cell depletion is described (section 5.5.1). Rapid red cell depletion for metabolomic analysis was performed by diluting Red Cell Lysis Solution 10x (Miltenyi) in ddH_2O as per manufacturer's instructions, and incubating cells in this solution at room temperature for 3 minutes. For later experiments, a 50μL aliquot each of spleen-derived and meningeal-derived ALL cell suspensions were taken and stored at 4°C for counting.

The cell solutions were then centrifuged at 4000g for 2 minutes, supernatants discarded, and snap-frozen in dry ice before storage at -80°C.

2.2.2.6 Retrieval of tissues for histology

Finally, for xenograft experiments where RNA or metabolite extraction was not required, mice were culled with increasing $FiCO_2$. Spleens were removed and weighed, then placed in 10% NBF and transferred to a specialist laboratory (Glasgow University School of Veterinary Medicine Pathology Department), fixed for over 48 hours. Mice heads were removed, stripped of soft tissues exterior to the skull and placed in 10% NBF for immediate storage and transport. On arrival at the specialist laboratory, these samples were decalcified in a Hilleman and Lee 5.5% EDTA in 10% formalin solution for 2-3 weeks, then fresh EDTA for 3-4 days prior to embedding in paraffin. Paraffin blocks were then cut into 2.5μm slides, dried, and stained with Gill's haemotoxalin and Putt's eosin prior to transfer back to the laboratory. Slides were scanned with a Hamamatsu NanoZoomer NDP scanner, and analysed using HALO v2.0.1061.3 software (Indica Labs Inc.).

2.2.3 RNA extraction

Two different techniques were used for RNA extraction. For RNAseq analysis, the University of Glasgow Glasgow Polyomics Department's protocol stipulated the use of Trizol™ for RNA extraction using the standard manufacturer's procedure. For other RNA experiments, RNA was obtained using the Qiagen RNeasy™ system which was simpler and safer.

2.2.3.1 Trizol™ extraction

To extract RNA using Trizol™, in a clean hood 1 ml of Trizol™ was added to cell pellets, and cell pellets resuspended using a pipette, and the mixture transferred to a "clip-lock" microcentrifuge tube. 0.2ml chloroform was added and mixed by inversion of the samples. After 3 minutes at room temperature, the mixture was centrifuged at 10,000g for 20 minutes at 4°C. After centrifugation the aqueous phase was removed from the resultant mixture and diluted in an equal volume 100% isopropanol (approx. 0.5ml), and mixed by inversion. After 10 minutes at room temperature, the mixture was centrifuged at 10,000g for 10 minutes at 4°C and the supernatant discarded. 1ml of 75% ethanol was added to the tube and mixed by pipetting, then centrifuged at 7,500g for 5 minutes at 4°C. The supernatant was discarded and the tube air-dried (approximately 45 minutes) then resuspended in 50μL nuclease-free water. This was incubated on ice for 10 minutes, then at 65°C for 5 minutes. The quantity and quality of RNA obtained was assessed using the Nanodrop spectrophotometer (ThermoScientific, MA, USA) to assess absorbance at 260 and 280nm wavelengths of light.

2.2.3.2 RNEasy

The RNeasy™ system was bought in as a kit, either the "micro", "mini", or "midi" kit depending on the cell number RNA was to be extracted from. The manufacturer's instructions were followed, but briefly:

- Cells were lysed using "Buffer RLT" (from the RNeasy™ kit), and homogenised by centrifuging through "QIAshredder" column.

- The homogenate was mixed with alcohol, and the RNA eluted onto a membrane in a "Spin Column" supplied with the kit by centrifugation.

- The column was washed with a proprietary buffer, and DNAse added to the membrane with RNA to remove DNA. This was incubated for 15 minutes, then washed again.

- The RNA wass retrieved in a microcentrifuge tube by centrifuging RNA/DNA-free water through the RNA-impregnated membrane.

- The quantity and quality of RNA obtained was assessed using the Nanodrop spectrophotometer as above.

2.2.4 RNAseq

RNASeq is a technique used to identify and quantify RNA using massive parallel sequencing of cDNA(Wang et al. 2009). RNA was obtained from SEM and REH cells retrieved from the CNS and spleens of mice as detailed above (sections 2.2.2.4 and 2.2.3.1). This was then passed to the University of Glasgow Glasgow Polyomics Department. The RNA was analysed using the NextSeq® 500 platform as (Illumina® Inc., CA) with poly-A mRNA selection and 75bp paired-end sequencing. Reads were aligned to the reference human and murine transcriptome using k-mer statistics with the Kallisto programme (Patcher Lab, Caltech, CA). Any reads aligning to the mouse genome were removed (100,000 of 30 $x10^6$ reads from control cells; 1,500,000 of 30 $x10^6$ reads from ALL cells retrieved from the spleen; and 800,000 of 30 $x10^6$ reads from ALL cells retrieved from the CNS). In total, RNA from 65,000 genes was analysed by the lab for differential expression using the DESeq2 package as discussed below (section 2.2.12.1).

2.2.5 Polymerase Chain Reaction (PCR)

PCR is a technique used to amplify a small sequence of DNA using two complementary-DNA primers (i.e. small molecules of DNA with sequences of bases that will bind one to the start and one to the end of a target sequence) and DNA polymerase. As the amplification of DNA is an exponential processes given sufficient resources, the time taken to reach a detectable amount of DNA can be used in a technique called quantitative PCR (qPCR) to calculate the original amount of the target DNA in a sample.

In this project, RNA from cells was converted to cDNA (i.e. each RNA molecule was translated into the complementary DNA sequence) using High-Capacity RNA-to-cDNA™ Kit (ThermoFisher Scientific), method below. This cDNA was then analysed using PCR to determine the original abundance of specific molecules of RNA.

2.2.5.1 Conversion of RNA to cDNA

Using a High-Capacity RNA-to-cDNA™ Kit (ThermoFisher Scientific), equal amounts of RNA for each sample in an experiment up to 2μg in a volume of nuclease-free water up to 9μL (as determined by nanodrop spectrophotometer absorbance) were mixed with a reaction buffer (including excess nucleotides) and the reverse transcriptase (RT) enzyme. Additionally samples were mixed with reaction buffer and nuclease-free water to provide a negative control (negative-RT control). The samples were then incubated at 37°C for 60 minutes to produce cDNA, then the reaction terminated by heating to 95°C for 5 minutes. The cDNA was stored at -20°C until use.

2.2.5.2 Primer design

Primers were designed using Primer 3 Plus web tool (www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi, Wageningen University, NZ), and analysed for target specificity using Primer-BLAST (National Center for Biotechnology Information (NCBI), US) to minimise the chance of off-target DNA/cDNA binding to sequences in the human or mouse genome.

In brief, the target DNA sequence (i.e. the sequence of DNA translated to RNA after splicing etc.) was loaded into Primer 3 Plus, and which was analysed for potential sequences with the following parameters:

- length 18-23 base-pairs
- 40%-65% of bases being guanine or cytosine
- predicted Tm ("melting temperature" – the temperature of maximal complementary DNA binding) 59.5°C-61°C

- Maximum self-complementarity of 2 (i.e. the primers will not have sequences larger than 2 base-pairs which are complementary to themselves or each other)

- Maximum 3’ self-complementarity of 1 (i.e. a complete lack of self-complementarity in the primer binding to the start of the sequence, where the polymerase reaction will start)

- Target sequence of less than 150 base-pairs

- Avoiding stretches of 4 or more G or C bases

- Avoiding a GC “clamp” on primers (i.e. avoiding 2 or more G or C bases within 5 bases of the 3’ terminus of each primer)

The primer pair with the characteristics closest to this was selected, and analysed using the primer-BLAST tool comparing the sequences to the human and mouse genome to ensure there was no ,minimal-target or cross-species complementarity, and to cover major splicing isotypes of the genes.

2.2.5.3 Standard PCR

2 sets of primers were designed for HMG CoA Reductase (HMGCR), HMG CoA Synthase (HMGCS), and TATA-box Binding Protein (TPB) (“housekeeping”) genes using the method described above. One set the “outer” set, amplified a sequence of the target gene containing the sequence amplified by the “inner” set of primers. Primers were obtained from Sigma-Aldrich.

Table 2-5 List of PCR primers for qPCR (HMGCR / HMGCS1 / TBP Housekeeping gene)

Target	**Sequence (forward/reverse)**
HMGCR Standard	GGTGATGGGAGCTTGTTGTG / AGTGCTGTCAAATGCCTCCT
HMGCR QPCR	CAATGGCAACAACAGAAGGT / GTGGAAGACGCACAACTGG

HMGCS1 Standard	AGGGTGGATGAAAAGCACAG / GGGCTTGGAATATGCTCAGTT
HMGCS1 QPCR	ATGGTTCCCTTGCATCTGTT / CTTCAGGTTCTGCTGCTGTG
TBP Standard	AGGATAAGAGAGCCACGAACC / GCTGGAAAACCCAACTTCTG
TBP QPCR	GGGCACCACTCCACTGTATC / CATCTTCTCACAACACCACCA

RNA was extracted from HEK293 cells, and converted to cDNA as above. A conventional polymerase chain reaction was then carried out for each primer pair using Phusion DNA Polymerase as per manufacturer's instructions. Briefly: cDNA was mixed with a reaction buffer, an excess of nucleotides, a primer pair and DNA polymerase. This mixture was placed in a thermal cycler for an initial 3 minutes of 98°C, then 30 cycles of 98°C for 5 seconds, 60°C for 20 seconds, 72°C for 15 seconds, a final 72°C for 5 minutes, then the products stored at 4°C.

A 2% agarose gel was prepared with TAE buffer and ethidium bromide, and the PCR products and a DNA ladder were run through the gel by electrophoresis at 100V for 20 minutes. UV light in a gel imager was used to visualise bands of oligonucleotides, confirming the presence of a single band of appropriate size for the target gene.

For the "outer" set of genes, bands were carefully cut from the gel, and DNA purified using a QIAquick kit (Qiagen) as per instructions (using a similar membrane-binding system to the RNeasy system). This DNA was quantified using Nanodrop spectrophotometer (with number of copies of the target sequence determined by calculating the mass of each copy of target sequence DNA based on length and the average mass of a DNA base-pair – 660da), and then a series of dilutions made to provide a standard for quantification of cDNA in samples for each of the target genes.

Quantification was done using quantitative PCR (qPCR) using a SYBR-green system. SYBR green is a dye which binds to DNA, and in its bound state can absorb blue light, and emit green light. The fluorescence of a mixture with an excess of SYBR green with DNA therefore depends on the concentration of DNA. This can be used to continually measure the DNA abundance during PCR reactions.

A master mix for each of pairs of “inner” primers together with SYBR-green (ThermoFisher Scientific)) and nuclease-free water was added to each sample in triplicate and to serial dilutions (each in duplicate) of the appropriate standard from the step above in a range from 10^8 copies to10^2 copies in dilutions of 10. Additionally the master mix was added to a “-RT” control for each gene for each sample.

PCR was then performed using a 7900HT real-time PCR machine (ThermoFisher Scientific). The mixture was heated to 94°C for 10 minutes, then 40 cycles of 94°C for 3 seconds and 60°C for 30 seconds, with fluorescence of the mixture measured at the end of each cycle. Finally, the mix was heated incrementally to 94°C over 25 minutes, with fluorescence continually measured for a melt curve. The period of exponential increase in DNA for each primer is automatically detected by the software, and an arbitrary threshold set. The time taken for each sample to reach this threshold (Ct) was determined and used to calculate the abundance of the target cDNA in the initial sample. A standard curve was calculated for the Ct’s of the standards, and the amount of each gene in the samples calculated from this. The abundance of each gene was normalised to that of the housekeeper gene (TBP).

2.2.5.4 Multiplex PCR

Multiplex PCR using the Fluidigm® Biomark HD system was used to facilitate the performance of multiple PCR reactions on single samples. The technique was theoretically very similar to previous techniques, but with some differences. Primers were prepared according the procedure above, and cDNA generated from each sample as above. The manufacturer’s instructions were followed, but in brief:

The abundance of each target cDNA sequence was amplified to improve sensitivity. A mix of each of the primers was prepared at 500nM and 0.5μL mixed with 2.5 μL TaqMan PreAmp Master Mix (ThermoFisher Scientific), 0.75μL nuclease-free water and added to 1.25μL of each sample. This mixture was placed in a thermal cycler at 95°C for 10 minutes, then 14 cycles of 95° for 15 seconds and 60°C for 4 minutes. Amplified samples were treated with Exonuclease I (New England Biolabs) – 0.4μL of 20 units/μL exonuclease I, 0.2μL reaction buffer and 1.4μL water were added to the 5μL samples, mixed and briefly centrifuged then heated in a thermal cycler to 37°C for 30 minutes, then 80°C for 15 minutes. The samples were then diluted in 18μL TE buffer and stored at -20°C.

Samples were mixed with SsoFast Supermix (BioRad) and DNA-Binding Dye Sample Loading Reagent (Fluidigm). Primer pairs were diluted to a concentration of 5μM in Assay Loading Reagent and DNA Suspension Buffer (Fluidigm). The samples and primers were then loaded onto a pre-primed 48.48 Dynamic Array IFC (Fluidigm), and the mixtures loaded using an IFC controller before PCR being performed and analysed using the BioMark HD.

Target	**Sequence (forward/reverse)**
HMGCR	TGTTTGCAGATGCTAGGTGTTC / AGTGACAATTCCCCAGCCATT
HMGCS1	GGTGTGCGCAGGACTTT / CAGGGCTTGGAATATGCTCAGTT
CYP51A1	ATTTGGAGCTGGGCGTCATC / TCAGTGGGAAAGTATCCATCA
DHCR7	GAGGCCAGGGGAAGGT / TCACAGCAGAACTTTT
DHCR24	CCCCGAGGTGTACGACAAG / TGAAAGTGTGGATCTAGGAAAGCA
FDFT1	GCCAACTCTATGGGCCTGTTT / AACATACCTGCTCCAAACCTCTT
LSS	GTACGAGCCCGGAACATTCT / CAGCCACATCTCTGGGAACA
MSMO1	ATTGGAATCGTGCTTTTGTGTGA / TGATGCCGAGAACCAGCATAG

MVK	GCTCTGGGTTGTGGGAGTTG / CGGCGCAGACACCAGYAGGA
SQLE	ATGGCAGAGCCCAATGCAAA / AACAGTCAGTGGAGCATGGA
TM7SF2	GGGGAACTCAGGCAATCC / CGGGTCGCAGTTCACAGAAA
PMVK	CCGCGTGTCTCACCCTTTTC / CTAAGCGGAGCGGCAACAAG
MVD	CTTTCCCCCAGGCTCGAATG / GGCCCCACCTGAGTGACAA
IDI1	ACCACCTCGACAAGCAACAG / AGCTCGATGCAATAATCCTTTCTCA
FDPS	TGACGGTGGTAGTAGCATTCC / CTTGCAGCAGTTCCACACA
NSDHL	CGCCTACGGACGGAAAAGAA / CGTGCGACTTGGTCTCTCA
SC5D	ACAAATGTTGGTGCTTACATCCT / TCTCTCGACGGACTTGATTCTTT
HSD17B7	AAGCAAAGGTTTCAGAGATTAGAC / CAGCTTGAAAAAGAGGCCAAAGA
EBP	CCGATACATCCTGGGTGACAA / GGCCCACAGAGACCACAAG
IDI2	GAGGAATTGCCATCTGAACGA / AATACCCAGGAAACGTGACT
SREBF1a	ATGGACGAGCCACCCTTCAG / GCCGACTTCACCTTCGATGT
SREBF1C	GCGCTCAACGGCTTCAAAAATC / GCATGTCTTCGAAAGTGCAATCC
SREBF2	GGACCTGAAGATCGAGGACTTT / TCATCTTTGACCTTTGCATCATC
SPP1	GAGGAGACACAGCCGAGATAC / GGCTTTTCCAAACAGAGTCACA
IRF4	ACCCGGAAATCCCGTACCAA / GGTGGGGCACAAGCATAAAAG
VEGFA	ACTGAGGAGTCCAACATCACCAT / CGGCTTGTCACATCTGCATTCAC
ACSS2	CCATTGTGTTTGCAGGCTTCT / CCAGCTCCTTCAGGTTCACA
IL-15	TGTTTCAGTGCAGGGCTTCC / CAACTGGGGTGAACATCACTTTCC
ACSS1	ACCTGGTGGCAGACAGAAACA / CAAAGAAGGGCCTCATCGCC

ICAM-1	TGGGCAGTCAACAGCTAAAAC / CCTGGCAGCGTAGGGTAAG
MERTK	AATCCCCCTCCGTGCTAACT / TGGGGAGGGAATTGCTTTGA
PBX-1	ATGACCATCACAGACCAGAGTT / AAAGAACAATGAGGGGCAGTT

2.2.6 Metabolomics

The analysis of large numbers of small-molecule metabolites from an integrated network in a biological system is known as metabolomics. Metabolomic analysis can be done in various ways, but the most common uses a combination of chromatography to separate metabolites in a sample over time, ionised to a charged state, and mass spectrometry to measure with extreme accuracy the mass (or more accurately the mass/charge ratio - m/z) of those metabolites by measuring the deviation of the ionised metabolite from a straight path under a known voltage. With this combination of techniques it is possible to detect many thousands of possible metabolites. In this project 3 complementary approaches to these techniques were used.

In the first, known in this thesis as “targeted” liquid-chromatography mass spectrometry (LC-MS), samples are separated on a liquid chromatography column then ionised and the masses of metabolites measured using mass spectrometry. Metabolites were identified by the chromatography time (known as retention time or RT), and sometimes the shape of the chromatography peak, and their mass by comparison to a regularly maintained database of pure metabolite standards’ RT and mass. This technique is relatively quick and provides excellent resolution of metabolite peaks.

The second technique, “untargeted” LC-MS is similar in that a liquid chromatography column is used to separate metabolites, though metabolites are diffused over a longer time to improve separation, and ionised only to either positive or negative polarity on a single run prior to detection. As an additional step, the top 10 highest abundance metabolites at any time are fragmented, and the m/z ratios of these fragments measured. Metabolites are identified based on retention time, mass and the fragmentation pattern. This technique allows the identification of a huge number of metabolites, not only those that have previously been run as a standard.

Finally a third technique, gas-chromatography mass spectrometry (GC-MS) was used. This vaporises samples and separates metabolites using gas-chromatography prior to mass spectrometry. This technique allows better detection and quantification of volatile and lipid molecules.

To prepare samples for analysis, the metabolites first need to be prepared or “extracted.”

2.2.6.1 CSF and plasma extraction

At 4°C 50μL of human CSF samples were mixed 1 in 20 in a extraction solution (50% methanol/30% acetonitrile/20% deinoised water) with a final concentration of 0.5μM ^{13}C-labelled (i.e. standard molecules with normal ^{12}C carbon atoms replaced with detectably heavier but chemically identical ^{13}C carbons) pyruvate, arginine, alanine, lactate and 5μM ^{13}C Glucose.

Samples were mixed thoroughly in a vortex for 30 seconds then centrifuged at 16,000g for 10 minutes at 4°C, then the supernatant transferred to a glass vial and stored at -80°C until analysis.

Mouse CSF and plasma samples were extracted using the same method, though 1μL of sample was mixed 1 in 50 in extraction solution.

Extraction using methanol/acetonitrile/water was used for analysis of small polar metabolites. For analysis of more lipid molecules two different sample extraction methods were used. For LC-MS lipid analysis, 100 μL of CSF was placed in a microcentrifuge tube at 4°C then 300μL of chloroform and 600μL of a 50:50 methanol/deionised water mixture added. Samples were mixed in a vortex for 2 minutes then the bottom layer (i.e. the chloroform layer) carefully pipetted into a glass vial. This vial was then placed under nitrogen gas until dry then stored at -80°C.

For GC-MS, samples were processed as above, then once dry under N_2 gas the samples were resuspended in 700μL 50:50 methanol:PBS and 75μL potassium hydroxide, then incubated at 80°C for 1 hour in order to hydrolyse lipids. After incubation 200μL water and 500μL of hexane was added. The samples were vortexed for 3 minutes and the top layer (hexane) carefully collected. This layer was then dried under N_2 gas. Once dried, the samples were resuspended in 40μL of dry pyridine (stored with molecular sieves to ensure no water contamination) and 50μL MSTFA to derivate cholesterol. Samples were added to glass vials, capped and incubated at 60°C for 60 minutes, cooled to room temperature and analysed immediately.

2.2.6.2 Cellular extraction

For extraction of cellular samples, cells in microcentrifuge tubes were retrieved from storage at -80°C, and extraction solution (50% methanol/ 30% acetonitrile / 20% water as above) added to a cell concentration of 1-2 $x10^6$ cells/ml. Samples were mixed by pipetting to resuspend cells then agitated in a thermomixer at 1,400 rpm for 10 minutes at 4°C. Samples were then centrifuged at 16,000g for 10 minutes at 4°C, the supernatants transferred to glass vials and stored at -80°C until analysis.

2.2.6.3 Targeted LC-MS

For small molecule analysis, samples were loaded onto a Thermo Ultimate 3000 high-performance liquid chromatography (HPLC) system (ThermoFisher). 2μL of sample was loaded and run through a 150 x 2.1mm SeQuant ZIC-pHILIC column (Merck), with a preceding SeQuant 20 x 2.1mm guard column (Merck) with an aqueous and an organic solvent. The aqueous solvent used was 20mM ammonium carbonate, and the organic solvent was acetonitrile. A gradient from 80% organic solvent/20% aqueous to 80% aqueous solvent/20% organic was run over 15 minutes, then returned to starting conditions over 7 minutes, all at 45°C. The total run time was 23 minutes.

After separations by chromatography, samples were ionised using heated electrospray ionisation (HESI) and injected into a Q-Exactive Orbitrap mass spectrometer, where ions were scanned in a range 75-1000 (m/z) with a resolution of 35,000. Polarity switching was used to produce and analyse negative and positive ions. Lock masses were used, and mass error for the spectrometer was less than 5ppm, and usually less than 2ppm. Data were acquired using XCalibur software (ThermoFisher), and analysed using TraceFinder v4.1 (ThermoFisher), and metabolites were identified by the mass of singly-charged ions and by known retention time(Mackay et al. 2015). All metabolites detected had previously been analysed using commercial standards. All LC-MS was carried out by the Metabolomics Unit in the Beatson Institute, Glasgow.

Lipidomic analysis was also performed using a Q Exactive Orbitrap mass spectrometer. 4µL of samples were injected as above, and separated on a Waters Acquity CSH C18 column (Waters) at 50°C. A gradient of aqueous solvent water/acetonitrile (40:60, v/v) with 10 mM ammonium formate and organic solvent acetonitrile/2-propanol (10:90, v/v) with 10 mM ammonium formate, at a flow rate of 0.3 mL/min was used. The solvent gradient ran from 0% to 40 % organic over 6 min, then from 40% to 100% in the next 24 min, followed by 100% organic for 4 min, and then returned to 0% in 2 min where it was kept for 4 min column equilibration (40 min total). Samples were ionised using HESI. The mass spectrometry analysis was performed with a full scan range of 300-1200 m/z, with a resolution of 17,500, using MS and ddMS(top 3) analysis. Polarity switching was used to produce and analyse negative and positive ions. Lipidomic data analysis was performed using LipidSearch 4.0 software (Thermo Scientific). All lipidomic LC-MS and initial LC-MS data analyses were carried out by S Tumarov, Beatson Institute, Glasgow.

2.2.6.4 Untargeted

Untargeted mass spectrometry samples were run as from small molecule analysis as above, but with a chromatography time of 38 minutes (with a flow rate of 100µL/min, and a solvent gradient (solvents as above) changing from initially 80% organic to 20% organic over 30 minutes, then back to 80% organic over 8 minutes). There was no polarity switching. Metabolite fragmentation by ddMS2 of the top 10 metabolites at any one time was carried out in the Q-Exactive HCD collision cell. Peaks were analysed using Progenesis QI software, and using publicly-available databases (the Human Metabolome Database (HMDB)(Wishart et al. 2007), the METLIN database(Smith et al. 2005) and the MZCloud database (https://www.mzcloud.org)). There was limited automatic checking of the HMDB and METLIN databases with the Progenesis software, however the bulk of compound identification was done manually. The MZCloud database was found to most faithfully identify prospective compounds from these experiments. Compound masses and fragmentation peak m/z's were used to manually search the database and potential matches identified. These were confirmed by analysis of external standards. All LC-MS was carried out by the Metabolomics Unit in the Beatson Institute, Glasgow.

2.2.6.5 Gas Chromatography-Mass Spectrometry (GC-MS)

Samples were separated using gas chromatography on an Agilent 7890B system using a Phoenix ZB-1701 column then analysed using Agilent 7000 Triple Quadripole GC-MS system with a scanning range of 38-650Da. Gas flow through the column was 1ml of He/minute. 1 µL of sample was injected in splitless mode, with an inlet temperature of 280°C, interface temperature of 270°C and a quadrupole temperature of 200°C. Quadrupole temperature was kept constant for 1 minute then increased by 20°C/minute to 280°C, and was kept at this temperature for 9 minutes. Total run time was 14 minutes. The column was equilibrated for 2 minutes between samples.

Data were acquired between 5 and 14 minutes, in a mass range of 140-550Da at 1.47 scans/second. The internal standard d_7-cholesterol was used to allow absolute quantification of cholesterol. Hunter B.06.00 software (Agilent, CA, US) and MetabQ (R package) were used for data analysis. All GC-MS and initial GC-MS data analyses were carried out by S Tumarov, Beatson Institute, Glasgow, unless otherwise stated.

2.2.6.6 Amplex® Red Cholesterol Assay

An alternative approach to measure cholesterol using Amplex® Red Cholesterol Assay Kit (ThermoFisher) was also used as per the manufacturer's instructions. This technique uses hydrogen peroxide produced by the oxidation of cholesterol to react with Amplex® Red reagent and cause fluorescence proportional to the abundance of cholesterol in the sample.

In brief: samples and a range of cholesterol standards were added to a proprietary reaction buffer. Negative controls with only reaction buffer, and positive controls with 10µM hydrogen peroxide were also prepared. A mixture of cholesterol esterase, cholesterol oxidase, horseradish peroxide (HRP) and Amplex® Red reagent was then added to the solution and incubated at 37°C for 30 minutes protected from light, then fluorescence at 590nm measured. To measure free (rather than total) cholesterol, cholesterol esterase is omitted from the reaction solution and replaced with reaction buffer.

2.2.7 *In vitro* culture

2.2.7.1 Maintenance/Passage

Cells were incubated at 37°C in 5% CO_2. All manipulation of tissue culture cells was done using sterile technique in a sterile TC hood.

SEM and REH cells were maintained in “complete” RPMI or DMEM media at a concentration of 0.5 – 2 $x10^6$ cells/ml as per the supplier’s recommendations. DMEM was used to maintain these cell lines in lieu of RPMI to facilitate co-culture experiments. Cells were regularly tested for the presence of mycoplasma using Mycoalert mycoplasma detection kit (Lonza) as per the manufacturer’s instructions. For passage, cells in media were pipetted into a 15ml centrifuge tube, then centrifuged at 300g for 5 minutes, supernatants discarded, and cells resuspended in complete media to a concentration of $5x10^6$/ml.

HS5 cells were maintained in “complete” DMEM with ES-FCS at a confluence of 60-90%. For passage, media was removed and stored, and cells were washed in sterile PBS. PBS and 0.25% Trypsin/EDTA (ScienCell) was added to the cells. For a 75 cm^2 flask, 8ml PBS and 2ml Trypsin/EDTA were used, and equivalent amounts for other sizes of flask. The flasks were monitored for cell detachment using an inverted microscope. Once cells had detached (3-4 minutes), the Trypsin-cell solution was neutralised in complete media in a 50ml centrifuge tube, and the flasks were incubated a further 2 minutes then washed twice in complete media to collect any residual cells. Cells were centrifuged at 300g for 5 minutes, the supernatant discarded, and cells resuspended in complete ES-FCS/DMEM prior to seeding at a confluence of 60%.

Primary Human Meningeal cells (HMens) were maintained in proprietary Meningeal Cell Media (MCM), with 2% FCS, 1% Meningeal Cell Growth Supplement and 1% Pen/Strep (Sciencell) at a confluence of 60%-90%. Passage was carried out as per HS5 cells, except trypsin was neutralised and flasks rinsed with trypsin neutralising solution (Sciencell), and cells were resuspended into a flask coated in poly-L-lysine (Sciencell) as per supplier's instructions.

2.2.7.2 Freeze/thaw

To freeze cells, cells were prepared as above (i.e. SEM and REH cells were pipetted directly into a centrifuge tube, and HS5 and HMen cells detached using trypsin prior to pipetting into a centrifuge tube, then centrifuged at 300g for 5 minutes), then resuspended at a density of $1x10^7$ cells/ml (SEM and REH), or $1x10^6$/ml (HS5 and HMen) in 1ml of 70% media (the same media used for maintenance of culture as above), 20% FCS and 10% DMSO in a cryostorage tube. Cells were then frozen at -196°C in the gaseous phase of a liquid nitrogen tank.

To thaw SEM and REH cells, cells were warmed in a water bath at 37°C, then mixed in 10ml of complete DMEM at 37°C, centrifuged at 300g, then resuspended in complete DMEM at a density of $1x10^6$/ml. To thaw HS5 and HMen cells, cells were thawed at 37°C in a water bath, then mixed with complete ES-FCS/DMEM or MCM to a density of 25,000 cells/ml and seeded in 12ml in a $75cm^2$ flask.

2.2.7.3 Co-Culture

To perform co-culture experiments, adherent cells (HS5 or HMen) were grown to 80% confluence in T75 flasks. The media was removed, the cells washed once in PBS, and suspension cells (SEM or REH, pre-treated with e450 cell proliferation dye (eBioscience) as described below) were then added in complete DMEM at a concentration of $0.5x10^6$/ml.

2.2.7.4 Investigating cholesterol metabolism *in vitro*

To examine the effects of simvastatin *in vitro*, simvastatin was obtained from Fluorochem, and dissolved in DMSO, then diluted to 0.1% DMSO in complete media. Cholesterol-ß-cyclodextran was obtained from Sigma-Aldrich, and diluted in sterile water and used in concentrations indicated below (section 5.2.2.3).

Mevalonic acid-lithium salt was obtained from Sigma (US), and dissolved in sterile water.

2.2.8 Cell counts

Two methods were used for cell counts. For all tissue-culture related cell lines, an automated cell counter (CASY® TT Cell Counter, Roche) was used as per manufacturer’s instructions. For any benchside cell counts, cells were stained in trypan blue (Sigma-Aldrich) and counted using an inverted microscope using a haemocytometer.

2.2.8.1 Haemocytometer

For cell counts using a haemocytometer, 10µL of cell solution diluted 1 in 4 in trypan blue and loaded onto the counting chamber under a coverslip. Cells were visualised using an inverted microscope, and the live cells distinguished by the exclusion of trypan blue. The number of live and dead cells in each of the 4 corner 4x4 (10^{-4}ml) squares was counted, and multiplied by10^4 to determine the sample cell live and dead cell count/ml (i.e. sample cell count/ml = cells per corner square x dilution x 10^4, or for 1 in 4 dilution: sample cell count/ml = (cells in 4 corner squares /4) x 4 (dilution) x10^4).

2.2.8.2 Automated Cell Counter

The automated Casy® TT Cell Counter was maintained and used as per manufacturer’s instruction. In brief: samples were diluted usually 1:500 in 10ml of CASY® Ton propriety buffer to around 2,000 cells/ml, and mixed gently. Where cells would not be counted immediately, this solution was stored at 4°C and carefully mixed prior to analysis. 3 replicates of 400µL of this solution were loaded onto the machine and live and dead cell counts measured by electrical impedance and recorded (i.e. live cells are insulated at the plasma membrane, and the cell volume is calculated from the impedance and recorded; dead cells’ plasma membrane is disrupted and these cells are insulated at the nuclear membrane, and this smaller nuclear volume is calculated and recorded).

2.2.9 Flow Cytometry

Flow cytometry is a technique that measures laser-activated fluorescence on individual cells. Fluorochromatic labels can be targeted to markers on the cell surface or in intracellular cytoplasm and the intensity of fluorescence measured to provide data on the abundance of those markers in a cell population. In this project, flow cytometry was used to track cell proliferation and apoptosis/cell death.

For this project a FACScalibur (BD) flow cytometer was used. Unstained, single-stain and double-stain controls together with proliferation controls and samples were loaded onto the flow cytometer. Acquisition gates were set based on forward and side scatter and double stain controls.

Data were exported and analysed using FlowJo v10.1r5, then exported for statistical analysis in Microsoft Excel 2016 v16.0.4639.1000.

2.2.9.1 Cell proliferation

In order to track proliferation, and additionally for co-culture experiments to discriminate leukaemic cells from stromal cells, an inert cytoplasmic dye was introduced into the cytoplasm of SEM and REH cells. With each subsequent cell division, the dye in the cytoplasm splits evenly between the daughter cells (i.e. the amount of dye in the cytoplasm of a cell is proportional to the number of divisions between that cell and the original cell at the start of the experiment). Cell proliferation Dye eFluor® 450 was obtained from eBioscience(ThermoFisher Scientific) and introduced into cells as per manufacturer's instructions. Briefly, SEM or REH cells were washed and resuspended in PBS at 2x the target cell concentration (usually $1x10^6$ cells/ml) a 20μM solution of the cell preparation dye in PBS at room temperature was prepared and added to an equal volume of the cell solution whilst vortexing. This mixture was incubated at 37°C in the dark for 10 minutes. The labelling was stopped by adding 4x volume of cold complete media and incubating on ice for 5 minutes. Cells were then washed x3 and resuspended in complete media.

To provide a non-proliferation control, demecolcine (Sigma-Aldrich) at 20ng/ml was added to cells stained with eFluor®450 to stop the cell cycle.

Proliferation was determined by measuring the Mean Fluorescence Intensity (MFI) of cells. In brief, cells were prepared with APC-Annexin V and Propidium Iodide (staining methods below) and resuspended in annexin V Binding buffer (BioLegend) at a concentration of approximately 250,000 cells/ml, and loaded onto the flow cytometer as above.

2.2.9.2 Apoptosis and Cell death

Apoptotic and necrotic cell death was measured using Annexin V and Propidium Iodide (PI) staining. Annexin V binds to phosphatidylserine (PS) - a protein normally found on the inner leaflet of plasma membranes. In early apoptosis, membrane organisation is lost and PS is found on the outer leaflet of cells. By measuring Annexin V binding to non-permeabilised cells, the proportion of cells in apoptosis can be assessed.

Propidium Iodide (PI) is a fluorescent molecule that does not permeate the membrane of live cells. Measuring PI staining of cells therefore allows the detection of dead cells.

To carry out this analysis APC-Annexin V Apoptosis Detection Kit with PI was acquired from Biolegend (Cat# 640932), and cells processed as per the manufacturer's instructions, except Ca-free PBS was used in place of proprietary Cell Staining Buffer. Cells were washed x2 in PBS, then resuspended in Annexin V Binding Buffer at a concentration of $1x10^6$ cells/ml. A 100µL aliquot was taken and 5µL APC-Annexin V and 10µL PI was added. Cells were gently vortexed then incubated in the dark at room temperature for 15 minutes. 400µL of Annexin V binding buffer was added, then cells were stored at 4°C and analysed immediately by flow cytometry as above. Cells were considered to be in early apoptosis if they were annexin-V positive, but PI negative. Cell were considered to be in late apoptosis if they were both annexin-V and PI positive.

2.2.9.3 Discriminating human Precursor-B cells from murine blood cells

To discriminate human precursor-B ALL cells from mouse haematopoeitic cells in our xenograft models to determine blood engraftment levels, or the proportion of murine-derived cells in spleen or CNS samples, flow cytometry was used. Anti-human CD19 antibodies with conjugated PE-Cyanine 7 (Biolegend) were used to mark human cells (anti-human CD45 antibodies were less successful as many ALL cells have only low-level CD45 expression), anti-mouse CD45 with conjugated FITC was used to mark murine leukocytes, and anti-mouse Ter119 with conjugated VioBlue (both Miltenyi) used to mark murine red cells.

To assess purity of CNS or spleen samples, 100µL aliquots of cells in 2%FCS/PBS were taken and 5ml PBS added. Cells were washed x1 in PBS, then resuspended in 50µL PBS. 5µL each of anti-human CD19-PE/Cy7, anti-murine CD45-FITC and anti-Ter119 were added, gently vortexed, and incubated for 10 minutes at room temperature in the dark, 1ml of ice-cold PBS was added. Samples were stored at 4°C in the dark, and analysed immediately as above.

To assess lymphocyte engraftment in blood, 50µL of blood in an EDTA-coated tube was taken from the tail vein of mice and added to 0.5ml of Red Cell Lysis Buffer (Miltenyi) and incubated for 10 minutes, washed twice, then resuspended in 50µL PBS. 5µL each of anti-human CD19-PE/Cy7, and anti-murine CD45-FITC was added, gently vortexed, and incubated for 10 minutes at room temperature in the dark, 1ml of ice-cold PBS was added. Samples were stored at 4°C in the dark, and analysed immediately as above.

2.2.10 *In vivo* simvastatin experiments

The efficacy of *in vivo* simvastatin was assessed as described below. Simvastatin was obtained from Fluorochem, and dissolved in DMSO at 250mg/ml. This solution was diluted 1:20 in 0.5% methylcellulose (Sigma-Aldrich) (for lower doses of simvastatin, the DMSO solution was diluted 1:50 in 0.5% methylcellulose to keep volumes of drug similar) and sonicated until a fine slurry was formed. Mice were dosed daily by oral gavage in doses as described below (section 5.3.1).

2.2.11 Genetic manipulation

In order to investigate the role of cholesterol synthesis in CNS ALL in vivo, two techniques to reduce intracellular cholesterol synthesis were employed: CRISPR-Cas9 deletion of the gene coding for the rate limiting step in cholesterol biosynthesis (HMGCoA reductase, HMGCR gene), and shRNA production targeting HMGCR RNA. Approval for genetic manipulation was given by the Institute of

Cancer Sciences GM committee and all techniques were carried out in a category 2 tissue culture facility.

2.2.11.1 CRISPR-Cas9

CRISPR-Cas9 is a system used in certain bacteria and archaea for targeting and deleting foreign or virus DNA. By repurposing this system, it is possible to introduce the Cas9 gene into a target cell's genome together with an RNA guide that will target Cas9 to a particular gene, causing a double-stranded DNA breaks and deletion of part of a gene.

To delete the HMGCR gene from SEM cells *in vitro*, a "Double Nickase" system was used. This is similar in principle to the standard CRISPR-Cas9 technique, but relies on have two offset 20-nucleotide guides for a Cas9-nickase (a version of Cas9 that cuts only one strand of DNA) binding to both strands of DNA, which will only cause a double-stranded break and gene deletion if both guides bind to the target gene. This increases the specificity of the gene targeting, though at the expense of reduced efficiency.

In order to help SEM cells survive with reduced or absent HMGCR function, the cells were grown in complete media with 40% FCS and 5mmol additional mevalonate (the product of HMGCR). The lowest puromycin dose that would cause 100% cell death at 48 hours in SEM cells in this media was determined experimentally.

A HMGCR double-nickase plasmid pair was obtained from Santa Cruz Biotechnology. One of the plasmids contained a GFP-producing gene, and the other a puromycin resistance gene allowing for double selection. The sensitivity of SEM cells in 40% FCS and 5mmol mevalonate to puromycin was assessed and an optimal dose chosen. The plasmid transfection was carried out as per the manufacturer's instructions. In brief: SEM cells were seeded in 3ml of complete media in a 6-well plate at a density of 1x105/ml, and left for 24 hours, washed and then resuspended in antibiotic-free complete DMEM. Plasmids were resuspended in DNAse-free water to form a 1g/L plasmid, 10mmol TRIS EDTA, 1mM EDTA solution, and added to proprietary "Transfection Medium" (SC-

108062), which was then added dropwise onto proprietary “Transfection reagent” (SC-395739) – an agent for the lipofection of plasmid into the cells.

After 24 hours, the culture media was replaced with DMEM with 40% FCS and 100μM mevalonate in addition to 5% glutamine and 5% penicillin-streptomycin. GFP-fluorescence was used to monitor for successful tranfection, and after 72 hours puromycin at the previously-determined optimal concentration was added to select for successfully-transfected cells.

2.2.11.2 ShRNA

An alternative method of genetic manipulation was the introduction of HMGCR shRNA into SEM cells. This was done using a lentiviral vector.

Three validated HMGCR shRNA-pLKO plasmids and a scramble-sequence shRNA-pLKO plasmid in the form of frozen bacterial glycerol stock were obtained from Sigma-Aldrich. Samples from these stocks were incubated and gently agitated overnight in Luria Bertoni Broth (E&O Labs) with amoxicillin at 38°C to grow the bacteria and plasmids purified from this bacterial culture using a Purelink HiPure Plasmid Filter Maxiprep Kit (ThermoFisher Scientific) as per manufacturer’s instructions. Briefly: bacterial cells were concentrated by centrifugation and chemically lysed. The lysate was drained by gravity though a membrane where DNA (i.e. plasmids) was bound. This membrane was then washed, and DNA eluted, the DNA washed in isopropanol, then 70% ethanol, then resuspended in Tris-EDTA buffer. DNA content and purity was measured using a Nanodrop spectrophotometer (ThermoFisher Scientific).

Lentiviral particles were then generated from each pLKO plasmid and a scramble-GFP plasmid, by transfection of HEK293 cells with the target plasmid, VSV-G envelope plasmid and PsPax2 packaging plasmid.

HEK293 cells were grown to 90% confluence in 100mm petri dishes, and media replaced with complete DMEM with 15μM chloroquine.

Each “target” plasmid DNA was mixed with packaging and envelope plasmid DNA in DNAse-free water, then calcium chloride added dropwise. This solution was then added dropwise onto agitated HBSS solution. This was left to settle for 20 minutes.

Each HBSS-DNA solution was then added dropwise onto a separate petri dish with HEK293 cells. The media was changed after 8 hours, then again after a further 16 hours. After another 24 hours, the media was collected, centrifuged at 300g for 5 minutes to pellet any non-adherent cells, the supernatant removed and run through a 45µm filter. This filtered media contained reconstituted lentivirus containing the target gene for insertion into the genome of SEM cells.

The lentiviral media was then added to SEM cells to a cell density of 2.5 x10^5/ml. These cells were plated in duplicate into 6-well (i.e. 2 wells per "target" plasmid – 3xHMGCR shRNA, 1x scramble shRNA, 1x scramble-GFP, for a total of 10 wells). After 24 hours, the lentiviral media was replaced with complete DMEM with 100µM mevalonate. Transfection efficiency was determined by assessing GFP-fluorescence in the scramble-GFP plasmid. After a further 48 hours, puromycin was added at on optimal concentration for these conditions (i.e. the lowest dose of puromycin causing 100% cell death at 48 hours in SEM cells in standard "complete" media and 100µM mevalonate) to select successfully transfected cells.

2.2.12 Bioinformatics

All raw data were converted to .xls/.xlsx/or.csv files and initial data manipulation and any statistical analysis using student's t-test carried out using Microsoft Excel 2016 v16.0.4639.1000.

Prism v 5.03 (GraphPad) was used to generate all Kaplan-Meier curves and calculate log-rank survival values unless stated otherwise.

Metaboanalyst web platform(Xia & Wishart 2011) (www.metaboanalyst.ca/) was used for ROC curve analysis of untargeted metabolomic data.

Figures were assembled using either Inkscape v0.92.2 or Microsoft Powerpoint 2016 v v16.0.4639.1000. Molecular models were made using ViewerLite™ v5.0 (Accelrys inc.).

Analysis of RNAseq data, and all other graphs and any other statistical analyses were carried out using R version 3.3.3 via Rstudio v1.0.136. Packages used included: DeSeq2, readxl, ggplot, rms, tidyr, grid, VennDiagram, together with dependencies. All R scripts and visual basic (MS Excel) scripts used for this study are published in succinct form in the appendix.

2.2.12.1 RNASeq

RNASeq analysis was carried out using the DeSeq2 package ((Love et al. 2014). Raw counts were obtained as discussed earlier from Glasgow Polyomics (section 2.2.4). Data were normalised using a median-of-ratios method. A list of genes was obtained with associated baseMean quantification of reads (i.e. a surrogate for RNA abundance for that gene), $\log_2$ fold-change between groups (CNS vs Spleen), standard error of the log-fold change, and statistical analysis for difference between groups in the form of a wald-statistic, p-value, and adjusted p-value.

2.2.12.2 Gene Set enrichment analysis

The Gene Set Enrichment Analysis v.3.0 java tool was downloaded from the Broad institute and used as per instructions (Subramanian, Tamayo, et al. (2005), PNAS 102, 15545-15550, http://www.broad.mit.edu/gsea/). “Hallmark Gene Sets”, a collection of 50 gene sets was downloaded from the broad institute website and used for analysis with an additional “cholesterol biosynthesis” gene set from KEGG pathways (https://www.genome.jp) human datasets “Terpenoid Backbone Synthesis” hsa00900, and “Steroid biosynthesis” hsa00100. 1000 permutations were analysed. All other setting were used as the default (including “dividing by mean” normalisation and “Classic” enrichment analysis). A table lisitng the most perturbed gene sets and a graphical representation of the top 20 most enriched genesets per phenotype (e.g. CNS vs spleen) was obtained.

2.2.12.3 Analysis of TARGET phase 1 data

Data were obtained from the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) initiative (managed by the United States’ National Cancer Institute (NCI)). The data used for this analysis are available ftp://caftpd.nci.nih.gov/pub/dcc_target/ALL/clinical/. Information about

TARGET can be found at http://ocg.cancer.gov/programs/target). Clinical data were downloaded from the TARGET website, and microarray data from the United States' National Centre for Biotechnology Information. Clinical file version: "TARGET_ALL_Phase_I_ClinicalData_5_8_2015_harmonized" and microarray: "GEO accession GSE11877," data were accessed on 22/03/2016.

Clinical and microarray data were merged by matching initially age in days at diagnosis, and any inconclusive results were then matched by white cell count at diagnosis. A list of genes involved in cholesterol biosynthesis was retrieved from KEGG pathways as above (section 2.2.12.2). Data for these genes were pulled from the array. Where more than 1 probe existed for a gene, the probe with highest coefficient of variation was selected. For each gene a z score was calculated for each reading using the formula:

$$\left(Z = \frac{\chi - \mu}{\sigma}\right)$$

(i.e. z-score = (sample value – mean value for that gene across all samples) / standard deviation of values for that gene across all samples). Data were then $\log_2$-transformed as detailed below (section 3.4.1) prior to analysis.A score was calculated based on number of genes in each sample which were above a particular z-score cutoff.

Chapter 3: Transcriptomic adaptations of malignant BCP-ALL cells to the central nervous system niche

3.1 Introduction and aims:

Metabolic processes in a cell can have many layers of control. Metabolic adaptations can occur at a substrate level, a protein level, at the RNA translation level, at the RNA transcription level or at a genomic level. In order to examine the metabolic adaptations of BCP-ALL cells it was decided to start by looking at the transcriptional level using RNASeq. This was felt to be the most comprehensive, technically simple, reproducible, mature and cost-effective technology.

The aim in this part of the study was to identify, using this unbiased approach, changes in the transcriptome of BCP-ALL cells in the CNS compared with systemic disease. In particular, perturbations in any genes or pathways involved in cell metabolism will be explored. These changes will then be validated in primary human ALL cells, both using xenotransplantation models of human leukaemia in mice, and in publicly available transcriptome data from primary human CNS ALL. Finally, the possibility of stratification risk of CNS relapse in children with ALL based on transcriptional signatures will be explored.

3.2 RNASeq analysis of CNS and systemic BCP-ALL cells from a murine xenograft model:

Having decided on using RNASeq technology to interrogate the transcriptome of CNS and systemic ALL blast cells it was necessary to design a model system to address our aims.

CNS disease in ALL occurs in a stereotyped manner with leukaemic blasts infiltrating the leptomeningeal space, adherent in plaques to the meningeal membranes, bathed in CSF. As discussed previously (section 1.3.2), this is a low-nutrient environment with close cellular contact between the leukaemic and stromal cells. Building a faithful model of this complex environment *in vitro* was not feasible.

Equally, using primary human leukaemia CNS and bone marrow blast cells was not feasible. Less than 5% of children have detectible CNS leukaemia on cytospin at diagnosis. The numbers of ALL cells in the CSF, when present, are usually small (typically 5,000 to 20,000 cells/ml, with less than 1ml of CSF available for research analysis). In addition, the cells that are seen on cytospin have detached from the bulk CNS disease and so may not be representative, and isolating RNA from these cells quickly enough to obtain meaningful data would be technically challenging. Finally, while DNA and/or RNA is extracted from the bone marrow of every child at diagnosis with ALL, this is done for the creation of patient-specific primers for minimal residual disease detection, and may not be done quickly enough after retrieval of cells to ensure metabolic signatures (e.g. hypoxia inducible factor (HIF)) are not lost.

Given these large challenges to obtaining primary human CNS BCP-ALL cells, it was therefore decided to use a xenograft model of childhood ALL.

It has been shown that childhood ALL cells, when introduced into immunocompromised NSG mice (mice lacking endogenous B-cells, T-cells, NK-cells and IL2Rγ receptors), typically cause leptomeningeal infiltration regardless of which original cells are used. Further in experiments with serial dilution of ALL cells prior to introduction into NSG mice it has been suggested that at least 1-in-10 ALL blasts is capable of causing CNS disease(Williams et al. 2016).

The pattern of CNS infiltration in these xenograft models is leptomeningeal plaques with complete sparing of the brain/spinal cord parenchyma – a very similar pattern to that seen in human disease(Thomas et al. 1964). In addition it is possible to extract quickly, at high purity, and in good cell numbers BCP-ALL cells from the CNS and systemic compartments for extraction of RNA.

For this experiment BCP-ALL cell lines were used. Whilst as noted above it is entirely feasible to use primary human BCP-ALL cells in our model of CNS leukaemia, primary cells have a longer time until end-point, and a less consistent pattern of engraftment between and within experiments. By using cell lines it is possible to plan and execute experiments to maximise the quality of data produced.

Two cell lines were selected for these experiments: SEM cells, BCP-ALL cells driven by an MLL t(4;11) translocation, to represent high-risk disease; and REH cells, BCP-ALL cells with TEL-AML1 t(12;21) translocation, to represent standard-risk disease. Cells were injected into NSG mice, and cells retrieved as above (section 2.2.1.1)..

RNA extraction was carried out by Dr O Olivares (postdoctoral researcher in the research group) as detailed above (section 2.2.3.1). In order to maximise the robustness of our results while controlling cost, RNA from the CNS and Spleen from 3-4 mice was pooled and each xenograft experiment repeated twice for SEM and 3 times for REH. Unfortunately RNA of a sufficient quality could only be obtained from Spleen ALL cells for 2 of the replicates in the REH experiments. The RNA was then given to the University of Glasgow Glasgow Polyomics Department for processing and read alignment. As an additional control, RNA from cells retrieved from standard tissue culture was analysed in parallel.

The RNA was sequenced, aligned to human and mouse genomes (with reads aligning to the mouse genome then removed), and measured using the NextSeq® 500 platform, as detailed above (section 2.2.4) and analysed for differential expression using the DESeq2 package(Love et al. 2014) as described above (section 2.2.12.1).

3.2.1 Gene expression of ALL cells in the CNS compared with spleen.

Using these data, the results of these experiments were analysed at multiple levels. Firstly the results were examined at a transcriptome-wide level, with principle component analysis and hierarchical clustering models showing clear clustering of the CNS samples compared with the spleen, and a clear separation between both *in vivo* groups and the *in vitro* control (Figure 3-1).

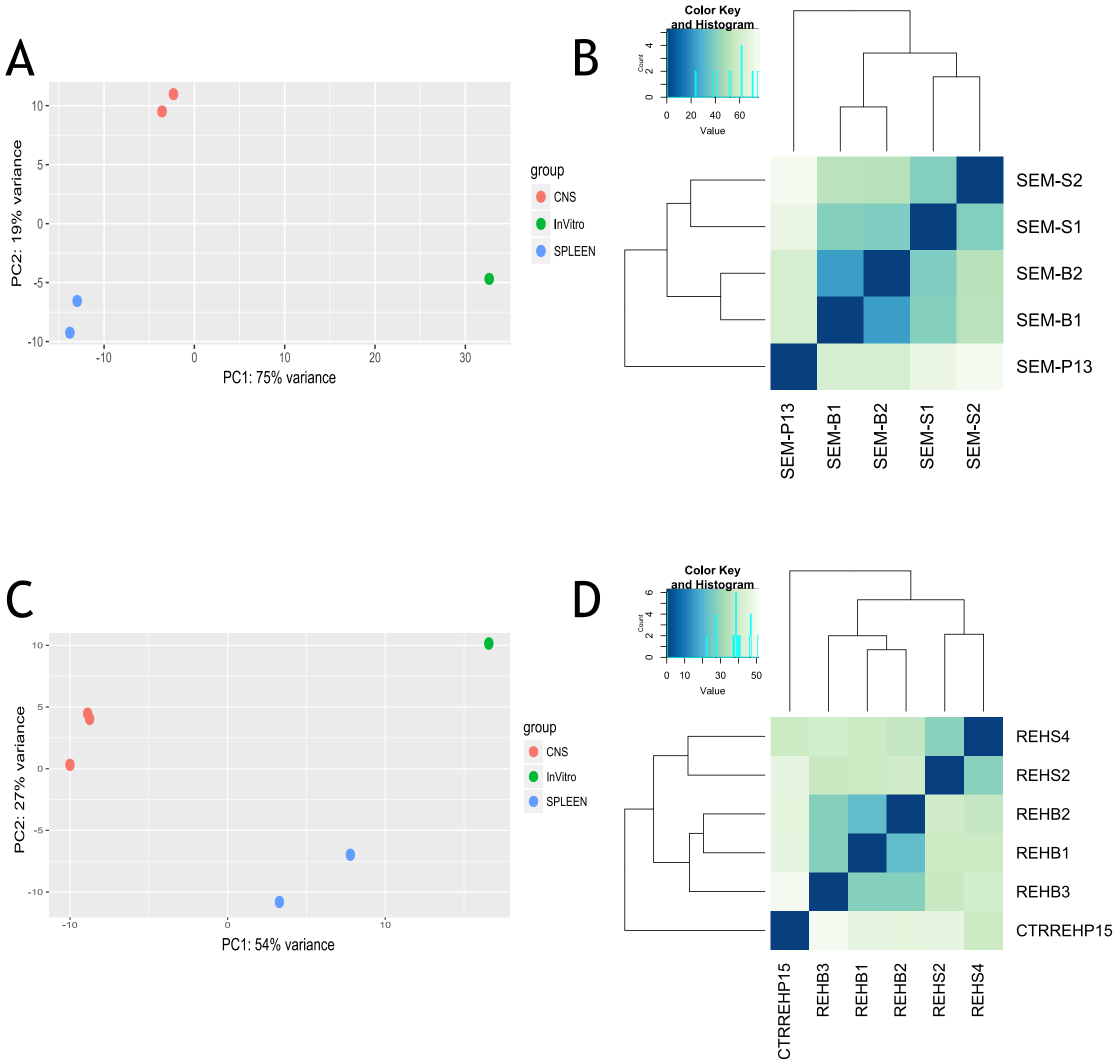


Figure 3-1 Principal Component Analysis (PCA) plot and Cluster Heatmap/Dendrogram for RNASeq datafrom A,B - SEM cells retrieved from murine CNS and Spleen, and from Tissue Culture (A – PCA, B - Cluster Heatmap/dendrogram. Data from the cells retrieved from spleens labelled SEM-S1/2, data from the cells retrieved from CNS labelled SEM-B1/2, data from the cells retrieved from tissue culture labelled SEM-P13); C,D – REH cells retrieved from murine CNS and Spleen, and from Tissue Culture (C – PCA, D - Cluster Heatmap/dendrogram. Data from the cells retrieved from spleens labelled REHS2/4, data from the cells retrieved from CNS labelled REHB1/2/3, data from the cells retrieved from tissue culture labelled CTRREHP15).

Given the significant shift between *in vivo* and *in vitro* results with these genome-wide perspectives it was decided to conduct further analyses using data from the *in vivo* CNS and spleen compartments with the exclusion of *in vitro* results (PCA and cluster Heatmap/dendrogram below).

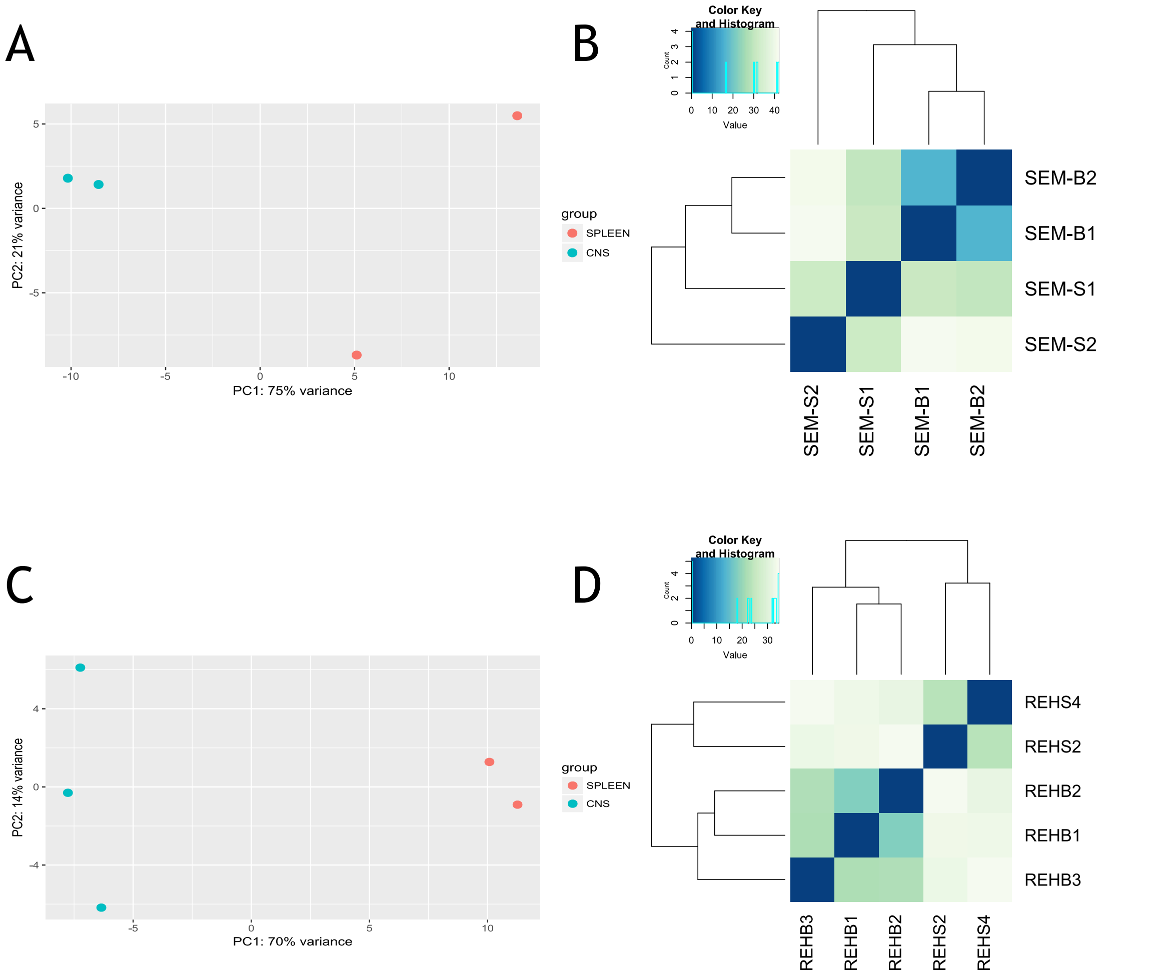
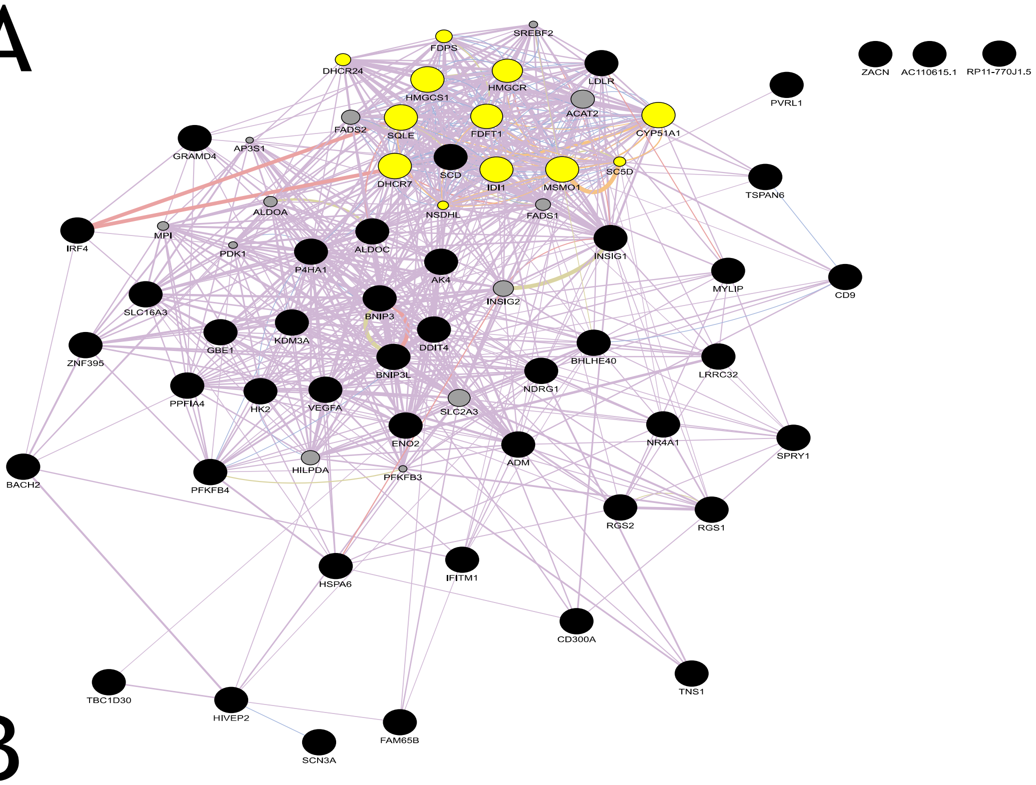


Figure 3-2 Principal Component Analysis (PCA) plot and Cluster Heatmap/Dendrogram for RNASeq datafrom A,B - SEM cells retrieved from murine CNS and Spleen (A – PCA, B - Cluster Heatmap/dendrogram. Data from the cells retrieved from spleens labelled SEM-S1/2, data from the cells retrieved from CNS labelled SEM-B1/2; C,D – REH cells retrieved from murine CNS and Spleen (C – PCA, D - Cluster Heatmap/dendrogram. Data from the cells retrieved from spleens labelled REHS2/4, data from the cells retrieved from CNS labelled REHB1/2/3).

Next the data were examined at a pathway and gene network level. Network analysis was performed on the GeneMANIA platform(Warde-Farley et al. 2010). The top 50 most significantly differentially expressed genes (ranked by adjusted p-value) were analysed using the Genemania algorithm(Mostafavi et al. 2008). This created a network map using gene-ontology (GO) database(Consortium 2015) to link these genes together (the gene lists are available below in Table 3-1 (SEM) and Table 3-2(REH)). This map was then interrogated to determine the predicted functions of the differentially expressed gene networks. All of the top ranked networks in this analysis were metabolic. The most statistically significantly altered network between CNS and spleen ALL was cholesterol biosynthesis (SEM: Figure 3-3, REH: Figure 3-4).

Next, pathway analysis was carried out using Geneset Enrichment Analysis(Subramanian et al. 2005). This method ranks all genes based on $\log_2$ fold-change, and uses this ranking to determine which pathways are "enriched" in one experimental group compared with another. For our analysis we used the GSEA "hallmark" genesets - 50 sets of curated gene pathways based on published data(available via the Broad Institute, MA)(Liberzon et al. 2015), plus an addition bespoke pathway containing only genes directly involved in cholesterol biosynthesis (determined using the KEGG database(Kanehisa & Goto 2000; Kanehisa et al. 2016; Kanehisa et al. 2017)). Using this approach, cholesterol metabolism (both the standard "Hallmark Cholesterol Homeostasis" geneset and the custom "TC Cholesterol Biosynthesis" geneset) were found to be highly significantly upregulated in cells from the CNS (SEM: Figure 3-5, REH: Figure 3-6).

A

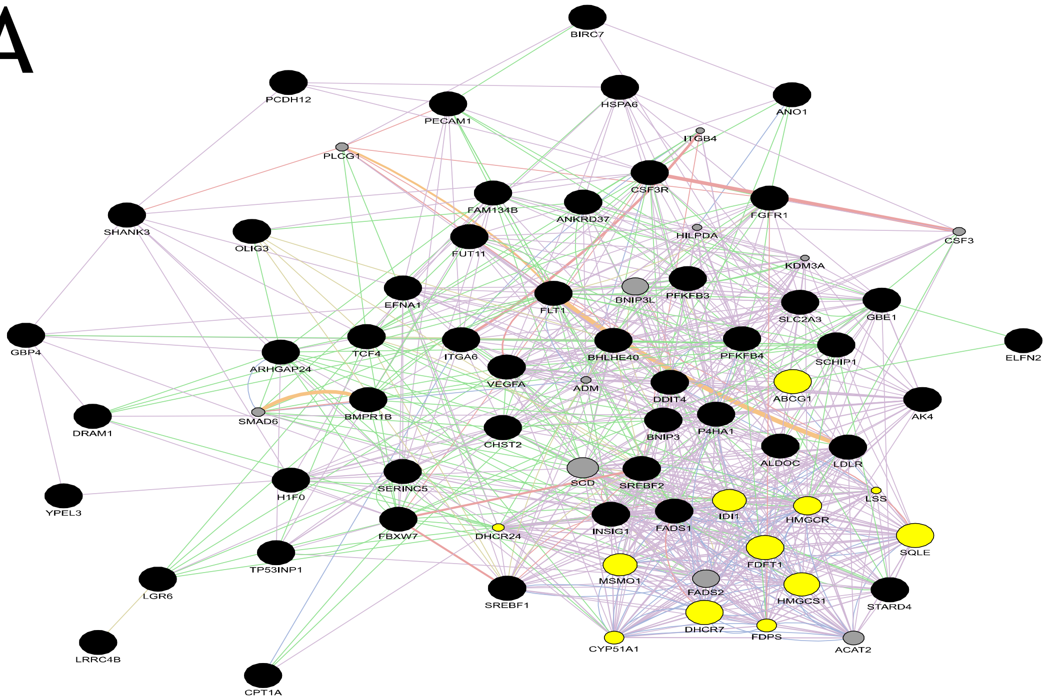


B

GO id	Description	q-value	Occurrences in Sample	Occurrences in Genome
GO:0006695	cholesterol biosynthetic process	3.58E-18	12	33
GO:0016126	sterol biosynthetic process	3.58E-18	12	34
GO:0008203	cholesterol metabolic process	3.51E-14	12	69
GO:0006694	steroid biosynthetic process	1.36E-13	13	107
GO:0016125	sterol metabolic process	1.40E-13	12	80
GO:0046165	alcohol biosynthetic process	2.19E-12	12	101
GO:1901617	organic hydroxy compound biosynthetic process	2.61E-11	12	125
GO:0008202	steroid metabolic process	2.02E-10	13	196
GO:0006066	alcohol metabolic process	8.07E-09	13	264
GO:0006096	glycolysis	2.43E-05	6	48
GO:0005996	monosaccharide metabolic process	5.67E-05	9	209
GO:0006007	glucose catabolic process	6.51E-05	6	58
GO:0006006	glucose metabolic process	9.79E-05	8	161
GO:0019320	hexose catabolic process	2.08E-04	6	72
GO:0019318	hexose metabolic process	2.14E-04	8	182
GO:0046365	monosaccharide catabolic process	2.14E-04	6	74
GO:0044724	single-organism carbohydrate catabolic process	0.001012	6	97
GO:0016052	carbohydrate catabolic process	0.001286	6	102
GO:0019200	carbohydrate kinase activity	0.024297	3	17
GO:0001666	response to hypoxia	0.042823	5	116
GO:0036293	response to decreased oxygen levels	0.044261	5	118
GO:0070482	response to oxygen levels	0.045782	5	120
GO:0071456	cellular response to hypoxia	0.067891	4	68
GO:0036294	cellular response to decreased oxygen levels	0.072883	4	70
GO:0071453	cellular response to oxygen levels	0.078112	4	72

Figure 3-3 Results of geneMANIA network analysis of RNASeq data from SEM cells retrieved from murine CNS vs spleen based on top 50 differentially expressed genes (ranked by adj. p-value) A –Network map, with cholesterol biosynthesis genes highlighted (yellow). Lines show gene interactions, node size is proportional to network interaction score; B - Table of predicted differences in cell processes between cells from CNS vs spleen.

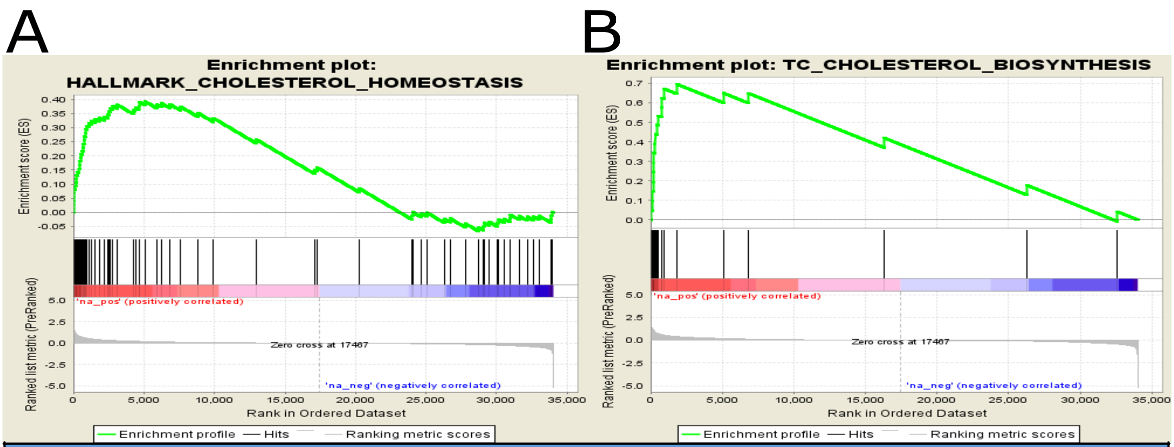
A



B

GO id	Description	q-value	Occurrences in Sample	Occurrences in Genome
GO:0006695	**cholesterol biosynthetic process**	**1.13E-17**	**12**	**33**
GO:0016126	sterol biosynthetic process	1.13E-17	12	34
GO:0008203	cholesterol metabolic process	1.10E-13	12	69
GO:0006694	steroid biosynthetic process	4.37E-13	13	107
GO:0016125	sterol metabolic process	4.37E-13	12	80
GO:0046165	alcohol biosynthetic process	6.79E-12	12	101
GO:1901617	organic hydroxy compound biosynthetic process	8.03E-11	12	125
GO:0008202	steroid metabolic process	6.77E-10	13	196
GO:0006066	alcohol metabolic process	1.45E-09	14	264
GO:0010883	regulation of lipid storage	0.0138	4	34
GO:0010866	regulation of triglyceride biosynthetic process	0.017944	3	12
GO:0019915	lipid storage	0.029849	4	43
GO:0090207	regulation of triglyceride metabolic process	0.065651	3	19
GO:0046982	protein heterodimerization activity	0.097163	6	190

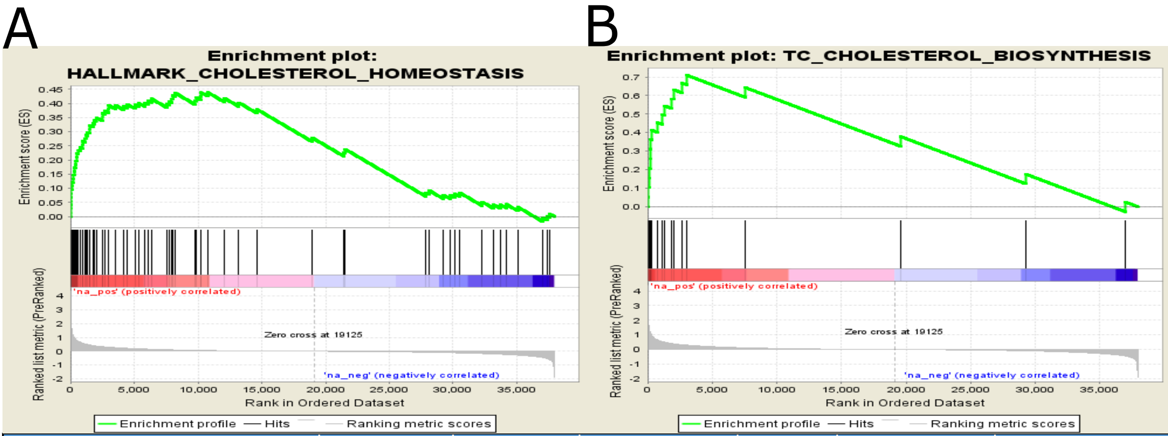
Figure 3-4 Results of geneMANIA network analysis of RNASeq data from REH cells retrieved from murine CNS vs spleen based on top 50 differentially expressed genes (ranked by adj. p-value) A –Network map, with cholesterol biosynthesis genes highlighted (yellow). Lines show gene interactions, node size is proportional to network interaction score; B - Table of predicted differences in cell processes between cells from CNS vs spleen.



C

Name of GeneSet	No. of genes in GeneSet	Enrichment Score	Normalised Enrichment Score	Nominal p-value	False Discovery Rate q-value	Familywise Error Rate p-value
Hallmark TNF-a signaling via NF?B	190	0.321	5.226	<0.00001	<0.00001	<0.00001
Hallmark cholesterol homeostasis	72	0.393	3.979	<0.00001	<0.00001	<0.00001
Hallmark interferon gamma response	191	0.243	3.921	<0.00001	<0.00001	<0.00001
Hallmark hypoxia	188	0.246	3.910	<0.00001	<0.00001	<0.00001
Tc cholesterol biosynthesis	20	0.697	3.844	<0.00001	<0.00001	<0.00001
Hallmark UV response DN	141	0.271	3.758	<0.00001	<0.00001	<0.00001
Hallmark mitotic spindle	198	0.222	3.657	<0.00001	<0.00001	<0.00001
Hallmark estrogen response early	183	0.217	3.475	<0.00001	<0.00001	<0.00001
Hallmark KRAS signaling up	170	0.232	3.425	<0.00001	<0.00001	<0.00001
Hallmark allograft rejection	178	0.197	3.072	<0.00001	<0.00001	<0.00001
Hallmark IL2 STAT5 signalling	197	0.185	2.981	<0.00001	0.0001	0.001
Hallmark interferon alpha response	93	0.262	2.903	<0.00001	0.00009	0.001
Hallmark inflammatory response	181	0.171	2.665	<0.00001	0.00009	0.001
Hallmark androgen response	97	0.220	2.55	<0.00001	0.00017	0.002
Hallmark KRAS signaling DN	147	0.171	2.409	<0.00001	0.00077	0.01
Hallmark complement	172	0.146	2.247	<0.00001	0.00197	0.028

Figure 3-5 Results of GeneSet Enrichment Analysis of RNASeq data from SEM cells retrieved from murine CNS vs spleen A –Enrichment plot for "Hallmark" cholesterol homeostasis pathway; B – Enrichment plot for bespoke "TC cholesterol biosynthesis" pathways; C – Geneset Enrichment Analysis table of the top 16 differentially expressed pathways using the MSigDB "Hallmark" pathways database and bespoke "TC" cholesterol synthesis pathways



C

Name of Geneset	No. of genes in GeneSet	Enrichment Score	Normalised Enrichment Score	Nominal p-value	False Discovery Rate q-value	Familywise Error Rate p-value
Hallmark TNFa signaling via NFKB	195	0.33557	5.376934	<0.00001	<0.00001	<0.00001
Hallmark hypoxia	194	0.330877	5.304321	<0.00001	<0.00001	<0.00001
Hallmark interferon gamma response	191	0.305322	4.907172	<0.00001	<0.00001	<0.00001
Hallmark cholesterol homeostasis	72	0.43975	4.361982	<0.00001	<0.00001	<0.00001
Hallmark interferon alpha response	92	0.344635	3.900943	<0.00001	<0.00001	<0.00001
Hallmark inflammatory response	190	0.238359	3.890412	<0.00001	<0.00001	<0.00001
Hallmark IL2 STAT5 signalling	196	0.230353	3.756217	<0.00001	<0.00001	<0.00001
Tc cholesterol biosynthesis	19	0.71073	3.716716	<0.00001	<0.00001	<0.00001
Hallmark UV response DN	140	0.267056	3.676804	<0.00001	<0.00001	<0.00001
Hallmark allograft rejection	179	0.22784	3.564309	<0.00001	<0.00001	<0.00001
Hallmark estrogen response early	187	0.2212	3.541693	<0.00001	<0.00001	<0.00001
Hallmark p53 pathway	194	0.210307	3.361116	<0.00001	<0.00001	<0.00001
Hallmark androgen response	96	0.277482	3.150325	<0.00001	<0.00001	<0.00001
Hallmark apoptosis	157	0.206014	3.015861	<0.00001	<0.00001	<0.00001
Hallmark complement	183	0.186569	2.935609	<0.00001	0.00008	0.001
Hallmark KRAS signaling up	171	0.178464	2.682808	<0.00001	0.00013	0.002

Figure 3-6 Results of GeneSet Enrichment Analysis of RNASeq data from REH cells retrieved from murine CNS vs spleenA –Enrichment plot for "Hallmark" cholesterol homeostasis pathway; B – Enrichment plot for bespoke "TC cholesterol biosynthesis" pathways; C – Geneset Enrichment Analysis table of the top 16 differentially expressed pathways using the MSigDB "Hallmark" pathways database and bespoke "TC" cholesterol synthesis pathways

With these results showing an enrichment of cholesterol biosynthesis in the CNS at a gene pathway level, the dataereas analysed at the level of individual genes. Of the 20 genes directly involved in cholesterol biosynthesis, all but 2 were upregulated, and 13 were statistically significant with an adjusted p-value of <0.05, with both downregulated genes being included in the 7 genes that did not have significantly differential expression between groups in SEM cells, with PMVK (phosphomevalonate kinase) being significantly reduced in REH cells (adj. p-value 0.017) (Figure 3-7).

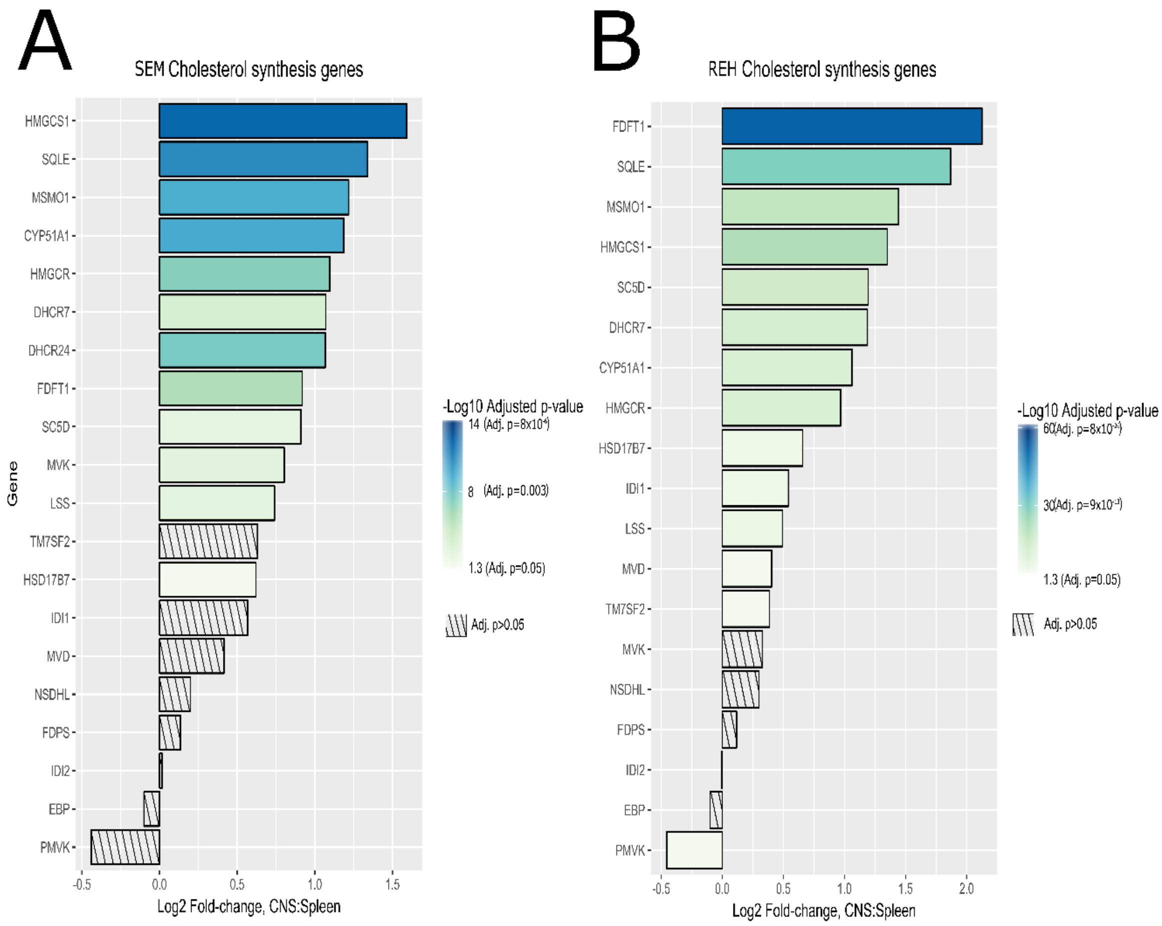


Figure 3-7 Waterfall plot of cholesterol synthesis gene differential expression by RNASeq for ALL cells retrieved from murine CNS and spleenA – SEM cells; B – REH cells; ordered by $\log_2$ fold-change CNS vs Spleen; $-\log_{10}$ p-value calculated using DESeq2 package and denoted by colour, genes without statistically significant different expression denoted by diagonal lines. Please note differences in colour scales between graphs.

The top 50 genes most significantly regulated in the CNS are listed in Table 3-1 (SEM) and Table 3-2 (REH).

Table 3-1 Top 50 differentially expressed genes in SEM ALL cells in the CNS compared to spleen

GENE ID	GENE EXPRESSION INTENSITY	LOG_2 FOLD CHANGE CNS:SPLEEN	ADJUSTED P-VALUE
BHLHE40	854.290057	3.579513952	3.72E-64
RP11-473I1.9	515.4523942	-5.189958922	4.82E-64
AK4	732.1970873	3.899323918	1.66E-63
H3F3AP4	266.5470379	4.876411497	1.80E-58
SLC16A3	420.6139571	4.061443941	5.51E-56
RGS1	241.6029986	4.704041172	1.20E-55
ALDOC	1379.277908	2.988659645	5.90E-52
SCD	2603.707357	2.652232456	3.17E-42
RP11-513I15.6	466.3420583	-3.18635547	1.09E-39
PFKFB4	661.1011637	3.357693729	2.62E-39
NDRG1	1941.714621	2.268400456	1.14E-35
CD9	703.3309409	3.167564359	2.24E-34
TBC1D30	183.3962127	3.388150826	2.05E-32
VEGFA	389.0034644	2.710617233	2.51E-27
RP11-386G11.10	2608.084034	-2.926810724	2.60E-27
ENO2	171.8322524	3.144849166	2.88E-26
RP11-452N17.1	86.91834009	-3.52970914	6.13E-26
BNIP3	1716.094338	1.961330535	3.27E-23
RP11-770J1.5	466.0528964	-2.532610169	3.34E-21
GBE1	2653.827839	1.681188404	1.80E-19
LRRC32	93.11256535	2.936056676	3.00E-19
MIR210HG	167.3653814	2.433883147	1.07E-17
CERS6-AS1	190.9033216	-2.821392463	5.33E-17
HK2	5808.813739	1.536358544	2.31E-16
LDLR	1988.388306	1.577382283	4.13E-16
AC110615.1	91.67609714	-2.87063744	5.40E-16
IRF4	535.4381629	1.905642971	5.40E-16
SPRY1	100.6324779	2.635676151	6.51E-16
BACH2	733.989599	1.844663082	9.41E-16
SCN3A	742.4200309	2.040778596	2.06E-15
RGS2	482.5503159	1.884242068	3.83E-15
CD300A	128.0680756	2.551295437	3.96E-15
INSIG1	4005.011772	1.58987599	5.40E-15
TNS1	801.272915	1.61716589	1.43E-14
HMGCS1	5900.975162	1.591277278	4.86E-13
P4HA1	2728.475608	1.333715342	1.02E-12
PVRL1	384.550124	1.861383131	5.79E-12

ZACN	271.2326783	-2.131024036	5.79E-12
HIVEP2	1001.984364	1.472539777	8.00E-12
PPFIA4	81.44481842	2.301019404	1.17E-11
SQLE	2936.861533	1.339166413	1.33E-11
ADM	48.68816604	2.435307116	1.59E-11
NR4A1	72.9122992	2.299248565	4.68E-11
MYLIP	1072.514875	1.379854974	5.33E-11
BNIP3L	4715.004603	1.387083128	6.94E-11
DDIT4	4483.92922	1.515689567	8.37E-11
FAM65B	2572.931109	1.598586739	1.72E-10
RP11-463I20.1	416.2678673	-1.475422195	3.11E-10
CYP51A1	2836.839519	1.185528456	3.41E-10
MSMO1	1933.47185	1.218348567	4.19E-10

Table 3-2 Top 50 differentially expressed genes in REH ALL cells in the CNS compared to spleen

GENE ID	GENE EXPRESSION INTENSITY	LOG_2 FOLD CHANGE CNS:SPLEEN	ADJUSTED P-VALUE
CSF3R	3250.179	2.96689	8.48E-96
BHLHE40	1036.656	4.341969	9.11E-91
LDLR	1972.6	3.095277	4.77E-90
ELFN2	1709.634	3.672371	2.81E-86
STARD4	538.2924	3.094963	1.59E-81
SHANK3	3086.683	2.665201	3.42E-72
SREBF1	3593.836	2.257933	1.09E-67
VEGFA	767.7651	3.519427	8.06E-67
SREBF2	8210.101	1.935486	9.54E-66
H1F0	2273.007	2.498565	2.77E-65
RP11-18H21.1	1323.096	2.067241	9.50E-57
FDFT1	6521.643	2.125536	2.17E-54
ALDOC	1002.504	2.676968	3.66E-51
HSPA6	245.2923	3.799504	2.38E-50
ABCG1	469.4563	2.882726	3.73E-46
AK4	307.1934	3.187793	2.43E-45
TP53INP1	1013.493	2.177063	8.37E-43
BNIP3	2902.897	2.06596	3.09E-41
CPT1A	8671.02	-1.4552	3.27E-41
OLIG3	1935.778	2.202999	3.28E-40
FBXW7	3573.935	1.375641	2.55E-38
LGR6	2439.537	1.595821	4.31E-38
PFKFB4	835.4546	2.372291	1.38E-36

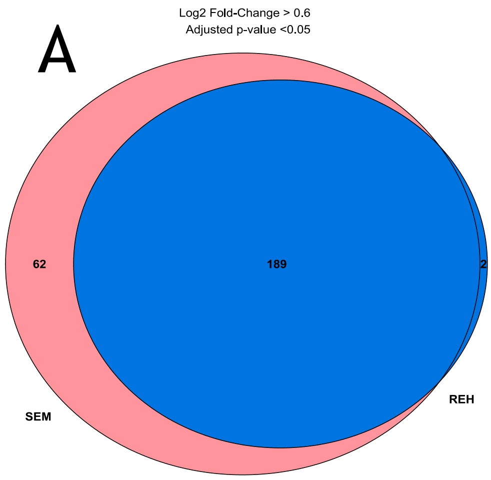
INSIG1	5191.641	2.06251	4.52E-35
ANO1	915.9114	1.809972	5.26E-35
DDIT4	2440.804	1.86262	6.18E-34
DRAM1	1351.463	1.697338	1.85E-32
FLT1	453.2486	2.481939	1.92E-32
YPEL3	2515.797	1.297044	2.23E-32
P4HA1	1993.925	1.769766	3.71E-32
FADS1	6938.226	1.260015	3.71E-32
EF1	436.4398	2.612551	4.39E-31
SQLE	2604.018	1.868471	3.92E-30
ANKRD37	230.0075	2.199052	2.55E-29
GBP4	2724.729	1.496766	3.10E-29
CHST2	231.1591	2.294324	5.34E-29
GBE1	2945.173	1.19739	5.71E-29
BIRC7	2310.026	1.674353	7.25E-29
ARHGAP24	328.4362	2.058197	1.05E-28
FGFR1	6803.806	-1.23488	5.88E-28
SCHIP1	257.5327	2.116252	3.14E-27
PECAM1	386.5941	2.470171	6.51E-27
SERINC5	4710.443	1.399024	6.71E-27
PCDH12	307.3392	2.193775	1.90E-26
PFKFB3	3137.987	1.416325	3.74E-26
FAM134B	1248.153	1.667611	1.01E-25
TCF4	12818.29	1.103448	1.53E-25
FUT11	890.3417	1.541721	2.48E-25
SLC2A3	908.1944	1.586512	3.81E-25
FLJ16779	1336.113	1.626996	8.33E-25

3.2.2 Combined analysis of genes differentially expressed between murine CNS and spleen in both REH and SEM cells by RNASeq

When the data from both SEM and REH cell lines were considered together, as suggested by the results above, there was significant overlap. Network analysis of the 189 genes differentially regulated between CNS and Spleen in both cell lines confirmed differential expression of cholesterol synthesis genes to be common to both cell lines (Figure 3-8).

B

GO id	Description	q-value	Occurrences in Sample	Occurrences in Genome
GO:0006695	cholesterol biosynthetic process	6.23E-18	15	33
GO:0016126	sterol biosynthetic process	6.23E-18	15	34
GO:0008203	cholesterol metabolic process	2.89E-14	16	69
GO:0016125	sterol metabolic process	2.79E-13	16	80
GO:0006694	steroid biosynthetic process	2.85E-11	16	107
GO:0046165	alcohol biosynthetic process	1.89E-10	15	101
GO:0006066	alcohol metabolic process	3.79E-09	20	264
GO:1901617	organic hydroxy compound biosynthetic process	3.79E-09	15	125
GO:0008202	steroid metabolic process	1.99E-08	17	196
GO:0001666	response to hypoxia	0.036047	8	116
GO:0036293	response to decreased oxygen levels	0.037112	8	118
GO:0070482	response to oxygen levels	0.038433	8	120
GO:0071456	cellular response to hypoxia	0.079572	6	68
GO:0036294	cellular response to decreased oxygen levels	0.087092	6	70
GO:0071453	cellular response to oxygen levels	0.095308	6	72



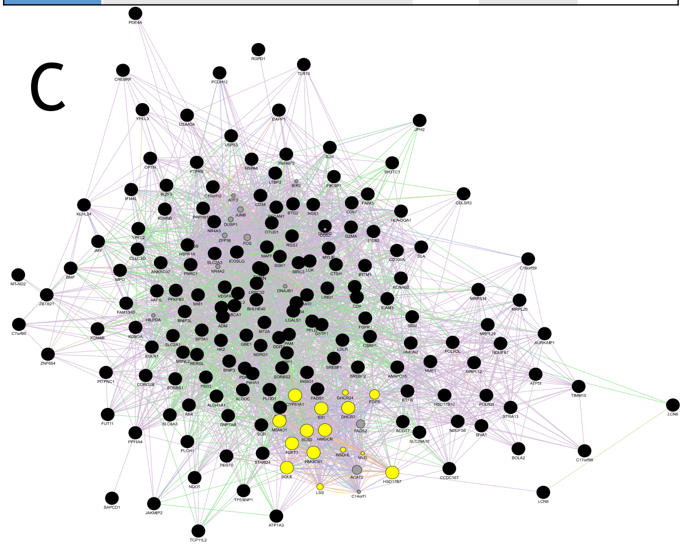


Figure 3-8 Combined analysis of genes differentially expressed between murine CNS and Spleen in both REH and SEM cells by RNASeq A – Venn diagram showing number of genes differentially expressed between murine CNS and spleen with Log_2 fold-change >0.6 or <-0.6, and adjusted p-value <0.05. B – Table of differentially expressed cell processes based on GeneMANIA analysis of genes differentially expressed between murine CNS and Spleen in both SEM and REH cells; C - GeneMANIA network map of genes differentially expressed between murine CNS and Spleen in both SEM and REH cells with cholesterol biosynthesis genes highlighted (yellow). Lines show gene interactions, node size is proportional to network interaction score.

3.2.3 Summary

RNASeq analysis of ALL cells retrieved from the CNS and spleen of xenotransplanted mice have shown clear transcriptomic changes between the two compartments. These changes are consistent between two different cell lines (SEM and REH), and support the hypothesis that ALL cells undergo metabolic adaptation to the CNS niche. In particular, pathway analyses using

GSEA and GeneMANIA platforms show increased cholesterol biosynthesis in ALL cells retrieved from the CNS.

3.3 Validation of RNASeq data in primary ALL cells

The next stage in the project was to validate these findings in primary human cells. As noted above, whilst using cell lines gives more consistent engraftment and CNS infiltration using the NSG mouse model of leukaemia, it is possible to engraft primary human leukaemic cells. Primary human ALL cells were taken from the bone marrow of children with ALL at diagnosis then stored in liquid nitrogen. Two sources of primary human ALL cells were used in this analysis: cells thawed and directly xenografted into mice (primary cells), and cells that had previously been xenografted into mice, then had cells retrieved from the murine spleen and refrozen in liquid nitrogen (primagraft cells). In this project I used 3x primary, and 3x primagraft ALL cells (i.e. cells from 6 children). Patient characteristics are summarised previously (**Error! Reference source not found.**). As noted previously , REH cells carry a t(12;21) translocation similar to patients 1-3, and SEM cells carry a t(4;11) translocation similar to patients 4-6.

The only major difference in technique between using cell lines (as described above) and primary ALL cells is that to improve engraftment cells are injected directly into the femurs of the mice under general anaesthetic as previously described (section 2.2.2.2), and that the time to engraftment varies. In these experiments the time to engraftment, and ALL cell count and purity when retrieved at cull is detailed in the table below (Table 3-3).

Table 3-3 Details of primary and primagraft cells engrafted into NSG mice and harvested from CNS and spleen at end of experiment

Cell ID	Translocation	Number of Mice at end of experiment	Mean duration of experiment (Days)	Site	Mean Cell Count from each site ($x10^6$ cells)	Mean ALL Cell Purity
Patient 1 (primagraft cell) (1)	TEL:AML1 t(12;21)	3	135	CNS	6.2	90%
				Spleen	170	79%
Patient 1 (primagraft cell) (2)	TEL:AML1 t(12;21)	2	170	CNS	15	99%
				Spleen	264	99%
Patient 2 (primagraft cell) (1)	TEL:AML1 t(12;21)	3	188	CNS	12	99%
				Spleen	235	98%
Patient 2 (primagraft cell) (2)	TEL:AML1 t(12;21)	3	184	CNS	15	98%
				Spleen	60	97%
Patient 3 (primagraft cell) (1)	TEL:AML1 t(12;21)	3	196	CNS	3	
				Spleen	80	88%
Patient 4 (primary cell)	MLL t(4;11)	4	188	CNS	14	99%
				Spleen	143	97%
Patient 5 (primary cell)	MLL t(4;11)	3	125	CNS	0.8	81%
				Spleen	8.5	85%
Patient 6 (primary cell)	MLL t(4;11)	4	No engraftment			

RNA was extracted from these cells using the RNeasy system, and analysed using the Fluidigm® multiplex PCR system as detailed previously (section **2.2.5.4**). PCR primers were designed for each of the genes in the cholesterol biosynthesis pathway. In addition, primers were designed for genes involved in lipid metabolism – specifically a selection of genes involved in fatty acid metabolism, ACSS genes involved in the conversion of acetate to acetyl CoA, and the sterol-regulatory element-binding transcription factor (SREBF) subtypes involved in regulation of cholesterol and lipid synthesis - Interferon regulatory factor 4 (IRF4)(Tun et al. 2008) (an interferon signalling pathway gene noted to be differentially expressed in our RNASeq data and in published reports of CNS lymphoma), and genes that have been identified in the literature as differentially regulated between CNS and systemic ALL (SPP1/osteopontin(van der Velden et al. 2015), VEGFα(Münch et al. 2017; Kato et al. 2017), IL15(Cario et al. 2007b; Williams et al. 2014), ICAM-1(Holland et al. 2011; Mielcarek et al. 1997), and MERTK(Krause et al. 2015)). .

3.3.1 Results of multiplex PCR of cholesterol synthesis genes in primagraft cells

Focussing on the results of the cholesterol biosynthesis pathway genes in the first instance, not unexpectedly the results were not as consistent with primary cell lines compared with the RNASeq data on cell lines. Only 2 genes in the cholesterol biosynthesis pathway (IDI1 (p=0.04) and MSMO (p=0.03)) reached statistical significance even before statistical correction for multiple testing (Figure 3-9).

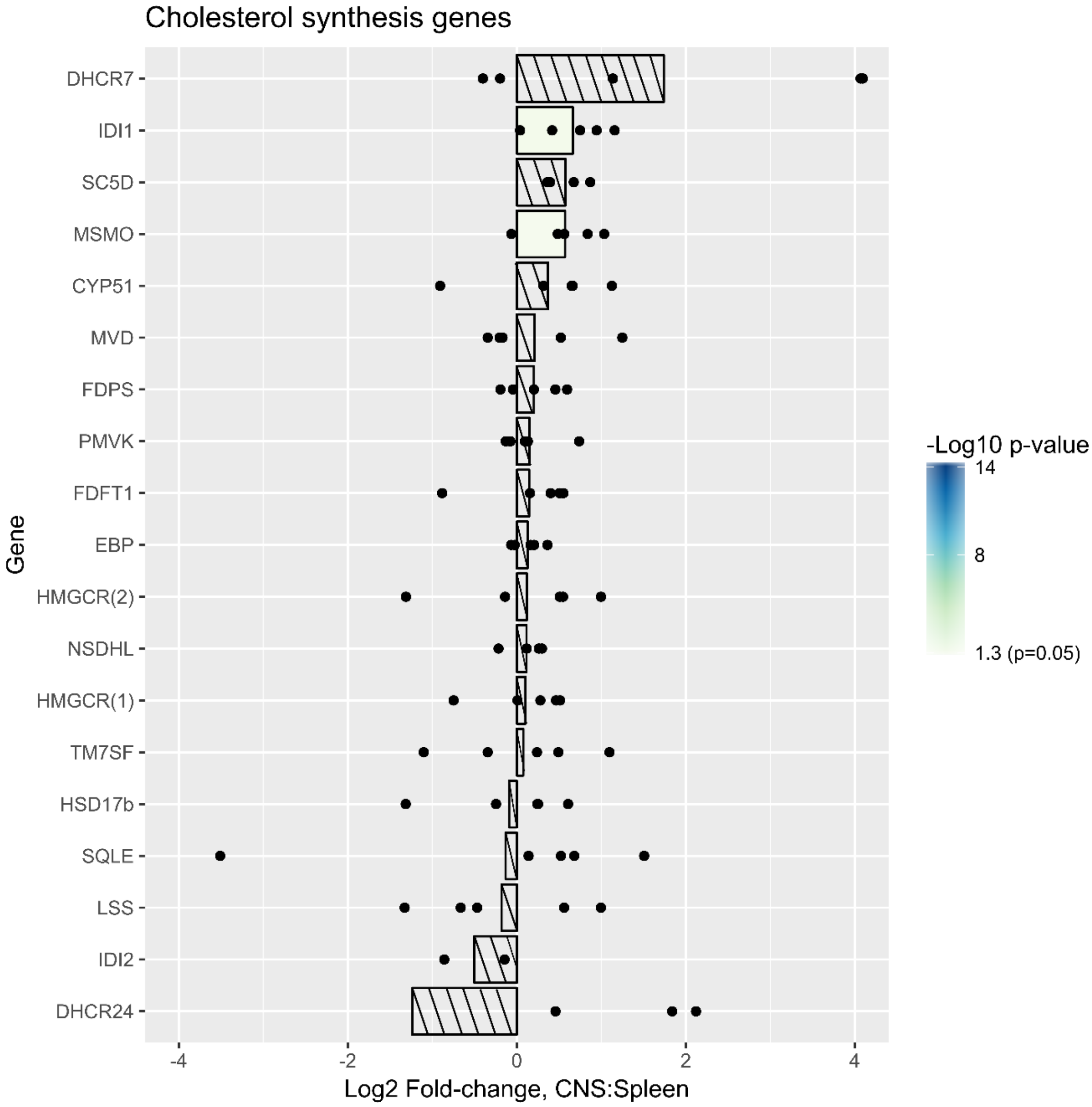


Figure 3-9 Waterfall plot of cholesterol synthesis gene differential expression by Fluidigm® multiplex PCR for primary and primagraft ALL cells retrieved from murine CNS and Spleen. Mean Log_2 fold-change noted by bars and individual patient data denoted by points. Genes ordered by log2 fold-change CNS vs Spleen; -log10 p-value (student's paired t-test) denoted by colour, genes without statistically significant differential expression denoted by diagonal lines.

Looking at these results in more depth, in 4 of the 5 patients there appears to be upregulation of cholesterol synthesis overall (patients 1, 2, 3, and 4; with 13/18, 13/18, 12/18, and 15/18 genes upregulated vs downregulated respectively). In the remaining patient (patient 5), the pattern is more balanced with 9/18 genes upregulated. These data were either from CNS and spleen of 3 mice (Patients 1 and 3), or from 1 mouse each (patients 2, 4, and 5), so statistical analysis was not performed (Figure 3-10).

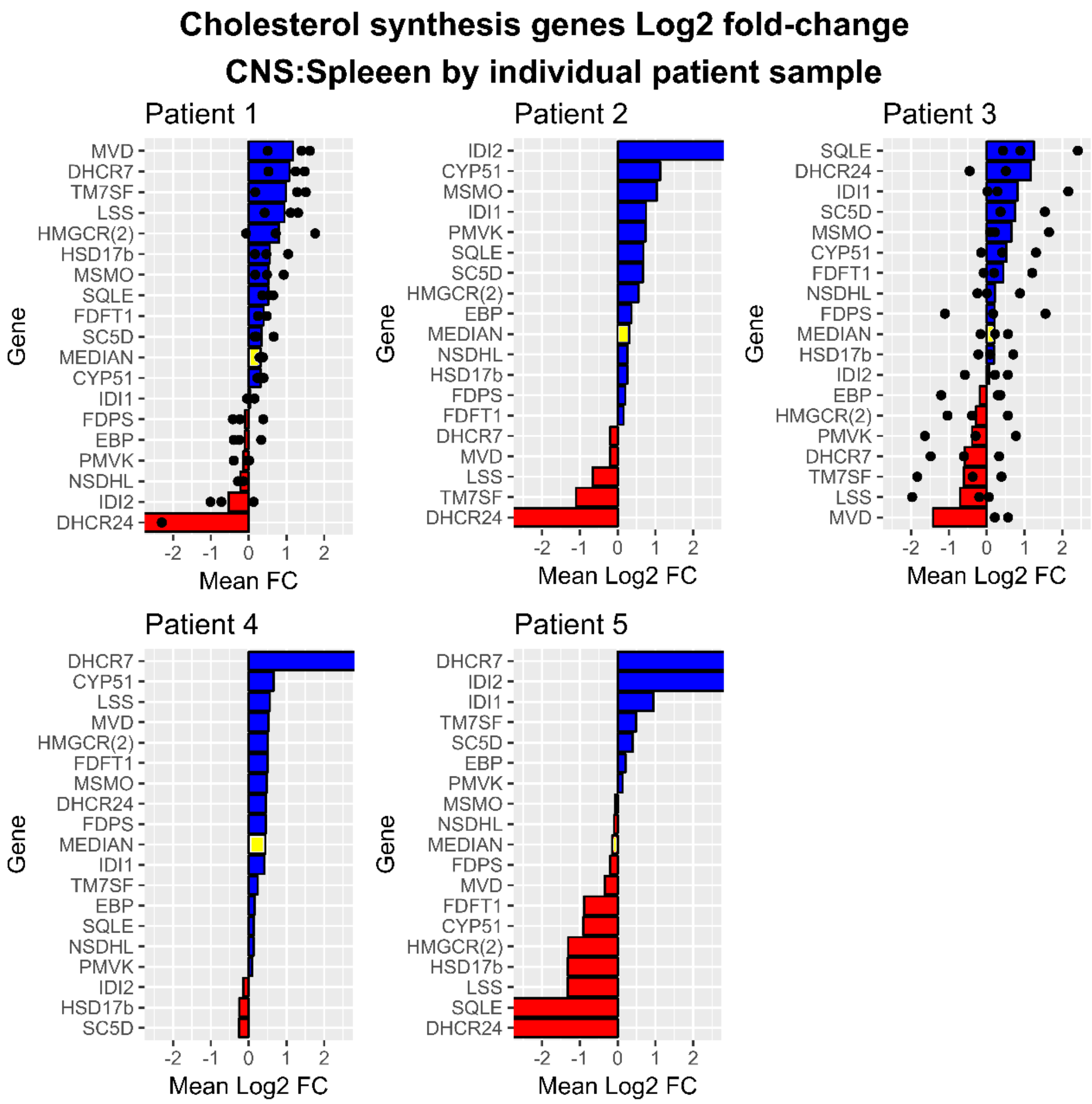


Figure 3-10 Waterfall plot of cholesterol synthesis gene differential expression by Fluidigm® multiplex PCR for primary and primagraft ALL cells retrieved from murine CNS and Spleen. Patients 1,2,3: t(12;21) ALL; Patients 4,5: t(4;11) ALL. Ordered by log2 fold-change CNS vs Spleen. Blue colour denotes upregulation; red colour denotes downregulation. Median Log2 fold change for cholesterol synthesis genes denoted by yellow colour. Dots represent results for individual mice in experiments with multiple mice included (individual data points beyond the limits of the graph are not shown). Note no statistical analysis performed on these data.

3.3.2 Analysis of lipid metabolism

3.3.2.1 SREBF and ACSS gene expression

As detailed elsewhere (section 1.2.4.3) the main regulatory system of cellular cholesterol and lipid metabolism revolves around sterol-responsive-element binding proteins (SREBPs); coded for with the SREBF1 and SREBF2 genes(Hua et al. 1993). SREBF1 has two transcriptional isoforms –SREBF1a and SREBF1c(Hua et al. 1995). These 3 SRBEBP proteins (SREBP-1a, SREBP-1c, and SREBP2) have subtly different roles: SREBP-1a is involved in both fatty acid and cholesterol synthesis, SREBP-1c is involved mainly in fatty acid synthesis, and SREBP2 is involved mainly in cholesterol synthesis(Horton et al. 2002).

The two ACCS genes (ACSS1 and ACSS2) convert acetate to acetyl-CoA which, among other roles, is a key lipid precursor metabolite. Given the potential importance of acetate as a metabolic precursor in cancer, particularly in nutrient poor environments(Schug et al. 2015) it was hypothesised these genes were likely to be upregulated in the CNS. There was not, however, evidence of increased ACSS expression in our RNASeq data, and data from this multiplex PCR experiment did not find evidence of upregulation (Figure 3-11).

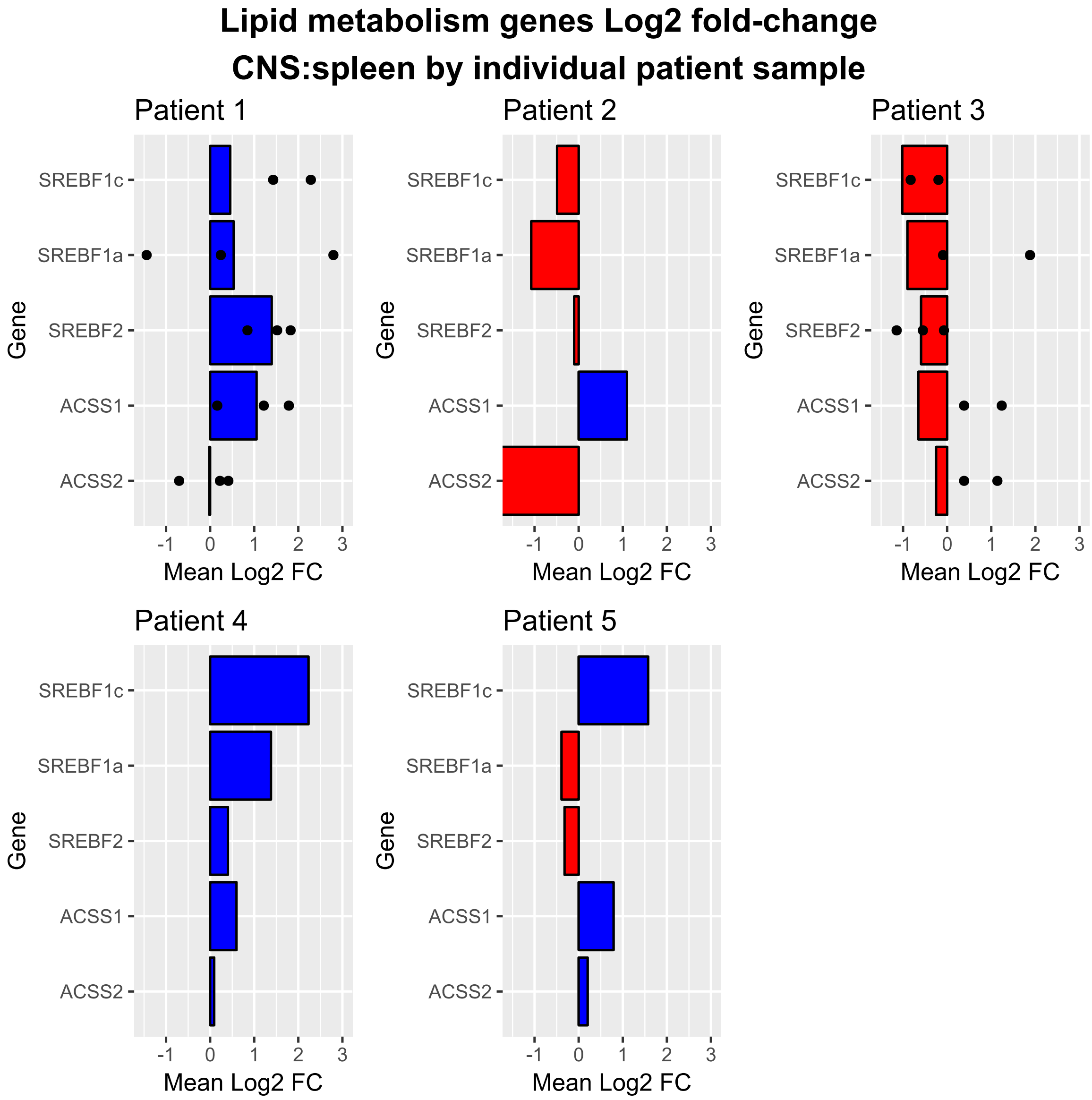


Figure 3-11 Waterfall plot of lipid metabolism gene differential expression by Fluidigm® multiplex PCR for primary and primagraft ALL cells retrieved from murine CNS and Spleen. Ordered by log2 fold-change CNS vs Spleen. Blue colour denotes upregulation; red colour denoted downregulation. Dots represent results for individual mice in experiments with multiple mice included (some individual data points are beyond the limits of the graph and not shown). There were no statistically significant differences between CNS and Spleen.

3.3.3 Fatty acid metabolism

In addition to cholesterol metabolism, this experiment provided data regarding fatty acid metabolism, showing two key genes (SCD and CPT1a) are significantly differentially regulated between the CNS and the spleen. SCD is upregulated and CPT1a downregulated in the CNS compared with the spleen (Figure 3-12). As noted previously (section 1.2.3) SCD upregulation has been shown to be associated with CNS relapse in ALL (van der Velden et al. 2015).

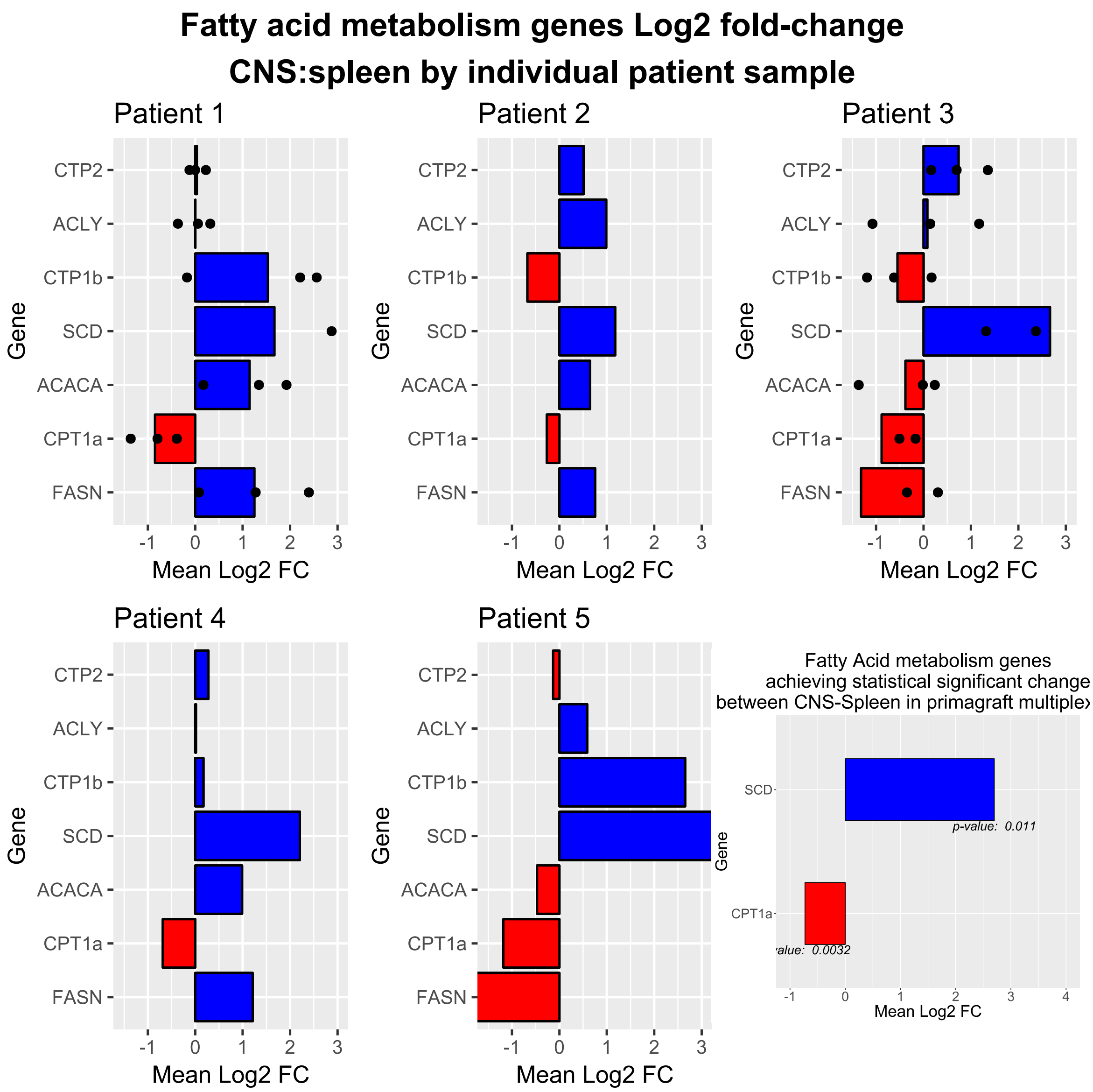


Figure 3-12 Waterfall plot of fatty acid metabolism gene differential expression by Fluidigm® multiplex PCR for primary and primagraft ALL cells retrieved from murine CNS and Spleen. Ordered by log2 fold-change CNS vs Spleen. Blue colour denotes upregulation; red colour denoted downregulation. Dots represent results for individual mice in experiments with multiple mice included (some individual data points are beyond the limits of the graph and not shown)

3.3.4 Analysis of other candidate genes in CNS ALL

On examination of the data for other putative genes differentially expressed in CNS ALL from this Fluidigm® multiplex PCR experiment, there is confirmation of some previously described findings. In particular, in all 5 patients VEGFα was detected in the CNS but not the spleen in keeping with published data. Interferon regulatory factor 4 was also upregulated in all 5 of our patient samples. The other factors tested (osteopontin/SPP1, IL-15, ICAM1, and MERTK) did not show any consistent pattern (Figure 3-13).

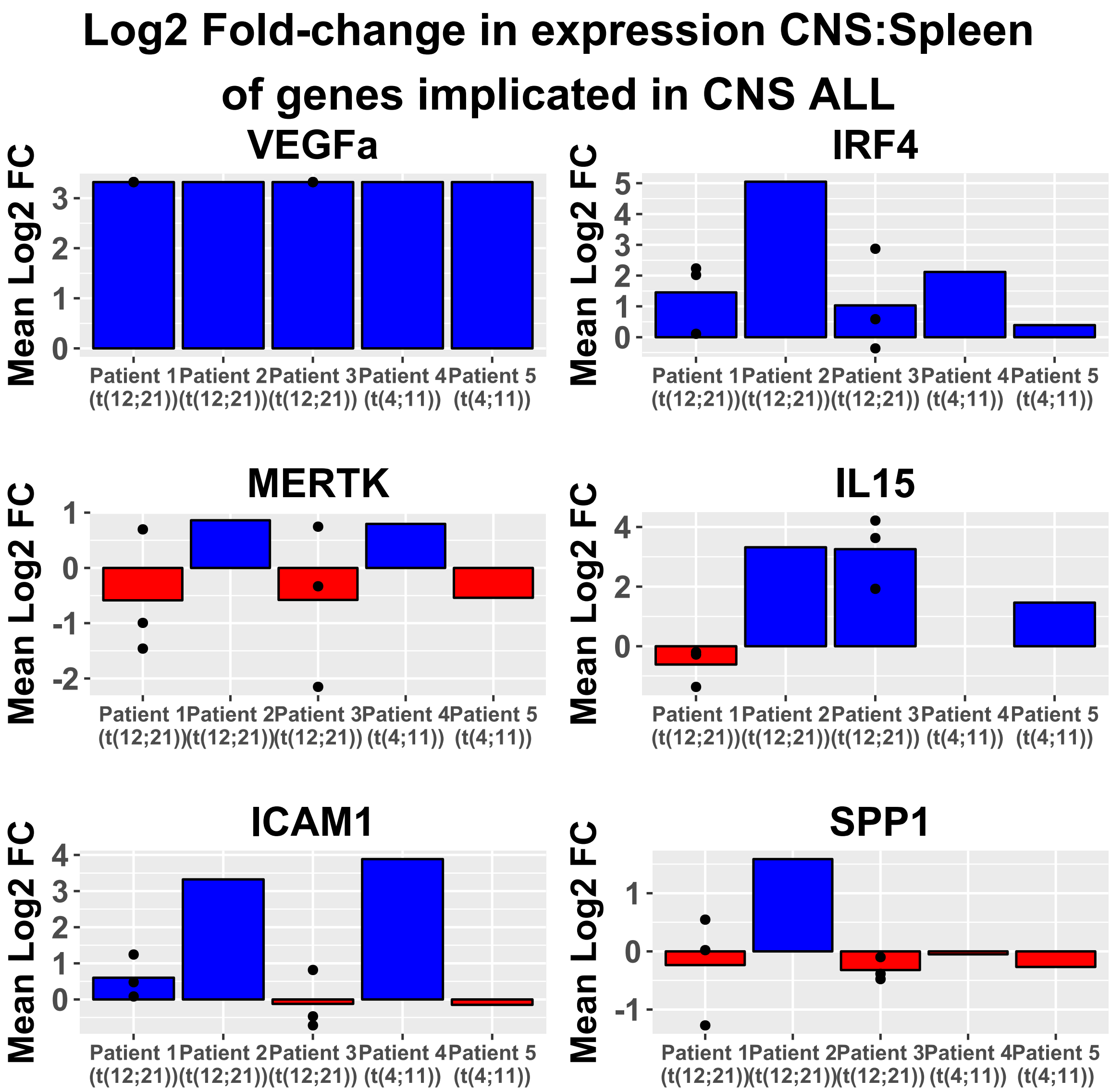


Figure 3-13 Waterfall plot of putative "ALL CNS-phenotype" gene differential expression by Fluidigm® multiplex PCR for primary and primagraft ALL cells retrieved from murine CNS and Spleen. Blue colour denotes upregulation; red colour denoted downregulation. Dots represent results for individual mice in experiments with multiple mice included (individual data points beyond the limits of the graph not shown). Note VEGFα bars shown do not show extent of fold-change – RNA was found in CNS but not spleen so fold-change is infinite. IRF4 showed a statistically significant increase in abundance in CNS compared to spleen (p=0.018). IL15 not detected in CNS or spleen in Patient 4.

3.3.5 Summary

Multiplex PCR of cholesterol synthesis genes, selected lipid metabolism genes and genes previously proposed to be important for CNS ALL was carried out on RNA from primary or primagraft ALL cells retrieved from the CNS vs spleens of mice. This showed upregulated cholesterol synthesis in cells from the CNS for 4/5 cells used, together with evidence supporting changes in lipid metabolism in cells in the CNS (particularly upregulation of SCD and downregulation of CPT1α). In addition these data confirms upregulation of VEGFα in cells retrieved from the CNS, as well as upregulation of IRF4. Other genes that have previously been suggested to be important for CNS ALL were not shown to have a consistent expression profile in the CNS vs the spleen in this experiment.

3.4 Validation of RNASeq in publicly available primary CNS ALL data

The final validation of the RNASeq data was comparison of our findings with publicly available microarray transcriptomic data from 8 children with ALL at CNS relapse (cells retrieved from the CSF) compared with cells from the bone marrow at diagnosis and bone marrow relapse(van der Velden et al. 2015). As mentioned earlier this method of obtaining cells for transcriptional analysis has significant challenges, but nevertheless provides possibly the closest data to the true transcriptional profile of human CNS ALL cells.

Data were retrieved from the “GEO Data Sets” NCBI database GSE60926 and analysed for differential expression of cholesterol synthesis genes by sample clustering (Figure 3-14) and Gene Set Enrichment Analysis (Figure 3-15). Entirely in keeping with our RNASeq data, there was significant enrichment of cholesterol synthesis in cells retrieved from the CNS at relapse compared with bone marrow either at diagnosis or BM relapse. There were no significanct differences in cholesterol biosynthesis between samples from the bone marrow at diagnosis vs BM relapse.

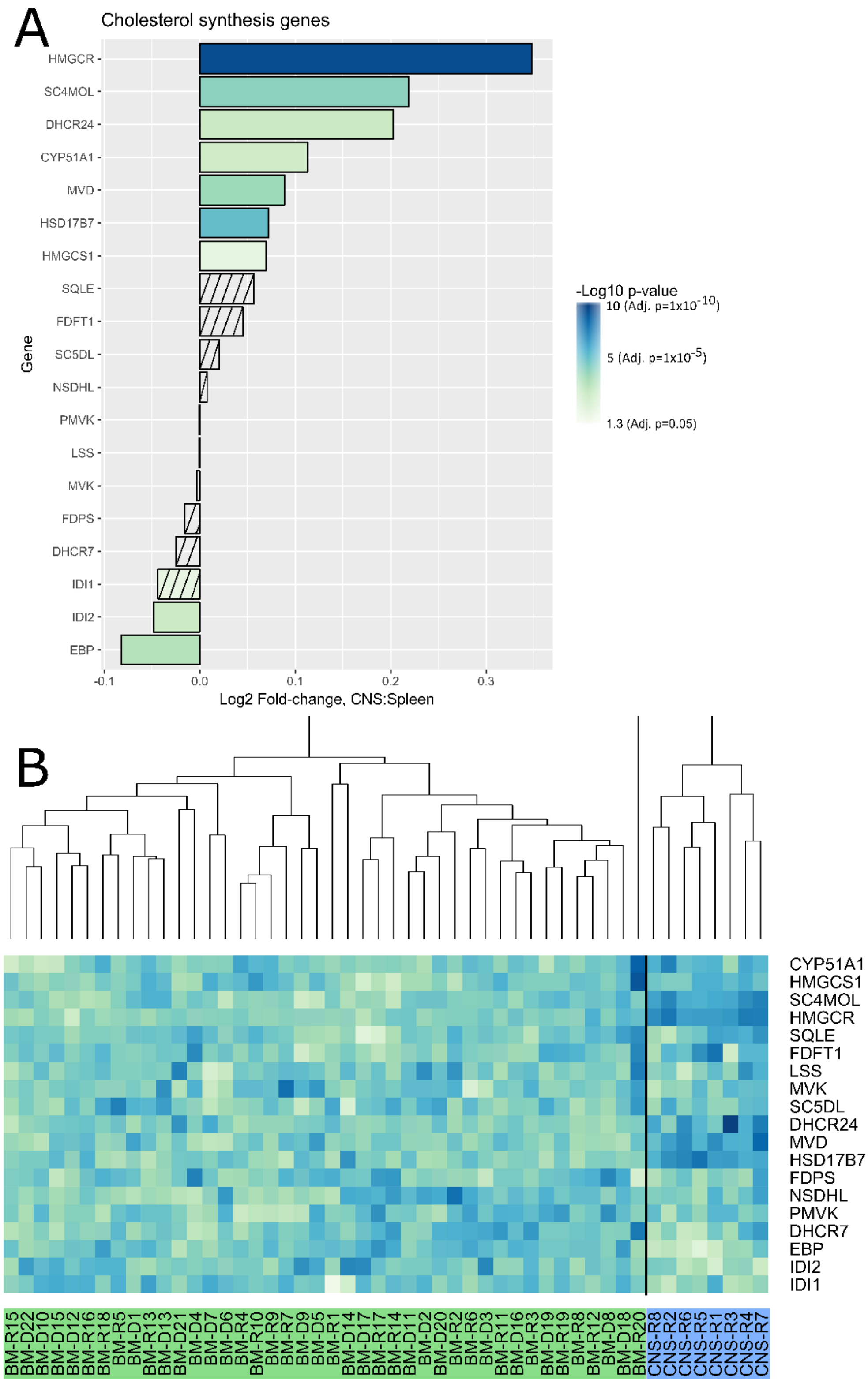
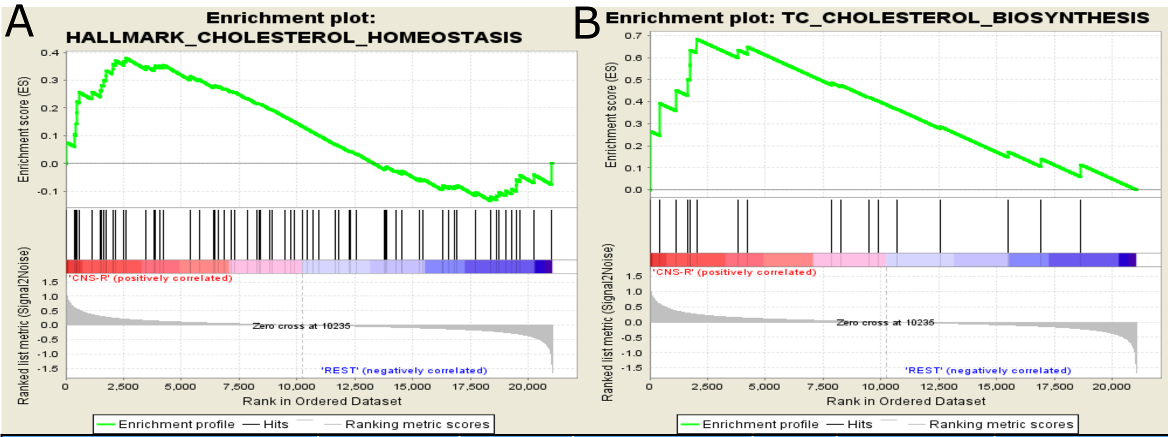


Figure 3-14 Cholesterol gene waterfall chart and Heatmap of transcriptomic data from human ALL cells retrieved from CNS at CNS-relapse (CNS-R) vs BM at diagnosis (BM-D) or relapse (BM-R). A -Waterfall plot of cholesterol synthesis gene differential expression between CNS and BM samples. Ordered by log2 fold-change CNS vs Spleen; -log10 p-value (student's t-test) denoted by colour, genes without statistically significant differential expression denoted by diagonal lines; B - Heatmap and dendrogram of cholesterol synthesis genes across all samples. CNS samples denoted in blue, BM samples in green.



C Name of Geneset	No. of genes in GeneSet	Enrichment Score	Normalised Enrichment Score	Nominal p-value	False Discovery Rate q-value	Familywise Error Rate p-value
hallmark allograft rejection	200	0.526	2.141	0	0	0
Tc cholesterol biosynthesis	18	0.684	1.771	0.004	0.006	0.013
Hallmark interferon gamma response	199	0.397	1.622	0	0.019	0.063
Hallmark complement	200	0.397	1.591	0.002	0.021	0.089
Hallmark interferon alpha response	97	0.425	1.551	0.002	0.024	0.125
Hallmark inflammatory response	200	0.345	1.396	0.007	0.085	0.435
Hallmark KRAS signaling up	199	0.339	1.368	0.016	0.096	0.522
Hallmark angiogenesis	36	0.441	1.367	0.08	0.084	0.526
Hallmark UV response DN	142	0.347	1.346	0.012	0.0916	0.603
Hallmark cholesterol homeostasis	73	0.38	1.337	0.063	0.09	0.636

Figure 3-15 GeneSet Enrichment Analysis of transcriptomic data from human ALL cells retrieved from CNS at CNS-relapse vs BM and diagnosis or relapse. A - Geneset Enrichment Analysis Enrichment plot for the "Hallmark" cholesterol homeostasis pathway; B - Geneset Enrichment Analysis Enrichment plot for the bespoke "TC cholesterol synthesis" pathway; C – Geneset Enrichment Analysis table of the top 10 differentially expressed pathways using the MSigDB "Hallmark" pathways database and bespoke "TC" cholesterol synthesis pathways

3.4.1 Summary

Analysis of primary gene expression data from ALL cells retrieved from the CNS at CNS relapse of ALL have confirmed upregulation of cholesterol biosynthesis in the CNS compared (non-paired) to the bone marrow at diagnosis and relapse.

3.5 Exploration of cholesterol synthesis upregulation as a marker of risk of CNS relapse

One of the key recent discoveries in the field of CNS ALL research was the detection of cells with specific markers that, if present in diagnostic bone marrow, predicted for future CNS relapse(van der Velden et al. 2015). Given the compelling evidence for enrichment of cholesterol synthesis gene transcription in the CNS, the possibility that increased levels of cholesterol synthesis genes in the bone marrow at diagnosis may predict CNS relapse was explored.

Publicly available data, with clinical annotation including survival and bone marrow/CNS relapse, for children with high-risk ALL treated on the Children's Oncology Group (COG) P9906 trial was obtained from the US National Cancer Institute TARGET phase 1 acute lymphoblastic leukaemia project and NCBI "Geo DataSets" GSE GSE11877.

Briefly, in the P9906 trial patients between the ages 1-22 years with high-risk leukaemia were recruited. High-risk disease was defined by any of:

- Age- and Sex-stratified WCC (Shuster criteria(Borowitz et al. 2003))
- CNS 3 status
- MLL positivity
- Testicular involvement (n=3, no CNS relapse)

Patients with Philadelphia chromosome or hypodiploidy were excluded. Patients with TEL-AML1, or trisomy 4 and 10, were excluded in the absence of CNS3 or testicular leukaemia(Bowman et al. 2011).

Data from 207 children were available in the form of microarray of ALL cells (from BM 131/207, peripheral blood 76/207). These were matched with clinical data from the TARGET database. A cholesterol signature was sought using a z-score approach The method employed is available above (section 0). Data were transformed to ensure normal distribution (Figure 3-16), In this study the cutoffs chosen were: 1.2 (approximately the top 10% expression levels for each gene), 1.5 (approximately top 5% of expression levels), and 2 (approximately top 2.5% expression levels). For simple two-group analyses, samples with 2+ genes upregulated with a z-score ≥ 1.5 were considered “cholesterol synthesis upregulated”.

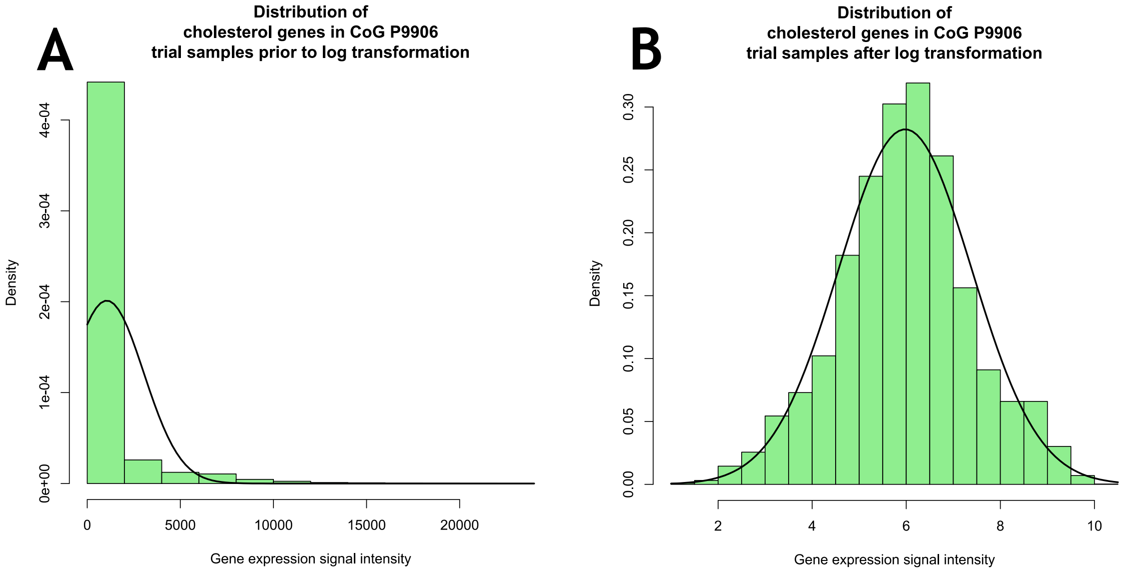


Figure 3-16 Histogram of distribution of cholesterol synthesis gene expression values across samples p9906 trial with overlaid normal distribution curve. A – before log transformation; A – after log transformation.

3.5.1 Correlations

The first step in this analysis was looking to see how a cholesterol biosynthesis signal was associated with other clinical factors linked to risk of CNS relapse: age at diagnosis, white cell count at diagnosis, MLL-translocation status and CNS-status (determined by whether there are red cells and/or leukaemic blasts and how many blasts are present in the CSF at diagnosis); and if the signal was associated with clinical outcomes. This was done by using the criteria above (2+ genes above a cut-off z-score of 1.5) to split the data set unto “cholesterol synthesis upregulated” and “cholesterol synthesis not upregulated” groups. Statistical analysis was carried out using a chi-squared test. The cholesterol gene signal was correlated with MLL status but not any other traditional risk factor for CNS relapse (Table 3-4). In terms of outcome, upregulated cholesterol synthesis correlated with CNS relapse (predominantly isolated CNS relapse) but not bone marrow (BM) relapse (Table 3-5).

Table 3-4 Correlation of upregulated cholesterol synthesis (defined as 2+ genes upregulated with a cutoff z-score of 1.5) with traditional risk factors for CNS relapse in childhood ALL. p-value calculated with chi-squared test.

Risk factor /Outcome	Number with cholesterol synthesis upregulated	% with cholesterol synthesis upregulated	Number without cholesterol synthesis upregulated	% without cholesterol synthesis upregulated	p-value
Total:	44	21%	163	79%	
WCC at diagnosis ≥ 50 $x10^9$/L	23	22%	82	78%	0.8
Age at diagnosis ≥ 10 yrs	23	18%	108	82%	0.09
MLL positive	10	50%	10	50%	0.0009
CNS 3 status	4	19%	17	81%	0.8

Table 3-5 Correlation of upregulated cholesterol synthesis (defined as 2+ genes upregulated with a cutoff z-score of 1.5) with death and relapse outcome in childhood ALL. p-value calculated with chi-squared test.

Risk factor /Outcome	Number with cholesterol synthesis upregulated	% with cholesterol synthesis upregulated	Number without cholesterol synthesis upregulated	% without cholesterol synthesis upregulated	p-value
Total:	44	21%	163	79%	
Death	16	31%	36	69%	0.07
Isolated BM relapse	10	26%	31	74%	0.4
Any BM relapse	10	23%	36	77%	0.7
Any CNS relapse	11	37%	19	63%	0.026
Isolated CNS relapse	11	44%	14	56%	0.003

Of note, MLL status in this dataset does not correlate with CNS relapse (p=0.54 univariate LogRank test) though numbers are small.

3.5.2 CNS relapse risk analysis

With this crude analysis showing an apparent correlation between samples having a cholesterol synthesis signal and CNS relapse, the next stage was to perform detailed survival analysis to assess whether this correlation persists in the presence of sample censoring, and how it develops over time.

To do this a Kaplan-Meier curve was created and statistical analysis performed using univariate LogRank test. This was done initially using the two groups above (Figure 3-17), showing that upregulation of cholesterol synthesis was strongly associated with CNS relapse (p<0.0001).

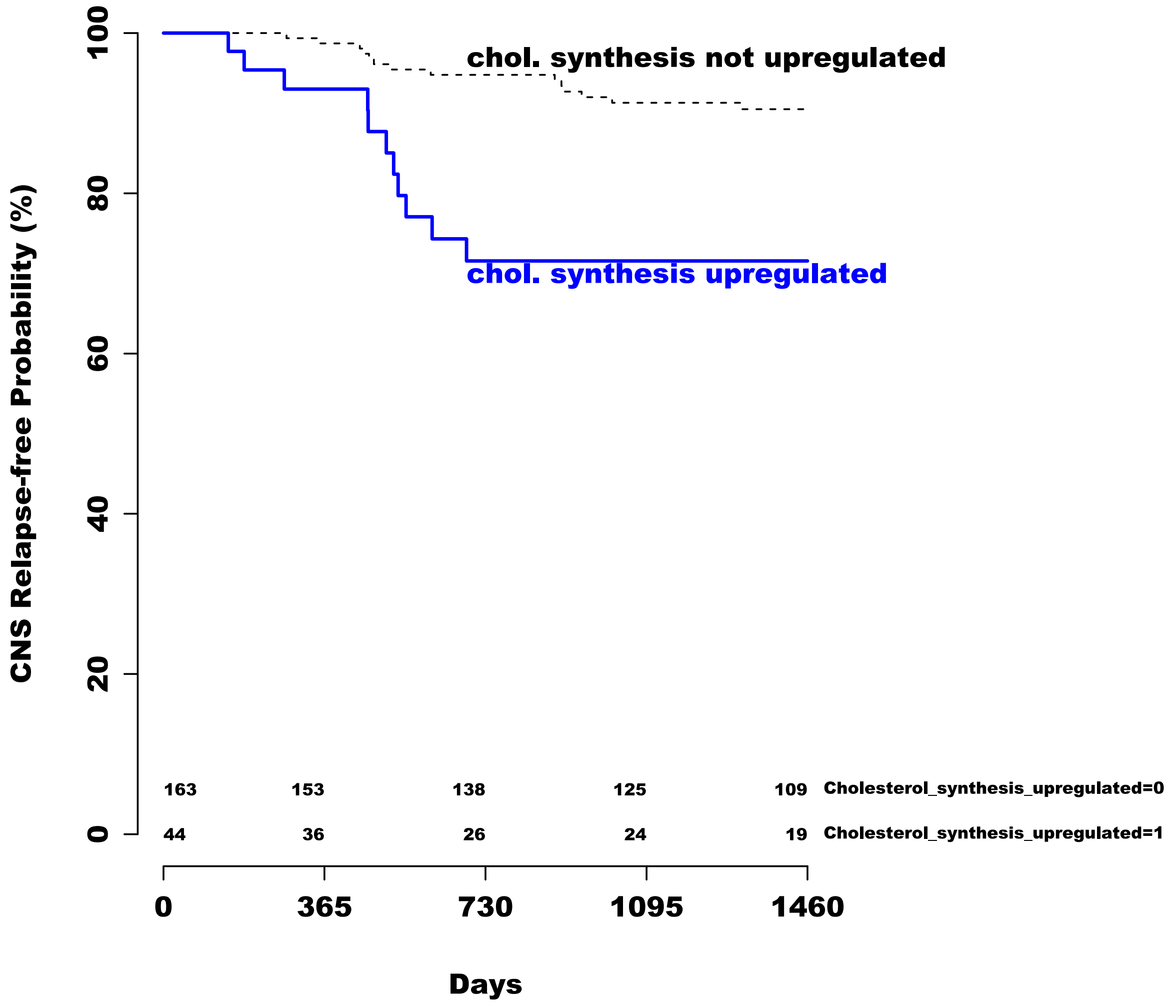


Figure 3-17 Kaplan-Meier survival curve for CNS relapse-free probability in COG P9906 trial by upregulation of cholesterol synthesis (defined as z-score ≥ 1.5 in 2 or more genes in cholesterol synthesis pathway). Number of patients included in analysis at each timepoint in each group marked along x axis. p<0.0001 by univariate LogRank test.

Next, to ensure this was a true cholesterol biosynthesis signal, Kaplan-Meier curves were created using multiple groups corresponding to number of genes with z-scores above each of the cut-offs used. The increased risk of CNS relapse consistently increased as the number of upregulated genes increased, and as z-score was increased (p<0.0001) (Figure 3-18).

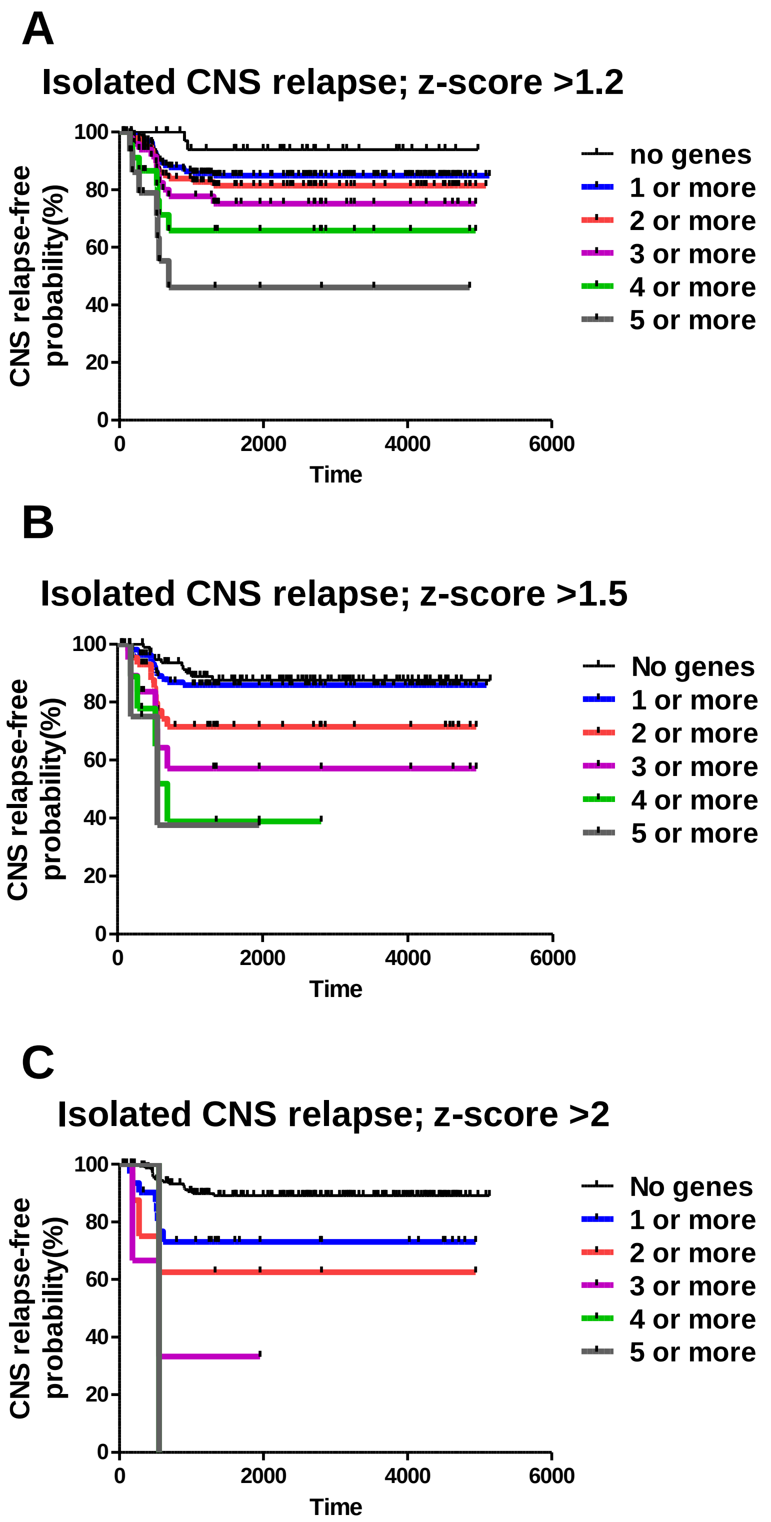


Figure 3-18 Kaplan-Meier survival curve for CNS relapse-free probability in P9906 data by upregulation of cholesterol synthesis. A - z-score ≥ 1.2; B – z-score >1.5; C – z-score >2; for cumulative increase in number of genes upregulated for each z-score. All analyses show significant differences between groups (LogRank test, p<0.0001).

3.5.3 Multivariate analysis

To confirm that this apparent increase risk is truly related to the cholesterol synthesis signal multivariate analysis with traditional risk factors of CNS relapse (section 3.5.1) was performed.

This was done by construction of a Cox Proportional Hazards model. As before, a simple two-group analysis was performed (Table 3-6), and analysis of a range of number of genes upregulated for a particular z-score – 1.2 (Table 3-7) and 2 (Table 3-8). For all of these analyses, upregulation of cholesterol synthesis retains high statistical significance for correlation with CNS relapse when analysed alongside the “high-risk” factors of the patients in the study. One small caveat worth noting is that the study protocol was changed early in the trial to increase CNS-directed therapy for patients with high WCC due to an early excess of CNS relapses in these children(Bowman et al. 2011).

Table 3-6 Table showing multivariate analysis of CNS relapse risk of traditional risk factors and cholesterol synthesis upregulation (defined as z-score ≥ 1.5 in 2 or more genes in cholesterol synthesis pathway)

	HR (95% CI)	p-value
Cholesterol synthesis upregulated (2genes; z-s ≥ 1.5)	3.33 (1.48-7.5)	0.00369
Day 29 MRD >0.01	1.23 (0.5-3.03)	0.65079
WCC at diagnosis ≥ 50 x10^9/L	1.29 (0.5-3.35)	0.59935
CNS status 3	0.71 (0.15-3.26)	0.65871
Age at diagnosis ≥ 10 yrs	0.45 (0.18-1.17)	0.10272
MLL positive	0.8 (0.21-3.01)	0.74187

Table 3-7 Table showing multivariate analysis of CNS relapse risk of traditional risk factors and cholesterol synthesis upregulation (defined as a continuous factor, the number of genes upregulated with z-score cutoff ≥ 1.2)

	HR (95% CI)	p
Number of cholesterol genes upregulated z-score ≥ 1.2	1.53 (1.21-1.94)	0.00037
Day 29 MRD >0.01	1.04 (0.42-2.59)	0.92801
WCC (diagnosis) ≥ 50 x10^9/L	1.27 (0.51-3.18)	0.6121
CNS status 3	0.66 (0.14-3.05)	0.59902
Age (diagnosis) ≥ 10 yrs	0.46 (0.19-1.12)	0.08719
MLL positive	0.64 (0.16-2.51)	0.52396

Table 3-8 Table showing multivariate analysis of CNS relapse risk of traditional risk factors and cholesterol synthesis upregulation (defined as a continuous factor, the number of genes upregulated with z-score cutoff ≥ 2)

	HR (95% CI)	p-value
Number of cholesterol genes upregulated z-score ≥ 2	2.25 (1.36-3.72)	0.00156
Day 29 MRD >0.01	1.12 (0.45-2.79)	0.815
WCC (diagnosis) ≥ 50 $x10^9/L$	1.9 (0.69-5.21)	0.21067
CNS status 3	0.63 (0.13-2.96)	0.55809
Age (diagnosis) ≥ 10 yrs	0.48 (0.19-1.2)	0.11771
MLL positive	0.33 (0.05-2.16)	0.2474

3.5.4 Bone marrow relapse

Having shown this correlation with CNS relapse, the next stage was to look for a similar correlation with (more common) bone marrow relapse. This was done in the same way: by constructing Kaplan -Meier curves and statistical analyses with univariate LogRank test (Figure 3-19).

In these data, upregulation of cholesterol synthesis does not correlate with bone marrow relapse (p=0.79(z-score 1.2); p=0.1(z-score 1.5); p=0.95(z-score 2).

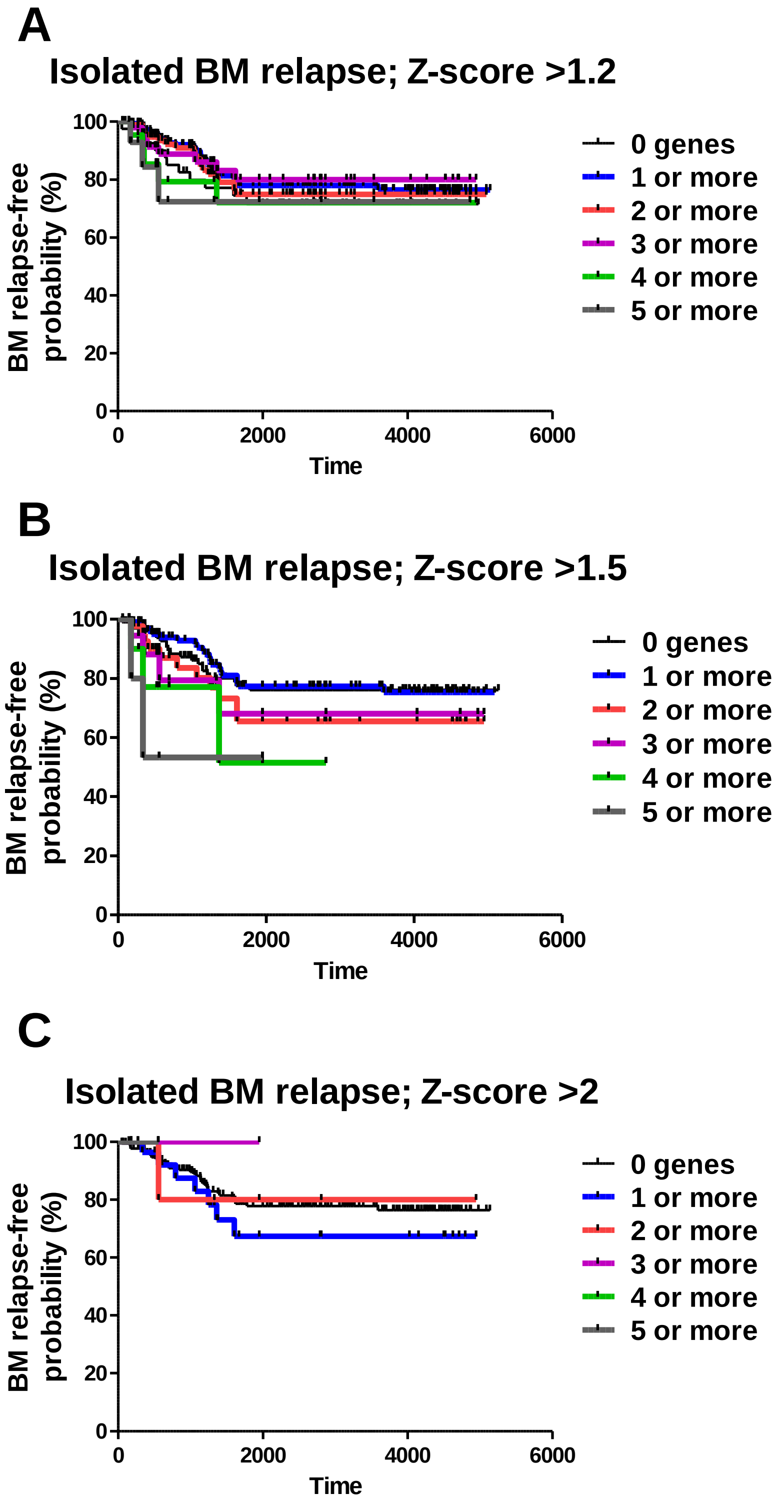


Figure 3-19 Kaplan-Meier survival curve for BM relapse-free survival in COG9906 data by upregulation of cholesterol synthesis. A - z-score ≥ 1.2; B – z-score >1.5; C – z-score >2; for cumulative increase in number of genes upregulated for each z-score. Note no significant difference between groups found for any of these z-score cutoffs.

3.5.5 Overall survival correlates with upregulated cholesterol synthesis

Finally, to complete this analysis the data were analysed for a correlation between a cholesterol biosynthesis signal and overall survival. Again, Kaplan-Meier curves were created and data analysed with univariate LogRank tests (Figure 3-20).On analysis using a single cutoff of gene number for each z-score (5 for the less highly upregulated z-score of 1.2, 2 for z-score 1.5, and 1 for z-score 2), there is a statistically significant difference in overall survival with a z-score cut-offs of 1.2 (p=0.0214 by LogRank test) and 1.5 (p=0.0319), but not 2 (p=0.288).

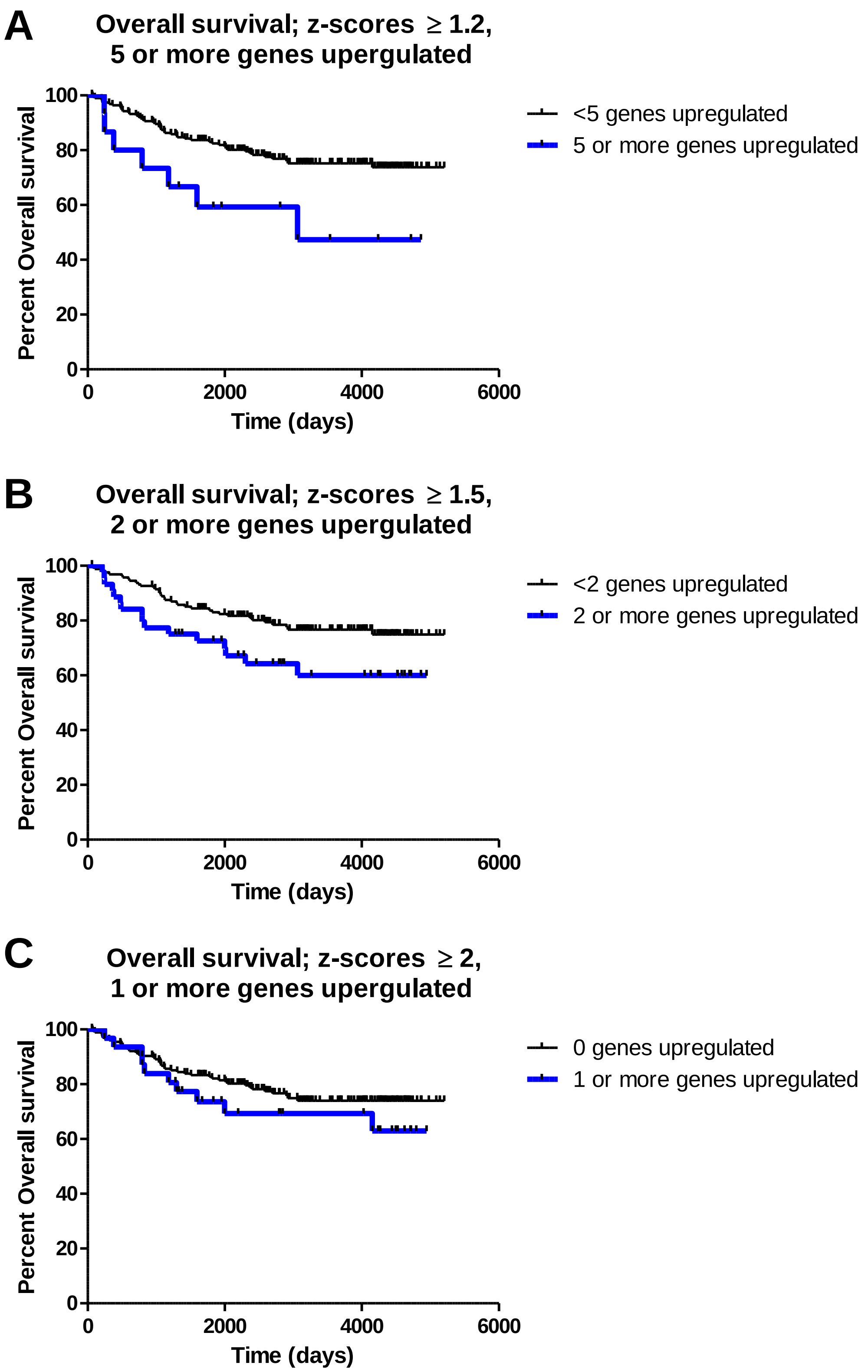


Figure 3-20 Kaplan-Meier survival curve for overall survival in P9906 data by upregulation of cholesterol synthesis. A - z-score ≥ 1.2, ≥5 genes upregulated (LogRank p-value=0.0214); B – z-score >1.5, ≥2 genes upregulated (LogRank p-value=0.0319); C – z-score >2, ≥1 gene upregulated (no statistically significant differences, LogRank p-value=0.288).

3.5.6 Summary

Analysis of publicly available gene expression data from bone marrow/peripheral blood ALL cells at diagnosis has shown that increased cholesterol biosynthesis specifically increased risk of CNS, but not BM relapse. This effect increased as the degree of upregulation and number of upregulated genes increased. The effect appears independent of traditional risk factors, though there was an association with MLL rearrangement. This translated into a significant reduction in overall survival for children with upregulated cholesterol biosynthesis in ALL cells at diagnosis.

3.6 IL7rα gene expression

More recently, IL7Rα gene expression (coding for part of the interleukin-7 receptor) has been shown to be correlated with CNS relapse in ALL (manuscript under revision). As this only came to light after this multiplex PCR experiment this gene was not included. However, on analysis of the our xenograft model RNASeq data (chapter 3.2), the GSE60926 microarray data from cells obtained from the CSF of children with CNS relapse of ALL (section 3.3.5) and analysis of the TARGET data for association with CNS relapse risk (section 3.4.1), we have confirmed elevated levels of IL7Rα gene expression are found in ALL cells in the CNS compared to systemic disease (Figure 3-21), and increased expression of IL7Rα in the bone marrow at diagnosis – determined, as above (section 0) using z-score, with a cut-off z-score of 1.2 chosen – was associated with increased risk of isolated CNS relapse in univariate (Figure 3-22) and multivariate analysis (Table 3-9). Interestingly, in the TARGET data, increased IL7Rα expression was correlated with the t(1;19) TCFR3-PBX1 translocation (Table 3-10). This fits with the initial studies and may help explain the increased risk of CNS relapse seen in children with this translocation (discussed previously in section 1.2.3).

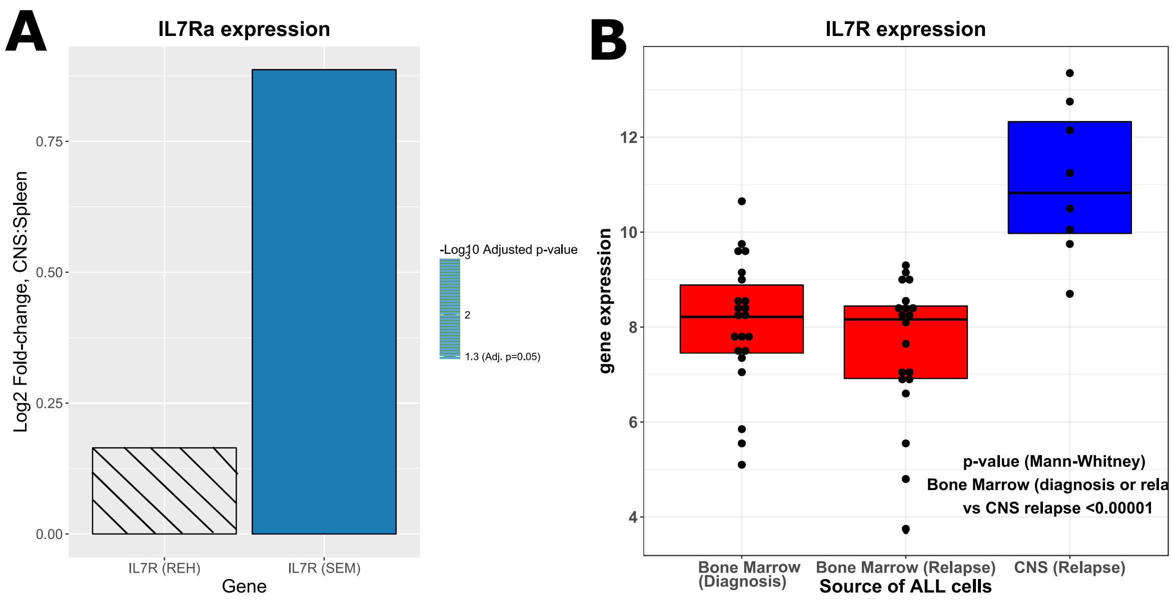


Figure 3-21 Log$_2$Fold-change in IL7R expression between: A - human ALL cells retrieved from the CNS:Spleen of NSG mice; B- ALL cells retriived from the CNS or bone marrow of children with ALL.

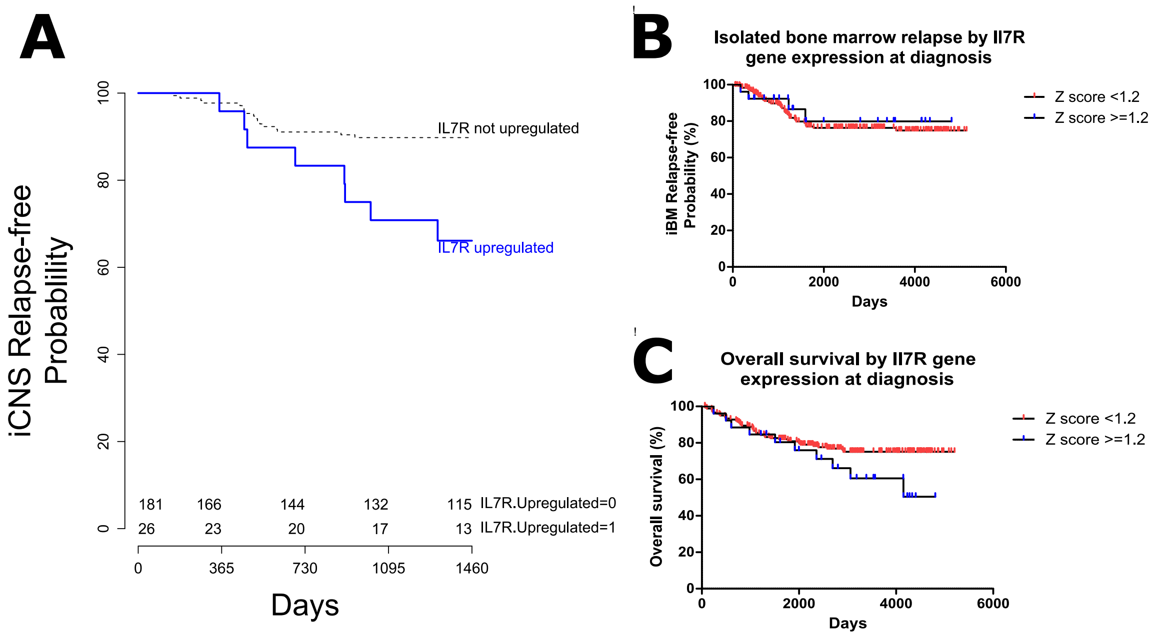


Figure 3-22 Kaplan-Meier curves for CNS relapse rate by upregulated IL7R expression in bone marrow ALL cells at diagnosis. A – isolated CNS relapse-free probability; B – isolated BM relapse-free probability; C – Overall survival.

Table 3-9 Multivariate analysis of impact of IL7R and cholesterol synthesis gene upregulation on CNS relapse risk in the TARGET dataset

Risk Factor	HR, (95% CI)	P-value
IL7R upregulation	4.74 (1.82,-,12.32)	0.00143**
Cholesterol synthesis upregulated	4.75 (1.99,-,11.33)	0.00044***
Day 29 MRD >0.01	1.19 (0.49,-,2.89)	0.69317
WCC (diagnosis) ≥ 50 x10^9/L	1 (0.37,-,2.68)	0.99764
CNS status 3	0.8 (0.17,-,3.84)	0.78291
Age (diagnosis) ≥ 10 yrs	0.47 (0.19,-,1.21)	0.11762
MLL Rearrangement	1.04 (0.28,-,3.89)	0.95782

Table 3-10 Correlation of cytogenetic subtypes with upregulated IL7R expression in the TARGET dataset

Cytogenetic subtype	Number with ILR7 upregulated	%	Number without ILR7 upregulated	%	p-value
Total	26	13%	181	87%	
MLL	2	10%	18	90%	0.7162
ETV6/RUNX1 Fusion	0	0%	4	100%	0.4440
TRISOMY 4 or 10	2	29%	5	71%	0.1934
TCF3 PBX1	7	30%	16	70%	0.0060

3.6.1 Summary

IL7R expression is increased in ALL cells retrieved from the CNS of NSG xenograft models and from human CSF at CNS relapse of ALL. Increased IL7R expression at diagnosis is associated with an increased risk of CNS but not BM relapse. This increased risk persists on multivariate analysis with traditional risk factors and increased cholesterol synthesis. There was an association between IL7R expression and t(1;19) TCF3-PBX1 translocation.

3.7 Conclusions

The aim of this chapter was to identify changes in metabolism of BCP-ALL cells in the CNS compared with systemic disease by analysing the transcriptome. The data presented show good evidence for metabolic adaptation in ALL cells in the CNS. These seem to focus mainly on changes in and cholesterol and lipid metabolism. These changes are in keeping with the CSF environment of these cells - low in glucose, low in oxygen and extremely low in cholesterol and lipids.

It is worth noting that a significant proportion of RNASeq data were not analysed as the reads that aligned to both the human and mouse genome and were excluded. This means there can be high confidence that the changes seen represent true findings from human ALL cells, but there may be significant changes between the CNS and spleen in genes partially conserved between humans and mice that are missed by this analysis.

Cholesterol metabolism may be a particularly interesting area of study. Cholesterol biosynthesis is consistently upregulated in the CNS in our xenograft model systems and this is confirmed in primary human samples. It appears as though it may be possible to detect a "cholesterol biosynthesis" signal in the bone marrow of children at diagnosis which correlates with a substantially increased risk of CNS relapse in a group of high-risk children, though further data would be required to assess the reproducibility of these data and how applicable it is to the wider population of children with ALL (discussed in more depth in Chapter 6:).

The data from primary and primagraft ALL cells in the NSG murine model did not as comprehensively show increased cholesterol synthesis as the experiments in the rest of this chapter, though there was an increase in cholesterol synthesis in the CNS in 4/5 samples tested. It is possible that the relatively subtle cholesterol signal was lost due to RNA degradation during storage – the cells were stored as cell pellets for up to 1 year at -80°C, partly because of the variability in time to engraftment. This is not entirely borne out by the RNA absorbance data prior to the experiment, but there was increased RNA retrieval from spleen ALL cells which may have caused the loss of difference between groups. Regardless, given the clear findings on RNASeq of two very different ALL cell lines with confirmation in cells from the CNS of 8 children, cholesterol metabolism was a very interesting area for further study and the results of further investigations *in vitro* and *in vivo* will be discussed in Chapter 5:.

Finally, analysis of these data adds supporting evidence to the upregulation of IL7Rα, IRF4 and VEGFα in ALL cells in the CNS, and raises the possibility of IL7R expression in the bone marrow at diagnosis as a risk factor for CNS relapse –

possibly correlated with the known risk factor for CNS relapse of t(1;19) TCFR3-PBX1 translocation.

Chapter 4: Metabolomic analysis of human and murine CSF in the presence and absence of CNS acute lymphoblastic leukaemia infiltration

4.1 Introduction and aims

Having examined the transcriptome of ALL cells in the CNS, the next step in this project was to look at the metabolome of CNS ALL. This was achieved through access to a biobank of CSF taken from children with ALL in Glasgow, UK, between 2009 and 2016.

As noted previously (section 1.2.2) ALL cells infiltrating the CNS lie almost exclusively in the leptomeninges, bathed in CSF. CSF is a nutrient poor environment, and in normal circumstances there are very few cells in this environment. In the context of CNS involvement with ALL however, there is a significant cellular burden. It was hypothesised that this would lead to changes in the composition of CSF reflecting increased cell metabolism within the leptomeningeal compartment that could be detected using untargeted LC-MS analysis.

This chapter focusses on detection of potential novel metabolic markers for CNS ALL in the CSF. Specific aims were:

1. To carry out metabolomic analysis of CSF, to determine whether samples taken at diagnosis with ALL can be discriminated from controls, and to identify potential markers for CNS ALL

2. To carry out metabolomic analysis of CSF and plasma from a mouse model of CNS ALL to identify similarities and differences from primary human samples, and provide additional evidence of the validity or otherwise of potential markers identified in primary samples

3. To determine the utility of potential markers to discriminate CSF samples taken at isolated CNS relapse from samples from patients with no relapse

4.2 CSF stability

The first step in this analysis was to look at metabolite stability over time. The CSF biobank is based on clinical samples. These would typically be taken in the operating theatre at the Royal Hospital for (Sick) Children, Glasgow, in the afternoon, kept a variable amount of time at room temperature in theatre, before being transported to the clinical laboratory where the samples would be stored at 4°C for usually 1-2 hours, but occasionally this may be as long as 72 hours (i.e. from Friday evening to Monday evening), before being centrifuged and supernatant stored at -80°C. There was concern that variable handling times before refrigerating and/or freezing samples might impact on sample quality for metabolomic analysis.

To assess the stability of CSF stored in this way, samples from 9 patients were stored either at room temperature for 2½ hours or on ice for: 30 minutes, 2½ hours, 24 hours or 72 hours. The samples were then centrifuged and supernatants stored at -80°C until analysis.

For analysis the samples were prepared as described above (section 2.2.6.1). Briefly, the samples were thawed and diluted 1:20 in extraction solution (methanol (50% volume): acetonitrile (30% volume): water (20% volume)). Samples were then centrifuged at 13,000 rpm at 4°C for 10 minutes and supernatants transferred to glass vials for analysis. Samples were stored at -80°C until LC-MS analysis.

These results show no obvious time-related changes in CSF metabolites overall, and no gross differences in stability between metabolites assayed (though the variation for valine was slightly higher than the other metabolites assayed). Samples in the biobank were therefore considered to have been processed in way consistent with good quality metabolomic analysis (Figure 4-1).

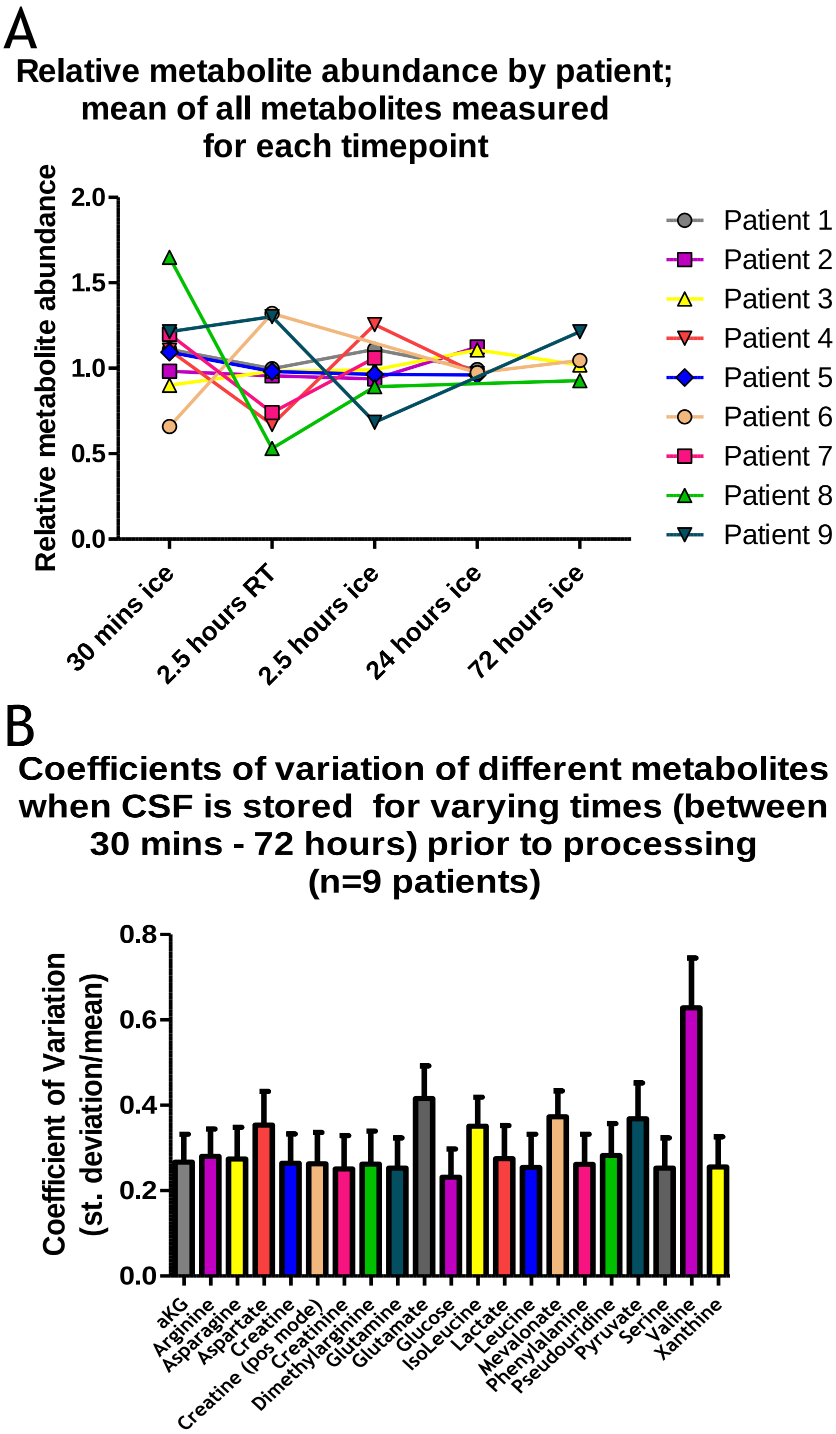


Figure 4-1 CSF stability after storage for various times (30 mins on ice, 2½ hours at room temperature, 2½, 24 or 72 hours on ice). A – Relative metabolite abundance (mean of all metabolites assayed) for each patient for each timepoint; B – Mean Coefficients of Variation across all timepoints per sample for each metabolite.

4.3 Untargeted: early vs late vs control

4.3.1 Background

It is possible to look at a large part of the metabolome using a technique called “untargeted” LC-MS. This is described in depth previously (section 2.2.6.4.This approach takes significantly more time and resources (each sample is run in positive and negative mode, at 34 minutes for each, as opposed to a single 22 minute run for a “targeted” approach), but provides better resolution of chromatography separation and a fragmentation signature – which can be compared to previously run standards and online metabolite libraries - to facilitate identification of interesting peaks. For this project we used an untargeted approach to generate candidate metabolites and pathways, and confirmed findings using targeted LC-MS.

The goal of the project was to compare CSF from children with CNS infiltration with ALL blasts to CSF from children without CNS infiltration (whether or not they remained on treatment for ALL). To do this we hypothesised (as discussed previously in section 1.2.4.1) that a significant proportion, if not all, of children have CNS infiltration at diagnosis with ALL.

We therefore compared the CSF of children at diagnosis (prior to any chemotherapy or steroid therapy) – “early” group - to CSF from the same children later in treatment (after completion of intensive chemotherapy, but while still receiving maintenance chemotherapy) –“late” group - and to control samples taken from children aged 12 and below who had CSF taken for investigation, but had no CNS pathology detected – “normal” group. Clinical characteristics for the children with ALL including a list of prescribed medications at the time of first lumbar puncture are listed previously (section 2.2.1.2).

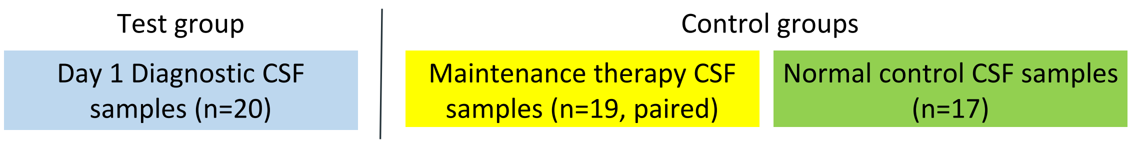


Figure 4-2 Experimental plan for untargeted metabolomic analysis of CSF samples from children at diagnosis with ALL and controls. Diagnostic samples form the "Early" goup, maintenance therapy samples the "Late" group and normal control samples the "Normal" group.

4.3.2 Initial Findings

.

Tthe data were analysed globally to look for gross differences between groups, showing clear separation by principle component analysis (Figure 4-3).

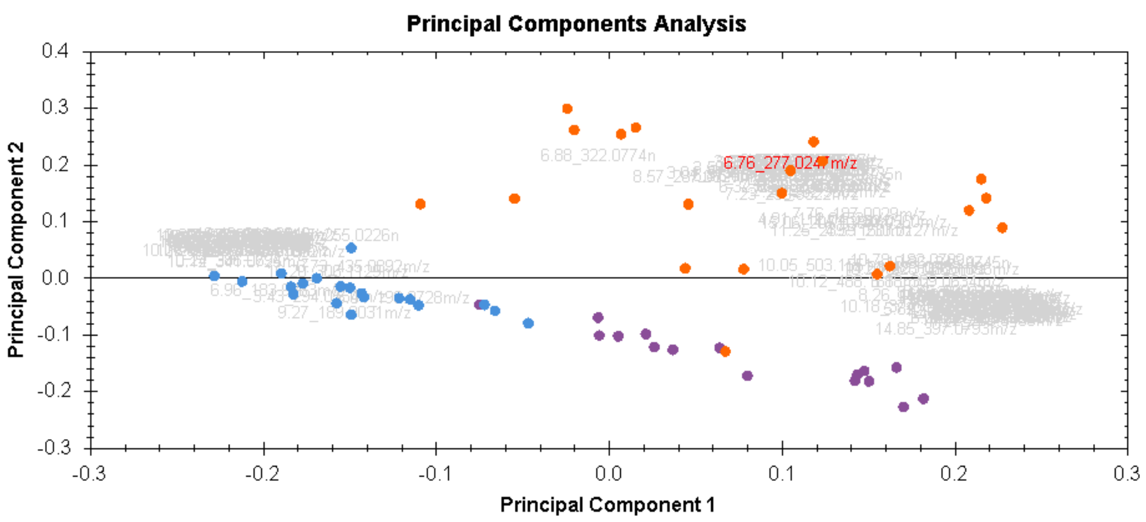


Figure 4-3 Principal Component Analysis of untargeted LC-MS data from CSF of children with and without leukaemia. Each dot represents a sample, with group denoted by colour: Light Blue – Early; Orange – Late; Purple – Normal.

4.3.3 Potential markers

The purpose of this project was to look for potential markers for CNS ALL. To do this the processed "untargeted" LC-MS data were analysed for potential metabolites with clearly higher or clearly lower concentrations in the “Early” group compared to BOTH the “Late” and the “Normal” control groups.

This was achieved with two complementary analyses. Firstly, peaks that had a ratio of the maximum (in both the control groups) to the mean of the “Early group that was less than 1.11 (i.e. the highest value in either of the control groups was less than 11% higher than the mean of the “early group), and a p-value by t-test (paired for “Early” vs “Late, unpaired for “Early” vs “Normal”) <0.05 for both groups, or peaks that had a ratio of minimum control/mean “Early” greater than 0.89 and p-values as above, were selected for further investigation **(**Equation 1**)**. Secondly, ROC curve analysis was performed using the Metaboanalyst(Xia & Wishart 2011) platform - separately for “Early” vs “Late” and “Early” vs “Normal”, and overlapping peaks were selected for further investigation. Perhaps not surprisingly, these analyses revealed very similar results (Figure 4-4, Table 4-1).

Equation 1 – Criteria for selection of potential markers for CNS ALL from metabolomic analysis of CSF

$$A\ metabolite\ was\ identified\ as\ a\ potential\ biomarker\ if:$$

$$(a)\ p - value < 0.05\ for\ Early\ vs\ Late\ AND\ Early\ vs\ Normal$$

$$AND\ either\ (b)\ or\ (c):$$

$$(b)\frac{Maximum(Late)}{Mean(Early)} < 1.11\ AND\ \frac{Maximum(Normal)}{Mean(Early)} < 1.11$$

$$OR$$

$$(c)\frac{Minumum(Late)}{Mean(Early)} > 0.89\ AND\ \frac{Minumum(Normal)}{Mean(Early)} < 0.89$$

These peaks were putatively assigned identities based on their m/z and fragmentation platform using the mzCloud™ database as described earlier (section 2.2.6.4). A literature review confirmed that while some of these targets had been described as tumour markers, and some had been associated with ALL, none had been associated with CNS disease in ALL. Of interest, using Progenesis' native metabolite identification software based on PubChem molecular masses and Human Metabolome Database(Wishart et al. 2007) and the Metlin Database(Smith et al. 2005) fragmentation spectra it was possible to identify multiple therapeutic drugs in the CSF including common antibiotics tazobactam and piperacillin, and xanthine oxidase inhibitor allopurinol.

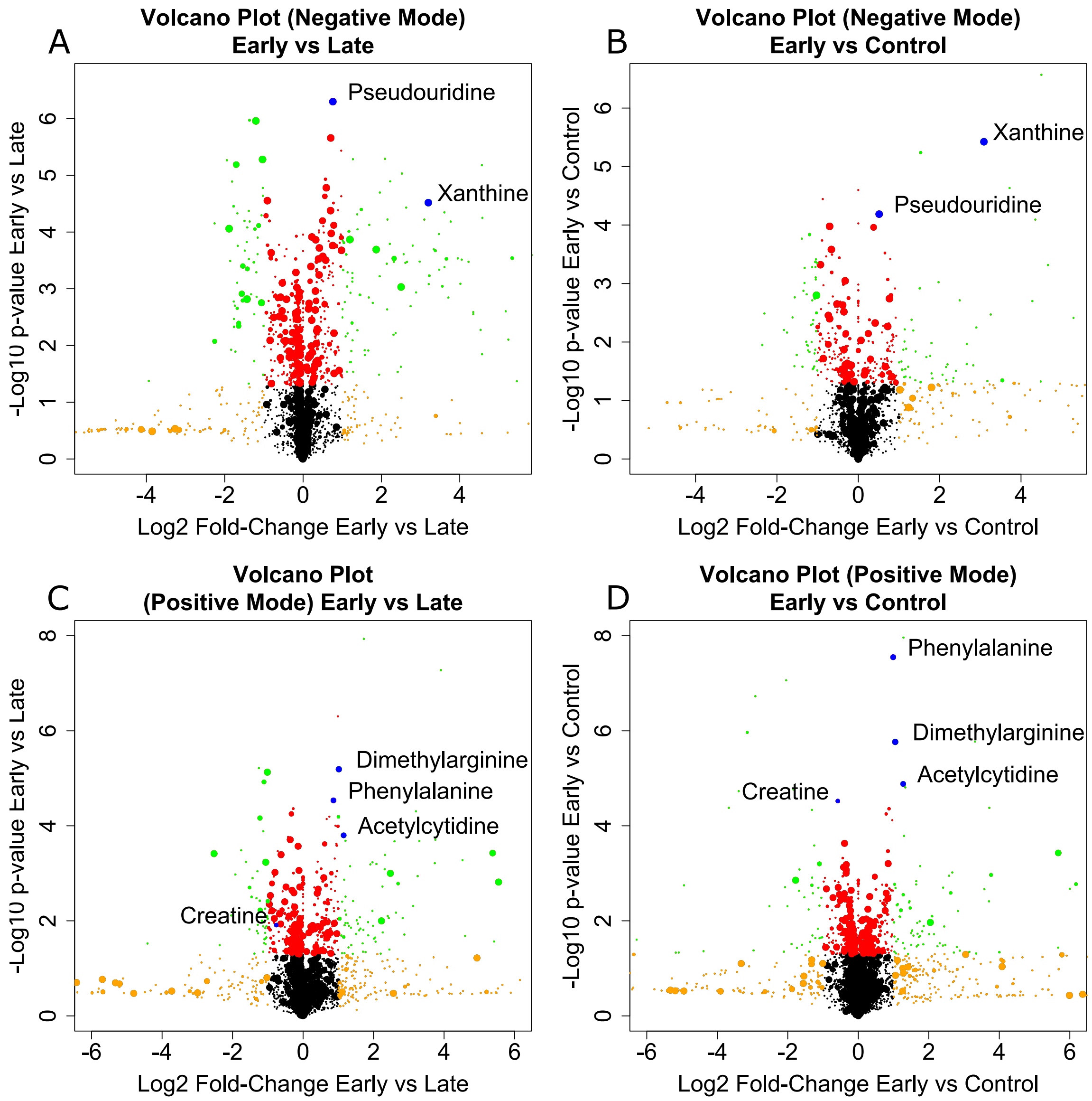


Figure 4-4 Volcano plots of untargeted metabolomic data of CSF. A,C - from children with ALL at diagnosis vs the same children on maintenance chemotherapy; B,D from children with ALL at diagnosis vs unmatched, normal controls. A,B show data for negative mode LC-MS and C,D for positive mode LC-MS.

Table 4-1 Top 22 Features by ROC analysis area under curve for Early vs Late and Early vs Normal with putative metabolite IDs shown. Features only seen in either Early vs Late or Early vs Normal analysis but not in both in grey. Low abundance metabolites had peak area $<1x10^{4}$ and were not investigated further.

Early vs Late					Early vs Normal				
Feature	Area under curve	t-test	log2 fold-change	Putative ID	Feature	Area under curve	t-test	log2 fold-change	Putative ID
6.96_283.0682m/z	0.989474	1.62E-05	18.45205	Xanthine	6.11_258.1086m/z	0.985294	1.09E-08	1.273087	5-methylcytidine
8.86_243.0621m/z	0.968421	1.96E-08	0.748173	Pseudouridine	3.13_430.2434m/z	0.979412	1.87E-07	-8.85663	
5.88_128.0705m-z	0.963158	1.03E-10	1.738966	Low abundance metabolite	3.20_455.2752n	0.979412	1.08E-06	-4.48734	
5.95_191.1389m-z	0.963158	1.46E-10	6.277073		3.30_243.1827m/z	0.979412	8.66E-08	-2.18896	
6.11_258.1086m-z	0.957895	3.27E-07	1.016278	5-methylcytidine	5.06_174.0761m/z	0.976471	1.73E-05	-2.06572	
6.66_166.0724m-z	0.952632	1.01E-07	0.840054	Phenylalanine	6.96_283.0682m/z	0.973529	4.20E-05	14.87414	Xanthine
6.64_279.1310m-z	0.934211	5.52E-06	5.896813	Low abundance metabolite	6.66_166.0724m/z	0.973529	2.83E-08	0.960477	Phenylalanine
5.72_283.1034m-z	0.931579	1.10E-06	1.031647		5.47_208.1291m/z	0.970588	0.00108	7.584732	
5.72_312.1300m-z	0.931579	2.23E-06	0.967139	Dimethylguanosine	6.64_279.1310m/z	0.970588	1.69E-06	8.281152	Low abundance metabolite
18.71_203.1501m-z	0.926316	1.88E-06	1.005927	Dimethylarginine	5.47_131.0448m/z	0.961765	0.00215	2.93144	
5.80_328.1612m-z	0.921053	3.55E-06	7.008089		5.88_128.0705m/z	0.958824	3.57E-10	1.818522	Low abundance metabolite
4.01_100.0757m-z	0.913158	0.000122	-2.25383		3.26_415.2113m/z	0.958824	1.87E-05	-11.2342	
5.18_145.0738n	0.910526	5.17E-07	-0.33699		5.95_191.1389m/z	0.95	3.43E-09	4.678004	
5.60_154.0610m-z	0.905263	1.40E-05	1.238755	Low abundance metabolite	18.71_203.1501m/z	0.944118	1.72E-06	1.009451	Dimethylarginine
14.90_194.0177n	0.902632	4.43E-05	10.06583		5.00_202.1801m/z	0.929412	0.0014	-1.64995	
5.57_286.1031m-z	0.9	1.59E-05	1.186994		5.57_286.1031m/z	0.9	1.31E-05	1.36419	Aceylcytidine
7.67_153.0407m-z	0.9	0.000322	1.768083		5.72_312.1300m/z	0.9	4.37E-05	0.853781	Dimethylguanosine
5.16_86.0600m-z	0.894737	1.81E-06	-0.28169		4.93_188.1392m/z	0.894118	0.000501	16.73596	
6.03_198.0873m-z	0.889474	2.14E-05	-1.14468		5.60_154.0610m/z	0.891176	1.56E-05	1.528614	Low abundance metabolite
11.41_146.0812m-z	0.889474	1.06E-05	-0.42146		5.72_283.1034m/z	0.888235	5.62E-05	0.824153	
3.68_149.0574n	0.884211	0.151283	-7.71376		4.96_143.0815m/z	0.885294	0.000372	4.258023	
11.92_132.0656m-z	0.881579	0.006222	-0.64479	Creatine	8.86_243.0621m/z	0.876471	6.52E-05	0.51311	Pseudouridine
4.93_188.1392m-z	0.878947	0.000245	15.78589		11.92_132.0656m/z	0.870588	3.01E-05	-0.56553	Creatine

The identities of these compounds were confirmed using standards obtained from suppliers (section 2.1.1.1). The retention time and fragmentation spectra of the standards were compared with those of the target peaks, and all of the peaks identities were confirmed with the exception of acetylcytidine.

As noted above, it is possible to identify allopurinol –a xanthine oxidase inhibitor – in the CSF from some of the “Early” samples which may provide an explanation for the increased xanthine seen in the CSF in this group - this is the most likely reason for the apparent difference. A correlation analysis of xanthine vs allopurinol abundance at first glance seems to confirm this – but this is heavily skewed by the complete absence of allopurinol from most samples. Interstingly, by analysing only samples with reliably detectable levels of allopurinol (taken as an arbitrary cut-off of peak area 1x10^4) there is no evidence of a correlation between CSF allopurinol abundance and CSF xanthine abundance (Figure 4-5)

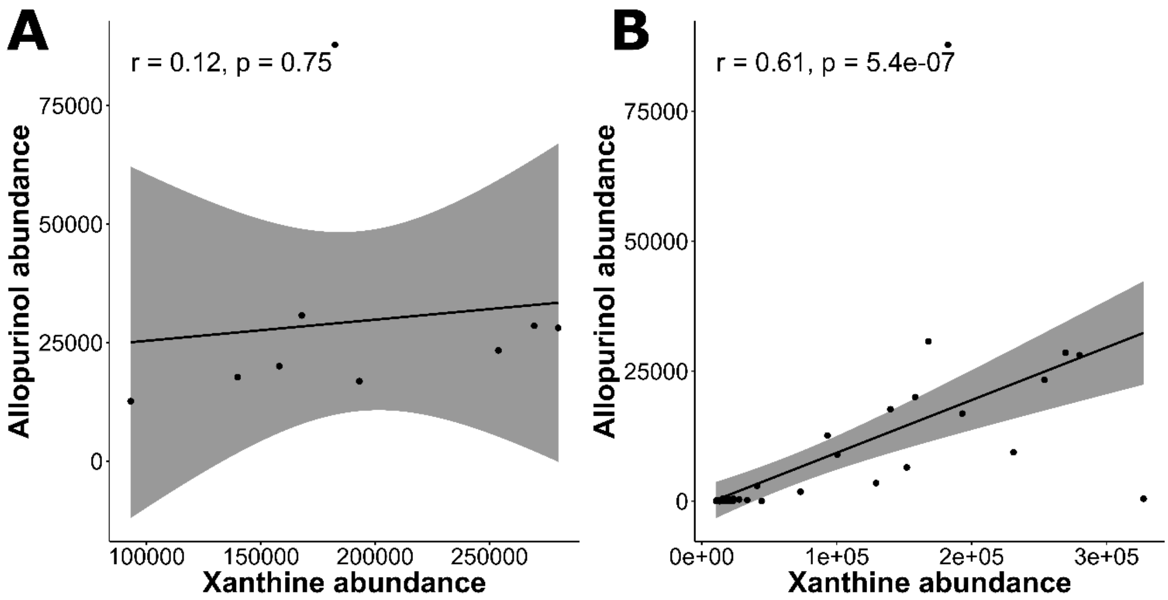


Figure 4-5 Scatterplots of allopurinol vs xanthine abundance in CSF showing best-fit regression line, confidence intervals in grey, with correlation coefficient and p-value at top for: A – samples with detectable allopurinol (n=8); B- all samples (n=56).

These metabolites were therefore taken forward for further investigation in a mouse model of CNS leukaemia and in CSF from patients with CNS relapse.

4.3.4 Summary

From these analyses it was concluded that it is possible to discriminate CSF taken at diagnosis with ALL from CSF taken from the same children on ALL maintenance chemotherapy, and from normal controls. In additions there are some features which are found at significantly higher or lower abundance in CSF samples at diagnosis which are potential markers for CNS ALL.

4.4 Targeted LC-MS analysis of leukaemic CSF using a murine model

To confirm the significance of these findings, metabolomic analysis of CSF and plasma from mice with and without leukaemia was carried out. Using a mouse model has several advantages over using primary data. There is certainty that the leukaemic mice have CNS leukaemia, and that the controls have no CNS pathology. There are no other complicating factors that could compound analysis (e.g. drug treatments, infections). It is relatively simple to obtain CSF in a controlled manner and sample preparation is much more controlled (though as noted above (Chapter 4.2) this does not seem to be a critical factor for CSF analysis), and in addition it is simple to obtain matched blood in a similarly controlled way. So despite the clear physiological differences between humans and mice, a xenograft model of human ALL cell line cells in immunocompromised NSG mice was used as in previous experiments (section 2.2.1.1).

9 experimental NSG mice were injected with SEM cells via the tail vein, with 8 mice without leukaemia injection as controls. After 28 days, CSF was extracted under terminal anaesthesia, then blood obtained via cardiac puncture; Cells were retrieved from the CNS and spleen as described above (section2.2.2.5). As this chapter is focussed on analysis of CSF and blood plasma, the results of analysis of cellular metabolites from this experiment are described in a later chapter (5.4.3).

Metabolomic analysis was carried out as described previously (section 2.2.6.1). The data from the CSF of 2 leukaemic mice were excluded from analysis – one had insufficient CSF for analyses (i.e <1μL), and one was grossly bloodstained, with clearly aberrant results. The results of plasma for 1 leukaemic mouse were excluded as the sample was insufficient. For CSF/Plasma comparisons, the data from all 3 of these mice were excluded.

4.4.1 CSF analysis

Looking first at the targets suggested by patient CSF, the most promising finding was creatine which was clearly lower in CSF from leukaemic mice (Figure 4-6). Pseudouridine was also found to be significantly lower in CSF from leukaemic mice.

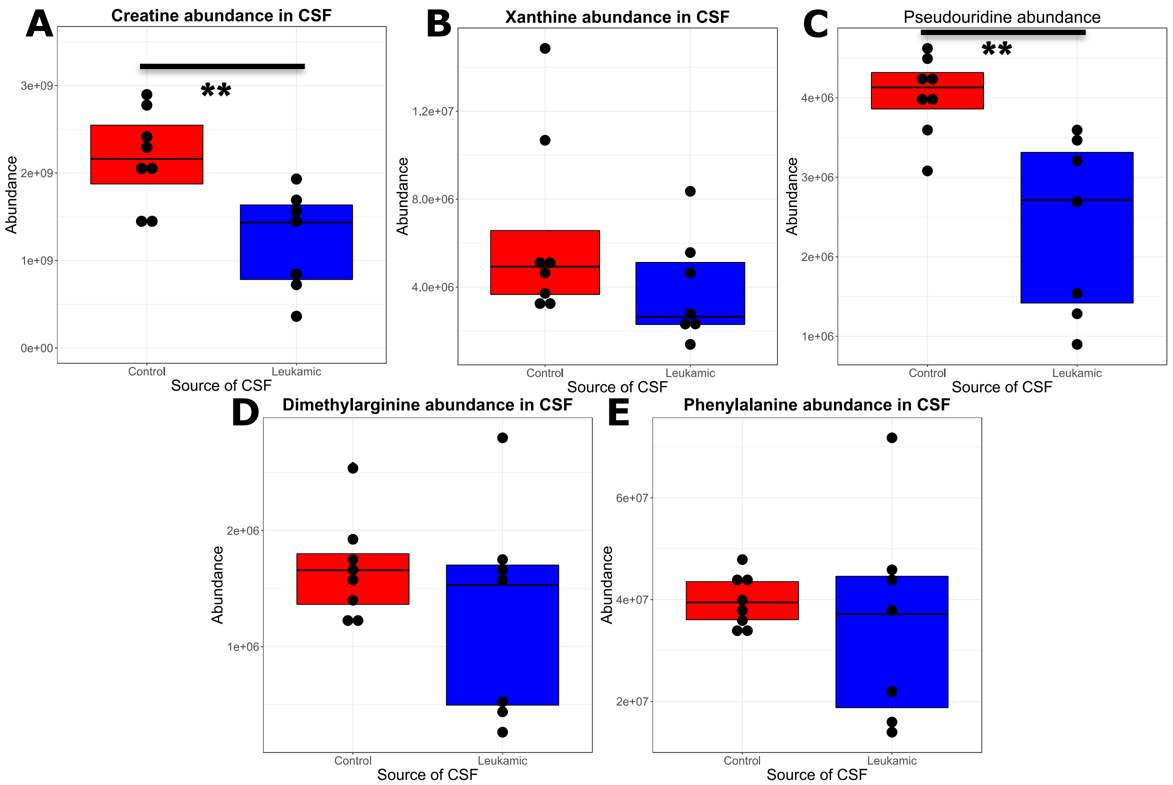


Figure 4-6 Boxplots of abundance of target metabolites in CSF from mice with CNS leukaemia vs mice without leukaemia. A – Creatine (p=0.0055); B – Xanthine; C – Pseudouridine (p=0.0023), D – Dimethylarginine; E - Phenylalanine.

With the consistent reduction in creatine between the initial untargeted study and this mouse model, related compounds creatinine and phosphocreatine were also analysed, showing a consistent reduction creatinine in CSF from mice with leukaemia (Figure 4-7).

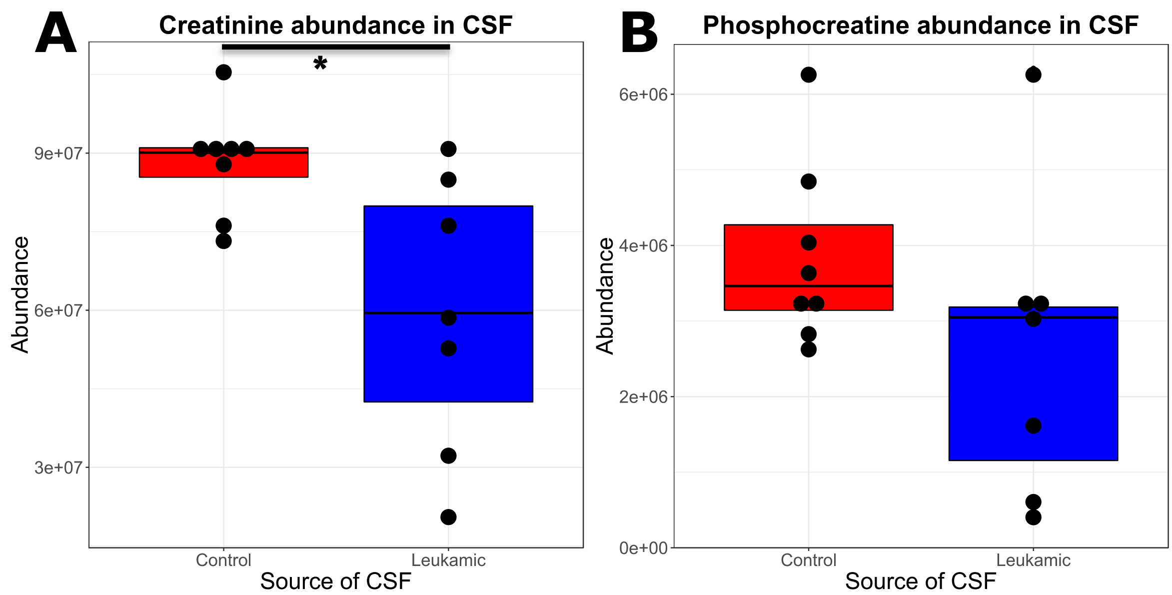


Figure 4-7 Boxplots of Creatinine and Phosphocreatine abundance in the CSF of mice with leukaemia vs mice without leukaemia. A – Creatinine (p=0.0108); B – Phosphocreatine.

4.4.2 Plasma analysis

Whilst the mouse model used undoubtedly was useful for investigating CNS leukaemia it was not a model for isolated CNS leukaemia. i.e. it is possible that the changes found in the CSF relate to wider changes in metabolism in mice with systemic leukaemia. To try to determine how much of the change in the metabolome in CSF was related to the systemic effects of leukaemia and how much may truly be a local effect of CNS leukaemia, firstly the abundance of these metabolites in murine plasma in the presence or absence of leukaemia was assessed (Figure 4-8).

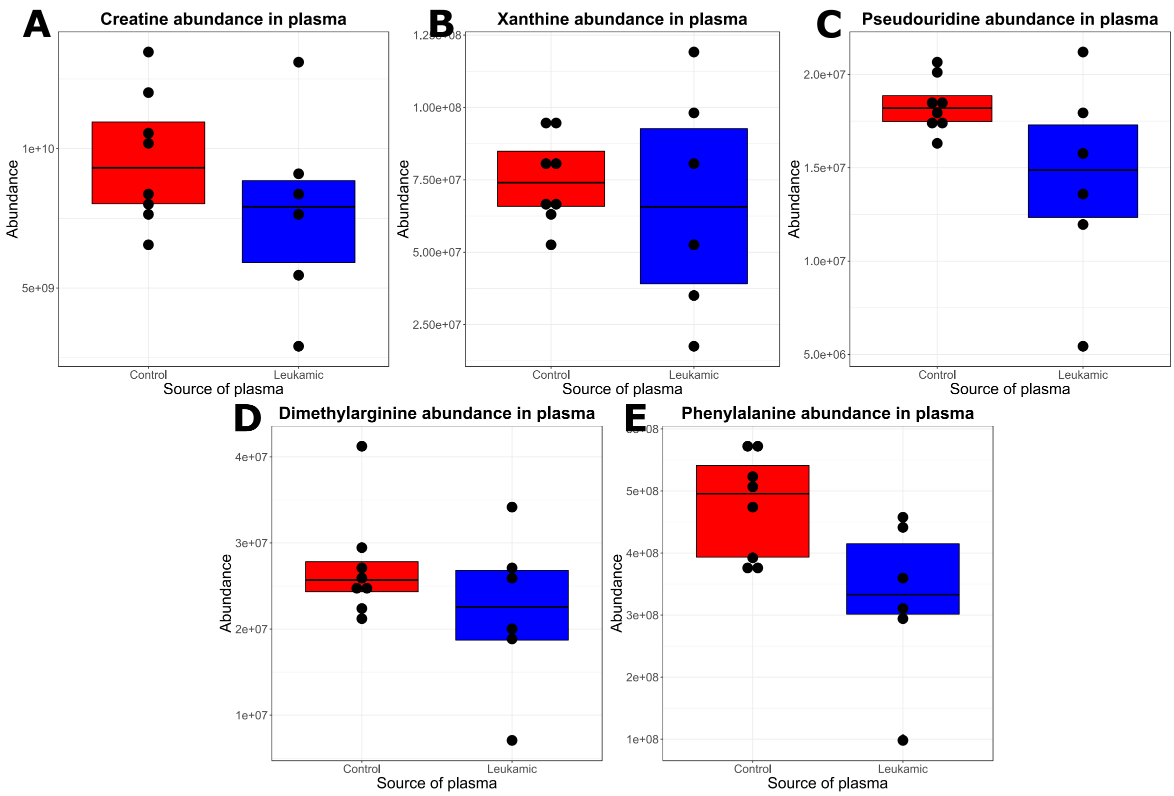


Figure 4-8 Boxplots of abundance of target metabolites in blood plasma of mice with and without leukaemia. A – Creatine; B – Xanthine; C – Pseudouridine; D – Dimethylarginine; E - Phenylalanine

Whilst none of these metabolites showed statistically significant changes in plasma abundance in the presence or absence of leukaemia in mice, creatinine (a metabolite closely linked to creatine) was significantly lower in abundance in plasma from mice with leukaemia (Figure 4-9).

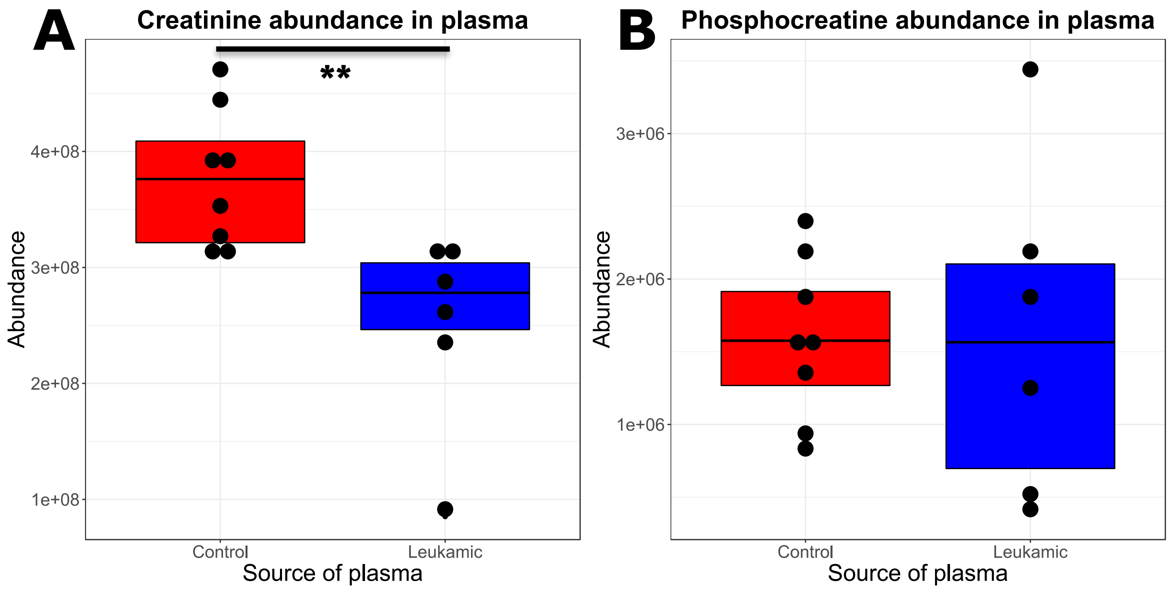


Figure 4-9 Boxplots of abundance of creatinine and phosphocreatine in blood plasma of mice with and without leukaemia. A – Creatinine (p=0.00615); B - Phosphocreatine

4.4.3 CSF and Plasma analysis

To look at more depth in to the relationship between levels of metabolites in the plasma and in the CSF, the ratio of these target metabolites in CSF:plasma was calculated and these ratios for normal and leukaemic mice compared. Whilst this did not show any statistically significant changes specific to the CSF of leukaemic mice, there was a trend towards a reduced pseudouridine CSF:plasma ratio (p-values=0.051). This lends some weight to the suggestion that changes in CSF abundance of theis metabolite may represent a CNS-specific effect of leukaemia.

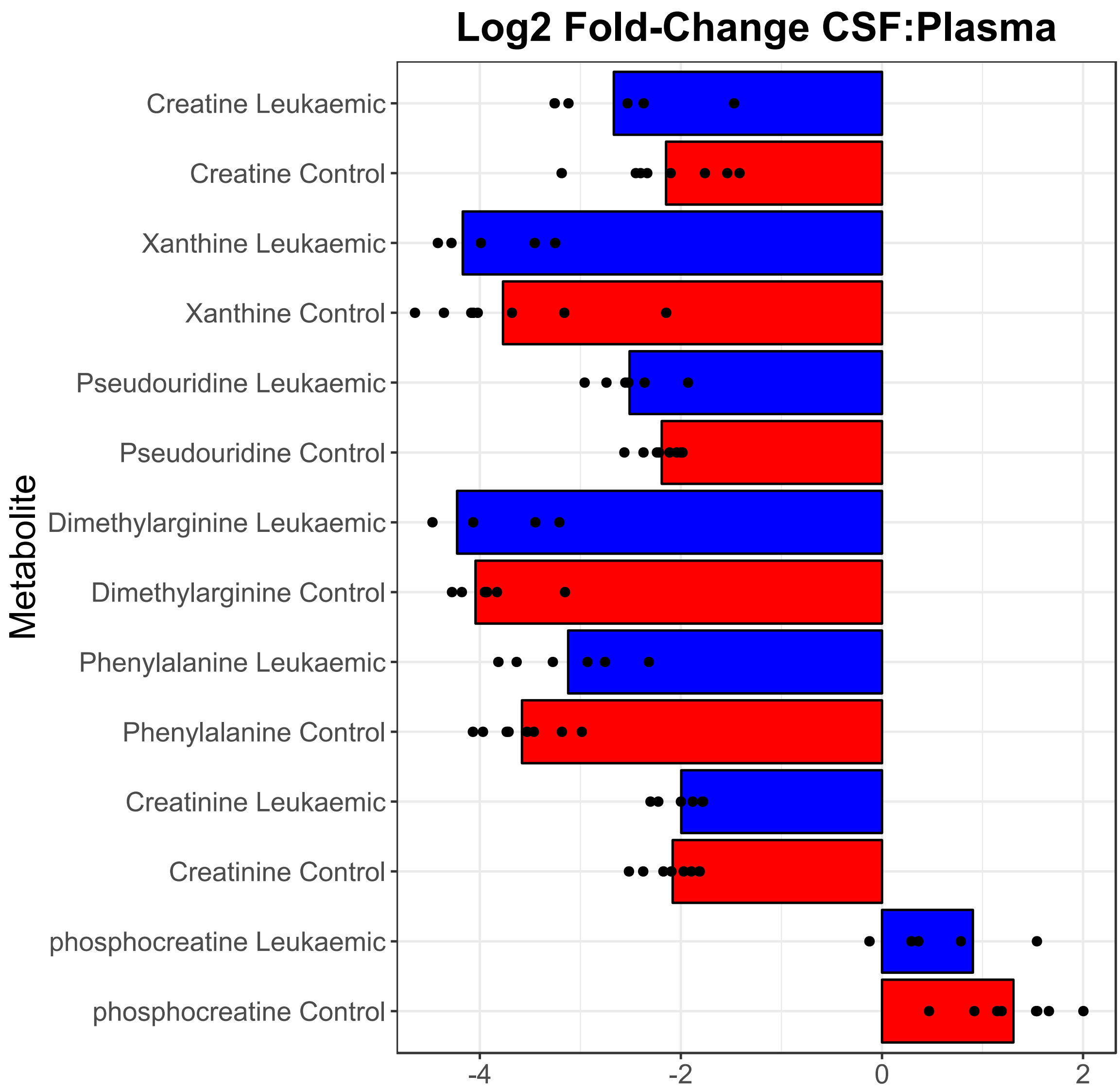


Figure 4-10 Waterfall plot of log_2 fold-change between CSF and plasma for target metabolites in mice with and without leukaemia

4.4.4 Summary

From this experiment it was concluded that there are clear metabolomic changes in the CSF and plasma of NSG mice with xenografted leukaemia, and that some of these changes are in keeping with findings from untargeted metabolomic analysis of human CSF from children with leukaemia. In particular a reduction in creatine seemed to be a consistent finding. There is a suggestion, though not confirmed, that some of these changes are more marked in the CNS and may be caused by the presence of CNS leukaemia.

4.5 Targeted LC-MS analysis of CSF from children at CNS relapse of ALL

The final experiment in this part of the project involved looking at the CSF of children with known isolated CNS relapse of ALL. 4 children with CNS relapse were identified. As noted in the patient characteristics table, the timing of CNS relapse varied widely, and two of the children had undergone bone marrow transplant for previous relapse prior to sample storage for this analysis. In addition only 2 of the patients had a sample stored from diagnosis with CNS relapse, the others had samples available from within a short space of time from the diagnostic sample but not from the actual diagnostic sample.

To provide a comparison, 10 patients who did not go on to develop CNS relapse had CSF samples from diagnosis, day 8, day 29 and 1 year analysed. This was done in order to provide a timeline of the abundance of target compounds during ALL therapy. Patient characteristics for relapse and control samples were described previously (Table 2-4). Some of the control patients also had samples analysed for the initial untargeted LC-MS experiment (section 4.3).

Samples were prepared as above (section 2.2.6.1) a “targeted” LC-MS set-up was used and results analysed. Initially looking at the relapse samples on their own, a timeline of abundance of each of the target metabolites from the untargeted LC-MS experiment above (section 4.3) was constructed (Figure 4-11).

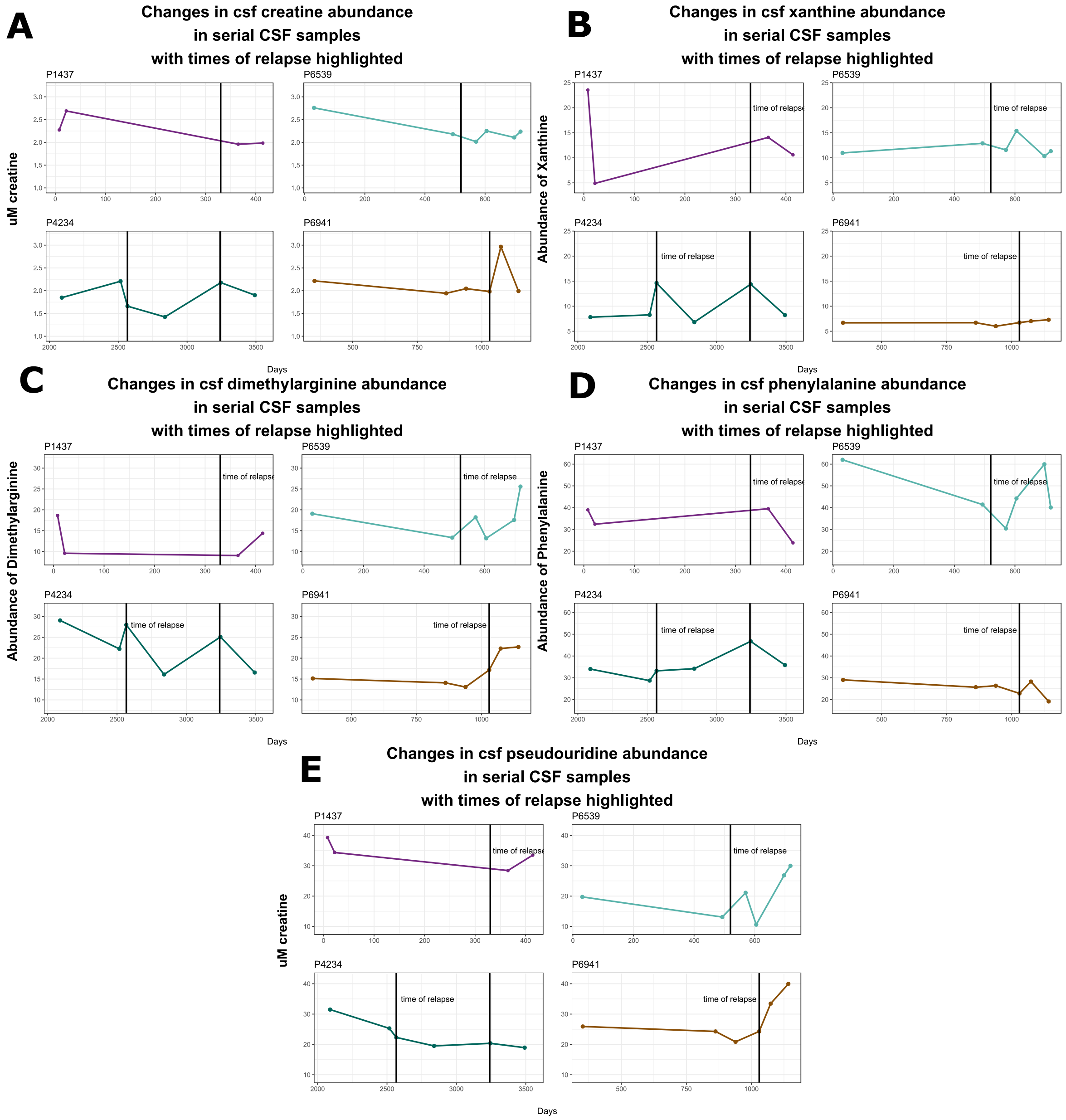


Figure 4-11 Changes in abundance of target metabolites in CSF of 4 children with isolated CNS relapse of ALL; metabolite abundance shown as a line chart with time of isolated CNS relapse denoted by vertical lines. One graph per child for each metabolite: A – Creatine; B – Xanthine; C – Dimethylarginine; D – Phenylalanine; E - Pseudouridine

From here the most promising metabolites were xanthine – which consistently increased around the time of CNS relapse in 3 of the 4 children, though as discussed previously is likely to be related to use of xanthine-oxidase inhibitor in 2 of these patients (P1437 and P6539) who had known CNS relapse prior to CSF sampling – and creatine, which appeared to decrease in the same children. Patient P6941 seemed to have little change in xanthine or creatine at the time of CNS relapse. Analysis of the data from children who did not go on to suffer CNS relapse confirmed the initial findings regarding creatine abundance at diagnosis vs later on in therapy (i.e. creatine was clearly decreased at diagnosis compared to later in treatment), and there appears to be a reduction in creatine abundance in 3 of 4 children at time of relapse, though the wide variation in background abundance of creatine in children who did not relapse makes potential use of creatine as a marker of CNS disease in ALL problematic (Figure 4-12).

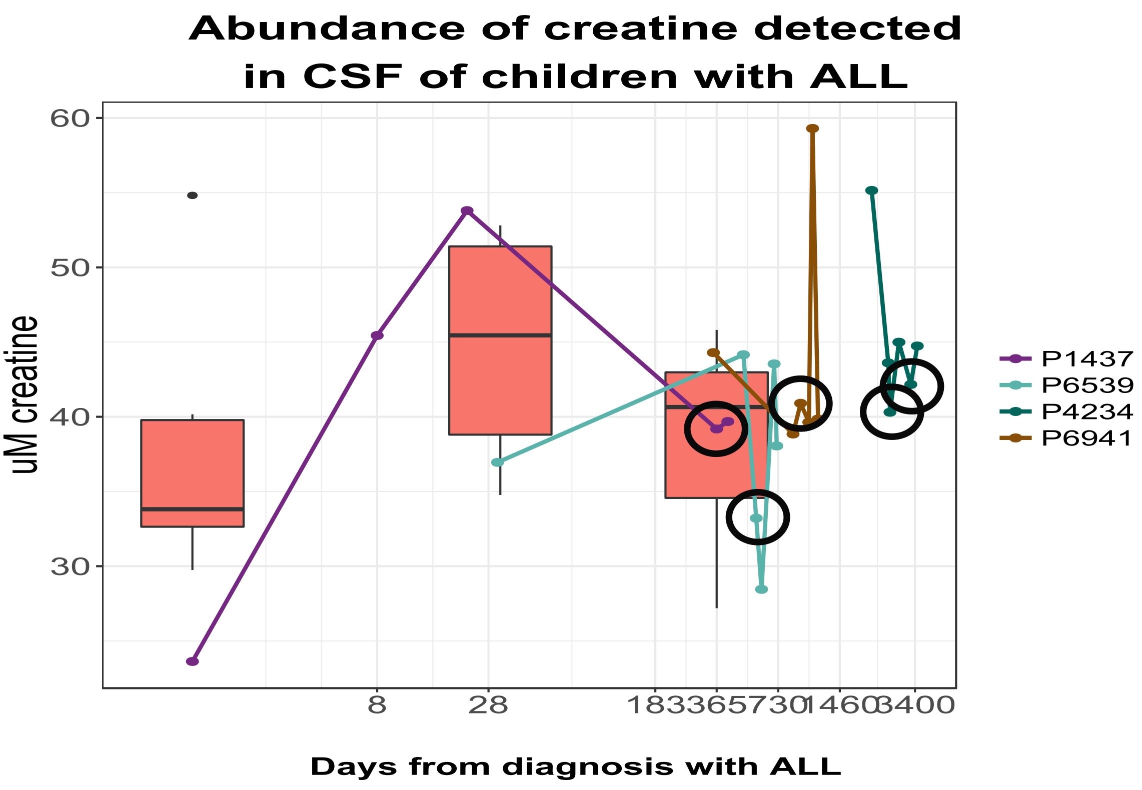


Figure 4-12 CSF Abundance of creatine as a timeline in 4 children who suffered CNS relapse of ALL. Boxplots show median abundance (line), quartiles (box) and range (whiskers, outlier as dot) in CSF from patients with ALL who did not suffer CNS relapse at days 1, 28 and 365 post-diagnosis. Sample taken closest to isolated CNS relapse in each child denoted by circles. Note time is on a log axis.

4.5.1 Summary

There are metabolomic changes in CSF at the time of CNS relapse that are consistent with changes between CSF at diagnosis and later in treatment in children who did not suffer CSN relapse, and between CSF at the time of isolated

CNS relapse and at other timepoints. In particular, creatine is a promising potential marker for CNS relapse in ALL, though this requires further evaluation.

4.6 Conclusions.

The aim of this chapter was to identify putative metabolic markers for the presence of CNS ALL. The data shows it is possible to detect metabolomic changes in CSF related to the presence of ALL using LC-MS.

As discussed previously (section 2.2.6), it is possible to examine large parts of the metabolome using a chromatography/mass spectrometry approach to discriminate and quantitate the abundance of metabolites. One of the first questions that needed to be addressed was what part of the metabolome was to be targeted. The more common and better characterised approach is LC-MS. This allows excellent discrimination and identification of small metabolites up to short peptides and is the optimal approach to examine many cellular processes e.g. glycolysis/TCA or one-carbon metabolism. This approach can identify some fatty acids, though retention times for all fatty acid and lipid metabolites tend to cluster very closely. This approach, however, cannot detect most lipids (e.g. cholesterol or its lipid precursors).

Likewise, the main alternative that is used for lipid metabolism analysis – GC-MS– is excellent for analysing lipid molecules but not suitable for most polar metabolites.

Both approaches require different sample preparation, and for GC-MS this is reasonably intensive, and in both approaches samples must be prepared quickly and efficiently in order to preserve the metabolome for analysis. This can make preparing the same sample for both concurrently difficult, so one approach must be prioritised over the other. GC-MS analysis may seem at first glance more suitable for this project, given the focus on cholesterol metabolism from previous results (section 3.2). However, lipidomics results can be complicated with multiple isomers and less differentiated retention times. In addition “lipidomics” is a less mature field than small-molecule metabolomics which can make interpretation of results more difficult.

On balance, it was decided to focus on “standard” LC-MS for this part of the project, in order to provide a more comprehensive and comprehensible picture of the metabolome. It does provide some data on aspects of lipid metabolism (e.g. fatty acids and polar cholesterol precursors) as well as allowing an in-depth look at the small and polar metabolites. The LC-MS technology is more mature and there are more public databases of more metabolite mass spectra to help with identifying compounds. In essence it was felt this approach provided the best chance of finding viable, clinically applicable candidate biomarkers of leukaemic CNS infiltration. Some simple lipidomic analyses were carried out and are detailed below (chapter 5.3)

This is the first time untargetd LC-MS analysis has been used for detection of ALL in the CNS. This is an extremely powerful technique, providing large amounts of data, and has given evidence for several potential markers of CNS ALL, with two different control groups allowing the detection of robust omarkers. It is possible that there are more sensitive and specific markers that could be detected as a metabolic signature of several metabolites, but network analysis was beyond the scope of this thesis – this work is ongoing.

There are clear limitations to this study. Firstly, while there is good evidence to assume that most or all children with ALL have CNS disease at the time of diagnosis, we have no way to confirm this with current diagnostic techniques. As discussed previously (section 2.2.6) this type of LC-MS detects only small molecules, and is not able to detect larger or lipid metabolites, which may have interesting changes given the findings of other parts of this thesis. Additionally this study uses primary human samples which is excellent for providing “real-world” clinically relevant data, but does allow for many more possible confounding factors (e.g. the administration of allopurinol as discussed previously). Additionally there is no way to discriminate a signal related to systemic ALL from a signal related directly to CNS ALL from these data.

The use of supporting data from mouse models and from primary CNS relapse CSF samples helps overcome some of these shortcomings. While mouse models provide very controllable data, this comes with the limitation that there are clearly large differences between mouse and “real-life” CSF, and any potential markers in CSF from murine models may not apply to human disease. CNS relapse is ultimately what this part of the project is trying to predict by looking at early and reliable markers in the CNS. It makes sense therefore that any metabolomic changes in the CSF that could be used as a marker should be detectable at the time of CNS relapse. In addition, in patients with isolated CNS relapse there is no evidence of systemic ALL, so any changes we find are very likely to be related to the presence of CNS leukaemia. A disadvantage of this approach is the fact that there were few suitable samples and not all samples (i.e. diagnostic, pre-relapse, relapse, post-relapse) are available for all children. In addition therapies such as prior bone marrow transplant may significantly alter the CSF metabolome.

There are some consistent changes between human samples, mouse xenograft model samples, and isolated CNS relapse samples, but there are also some interesting differences. Of particular note xanthine was markedly different in children at diagnosis compared with either control group, was unchanged in mouse model samples lending credence to the idea that this the changes in xanthine relate to allopurinol exposure rather than ALL.It should be noted that in addition to the lack of correlation between detected allopurinol abundance in CSF and xanthine abundance (chapter 4.3.3), patient P4234, who has clear increases in xanthine abundance at the time of CNS relapse would certainly not have had recent exposure to allopurinol prior to CSF sampling.

Metabolomic analysis is an extremely powerful tool for investigating the metabolome and the detection of novel biomarkers though more work is required to streamline the identification of targets and to develop tools for the integration of multi-omics data.

Chapter 5: Metabolic changes in Precursor B-cell ALL in the Central Nervous System

5.1 Introduction and aims

As noted in Chapter 3:, analysis of transcriptomic data from *in vivo* models and primary data showed that the cholesterol biosynthesis pathway is consistently upregulated in ALL cells in the CNS compared with systemic disease (BM/Spleen). In addition, analysis of CSF and plasma has shown metabolic changes in the presence of ALL. The next step in this project was to look more directly at the metabolic changes in ALL cells in the nutrient-poor (section 1.3.2) CSF, focussing on investigating the findings outlined in previous chapters.

Specifically in this chapter, the aims were:

1. To investigate the effects of modulating cholesterol metabolism in ALL cells *in vitro* and *in vivo* to determine if cholesterol metabolism was a critical pathway for CNS leukaemia that could be a potential clinical target

2. To investigate lipid and fatty acid metabolism more generally *in vivo* using human CSF analysis and human ALL cells in the CNS and mouse CSF from a xenograft model

3. To look more generally at the ALL metabolome in CNS ALL using metabolic tracers *in vivo* in a xenograft model

5.2 *In vitro* studies

Working with *in vitro* cell cultures allows precise control of experimental variables that is not possible with primary samples or complex *in vivo* models. This was used to investigate the importance of the cholesterol synthesis pathway in a childhood ALL cell line (SEM) cells' growth and viability. To do this, experiments were carried out to determine the cholesterol abundance of culture conditions, select appropriate conditions for further experiments, and clarify the effects of these conditions on cell growth and viability prior to directly investigating cholesterol metabolism.

5.2.1 Effects of increasing serum concentration on cell expansion and cell death

The first series of experiments explored the effects of increasing concentrations of FCS in DMEM cell culture medium on SEM cells. This is a crude but simple way of changing the abundance of cholesterol available to the cells. Firstly, the abundance of cholesterol in media with different proportions of FCS was measured using the Amplex™ Red Cholesterol Assay Kit (ThermoFisher) as described previously (section 2.2.6.6). Cholesterol was measured as free cholesterol and total cholesterol, with esterified cholesterol calculated from the difference. From these results, 1%-FCS-DMEM has a roughly equivalent concentration of total cholesterol to children's CSF and is therefore a reasonable benchmark around which to base this *in vitro* model system. There was a reduction in free cholesterol in 1% FCS-DMEM compared to CSF, but this makes up a small part of the cholesterol pool (Figure 5-1).

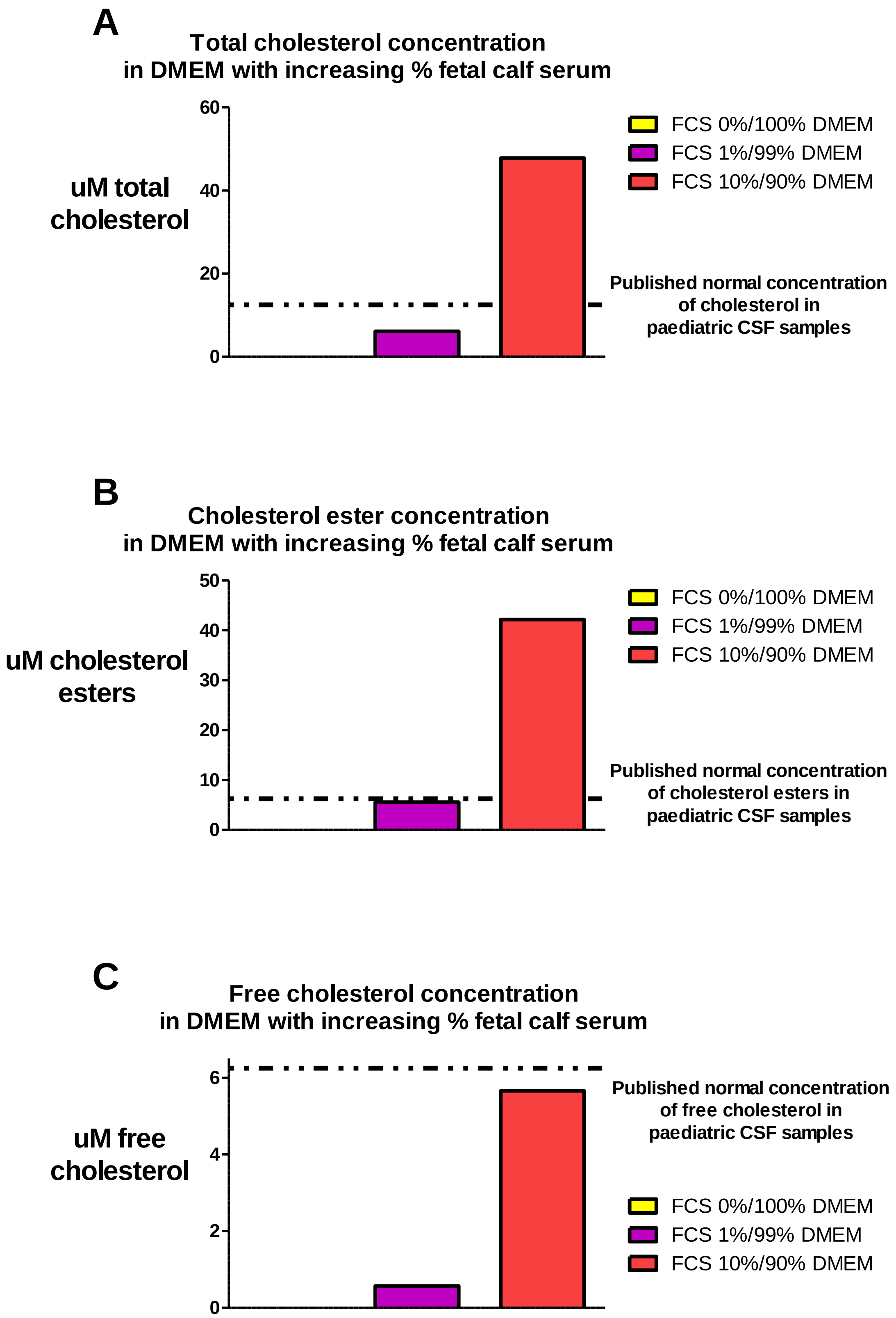


Figure 5-1 Cholesterol quantification in DMEM with increasing percentage of FCSA - Total cholesterol, B – Cholesterol esters; C – Free cholesterol. Data shown as mean of two

replicates. Dotted line shows published normal concentration of Total/Free/Esterified cholesterol in paediatric CSF samples(Illingworth & Glover 1971)**.**

The next part of this experiment was to explore the influence of increasing proportion of FCS in DMEM on SEM cells *in vitro*. Cell culture methods are described previously (section 2.2.7), briefly: cells were incubated with cell proliferation dye e450 then cultured in standard culture media with increasing concentrations of FCS:

- FCS 0%/100% DMEM
- FCS 1%/99% DMEM
- FCS 5%/95% DMEM
- FCS 10%/90% DMEM

After 24, 48 and 72 hours cells were counted, harvested and stained with annexin-V (conjugated to APC) and PI and analysed by flow cytometry as previously described (section 2.2.9).

Unsurprisingly, there was an increase in cell proliferation both by e450 staining and cell counts as FCS concentration increased. In the absence of FCS there was a sharp drop in cell numbers, which slowly recovered over the next 48 hours. In 1% FCS there appeared to be steady cell proliferation, though at a slower rate than in 5% or 10% FCS. No significant differences were detected between cells grown in 5% or 10% FCS.

A similar pattern was seen when looking at cell viability and apoptosis – cells grown without FCS had markedly increased apoptosis. This was reduced in 1% FCS, and reduced further in 5% or 10% FCS with no significant differences between cells grown in 5% or 10% FCS (Figure 5-2). Similar results were seen at 24 hours and 48 hour timepoints.

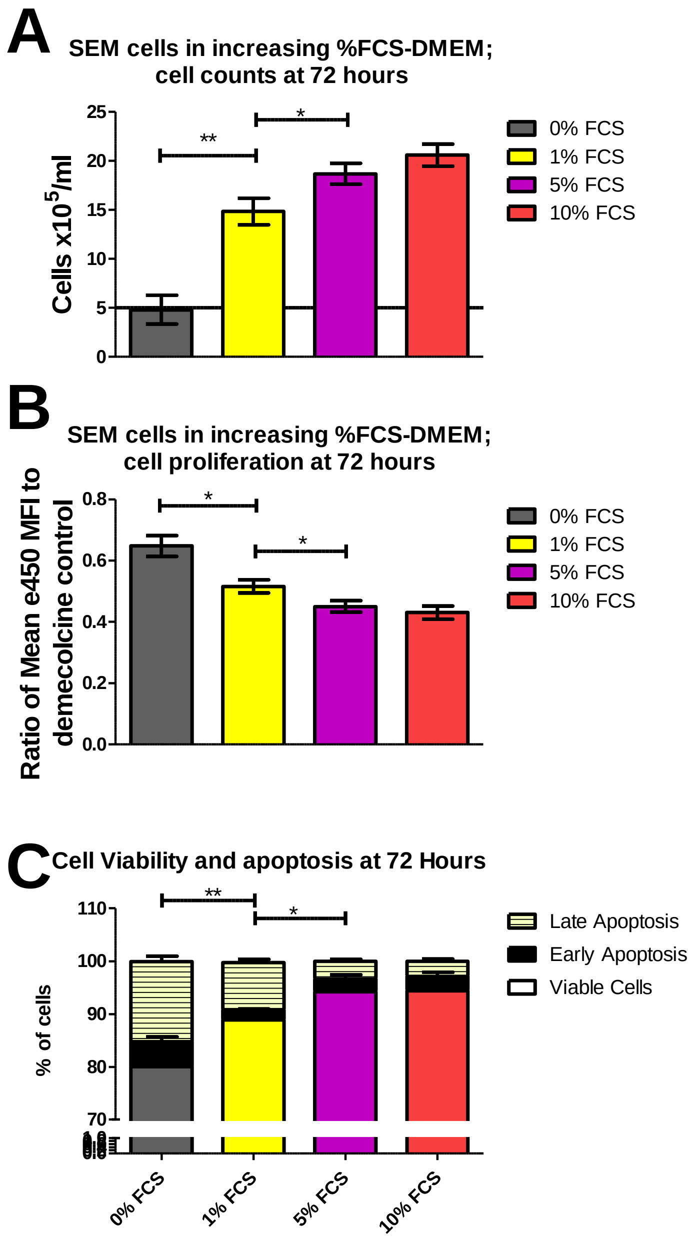


Figure 5-2 SEM cell counts, cell proliferation and cell viability in increasing concentrations of FCS after 72 hours. A – Cell counts; B – Cell proliferation; C – Cell viability/apoptosis. Proliferation measured by ratio of mean fluorescence intensity (MFI) of cell proliferation dye e450 to cell-cycle arrested control (i.e. lower value represents higher proliferation). n=3, Mean + SEM shown. Line on graph A indicates cell count at the start of the experiment.

5.2.1.1 Summary

Cholesterol abundance increases with increasing proportion of FCS in DMEM. 1% FCS has a roughly equivalent level of total cholesterol to published values for normal paediatric CSF. There is a clear improvement in SEM cell survival and increased proliferation between 0% FCS- and 1% FCS-DMEM, and between 1% FCS-and 5% FCS-DMEM with no clear difference between cells grown in 5% FCS- and 10% FCS-DMEM. This is not to suggest that cholesterol abundance is the limiting factor in lower-% FCS-DMEM, but it provides some evidence on which to base choice of FCS percentage in media. In particular in CSF-like levels of cholesterol (1% FCS-DMEM) there remains robust cell growth and cell viability.

For future experiments investigating cholesterol metabolism, media with 1% or 10% serum (the standard concentration of FCS in tissue culture media) were used for experimental models of leukaemic cells in CSF-like abundance of cholesterol (1% FCS) vs more abundant cholesterol (10% FCS).

5.2.2 Effects of simvastatin on cell expansion and cell death

The next phase of this series of experiments was to look at the direct effect of statin therapy on SEM cells *in vitro*. Statins are small molecule inhibitors of the enzyme HMG-CoA reductase, which catalyses the cleavage of HMG-CoA to create the first committed precursor in cholesterol biosynthesis, mevalonate. Statins competitively inhibit the active site of the HMG-CoA reductase molecule preventing the cleavage of mevalonate (Figure 5-3).

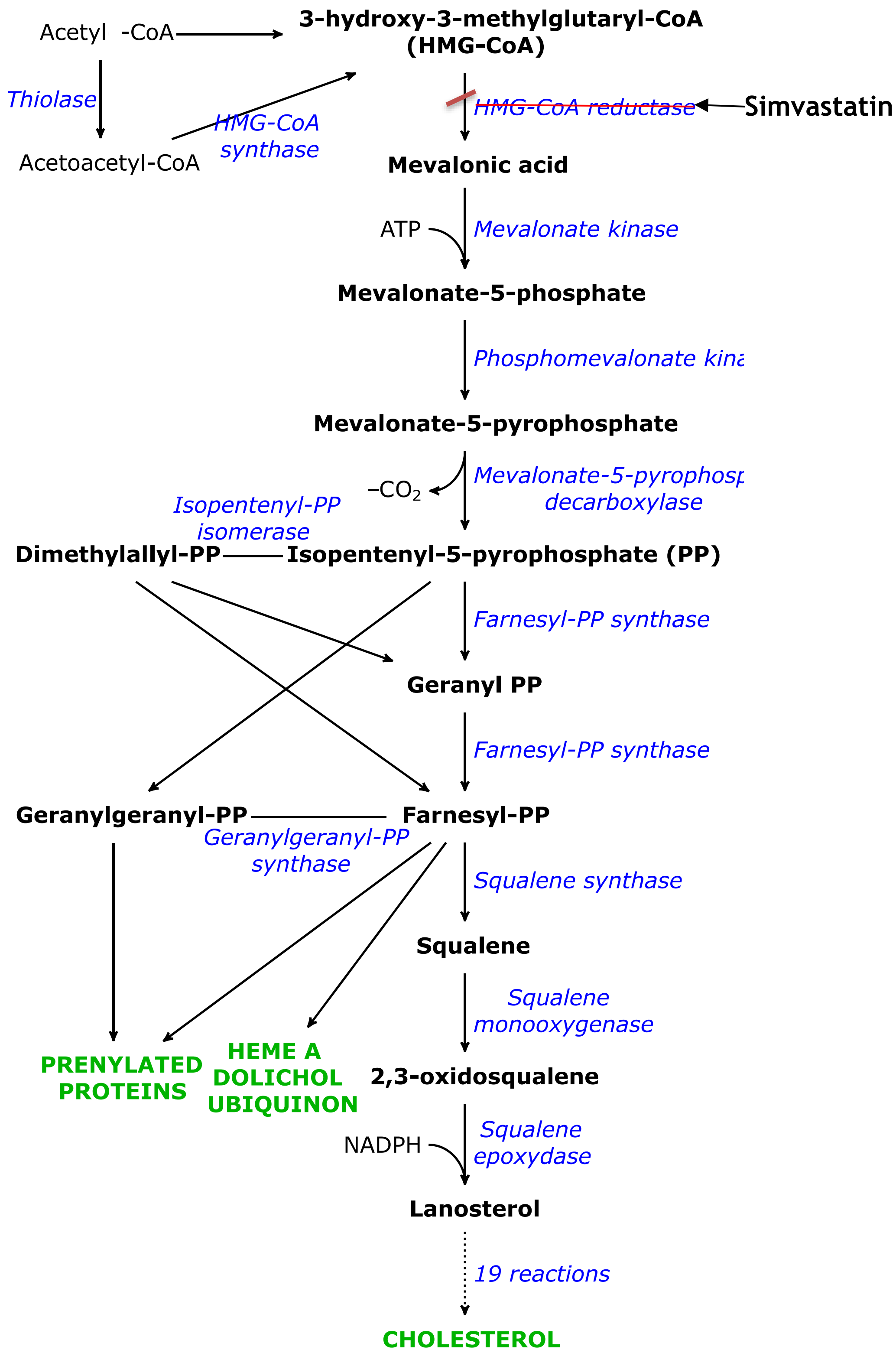


Figure 5-3 Diagram of the mevalonate pathway with the effects of simvastatin shown adapted from "HMG-CoA reductase pathway" by user Krishnavedala used under a creative commons licence (https://commons.wikimedia.org).

Simvastatin was selected as the statin of choice based on its superior penetration into CSF, and therefore greater potential for efficacy against CNS disease *in vivo*. Simvastatin has limited solubility in water, so was dissolved in DMSO as described above (section 2.2.7.4). Similar experiments have previously been published, though in different cell lines, in different culture conditions, with different methodology, and without cholesterol rescue (Sheen et al. 2011).

5.2.2.1 DMSO dose titration

A series of experiments were carried out to determine the optimal concentration of DMSO to use. These were performed initially in 10% FCS-DMEM to provide a baseline, then in 1% FCS-DMEM to ensure no differences appeared with lower levels of serum. These experiments used the same cell count, proliferation, and cell viability/apoptosis methods as above, and were performed in two independent experiments (and were therefore not suitable for statistical analysis). These results showed that there was little evidence of DMSO toxicity at concentrations below 1% after 72 hours of cell culture (Figure 5-4). For future experiments simvastatin was prepared in 0.1% DMSO. Similar results were found at 24 and 48 hour timepoints.

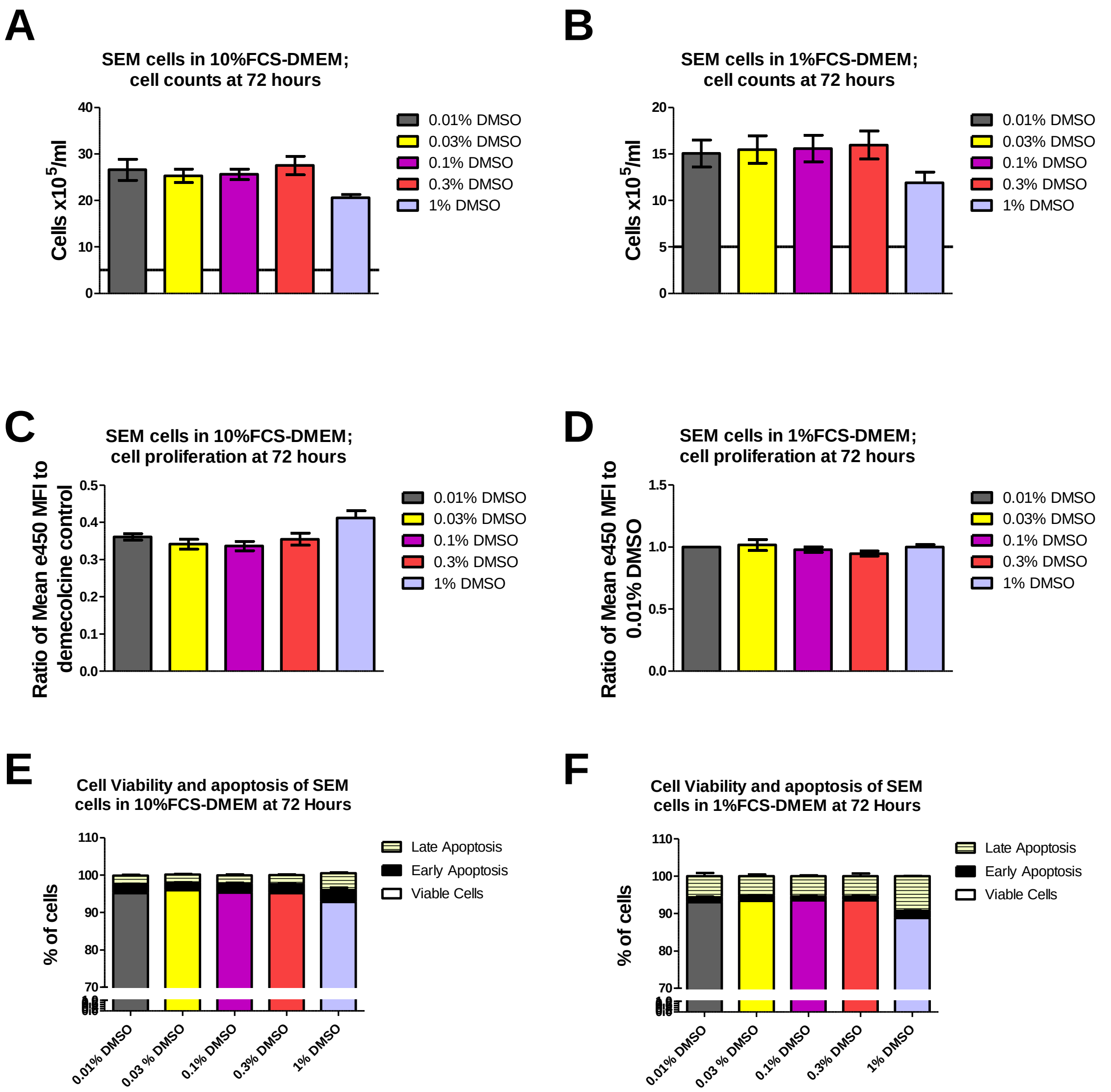


Figure 5-4 SEM cell counts, cell proliferation, and cell viability with increasing concentrations of DMSO *in vitro* in (A, C, E) 10% FCS-DMEM or (B, D, F) 1%FCS-DMEM after 72 hours. A, B – cell counts; C, D – proliferation (note proliferation compared to 0.01%DMEM internal control in 1% FCS experiments); E,F – Cell viability and apoptosis. n=2; Mean + SEM shown. Line on A, B indicates cell count at the start of the experiment

5.2.2.2 Simvastatin dose titration

The next step was to evaluate the effective *in vitro* dose of simvastatin. This was done in a similar way to previous experiments, with cell counts, cell proliferation and cell viability after 24, 48 and 72 hours of cell culture in a range of simvastatin doses from 1μM to 10μM in 1% FCS-DMEM and 10% FCS-DMEM. These experiments showed a clear impact of simvastatin on SEM cell proliferation and apoptosis at a dose of 10μM..

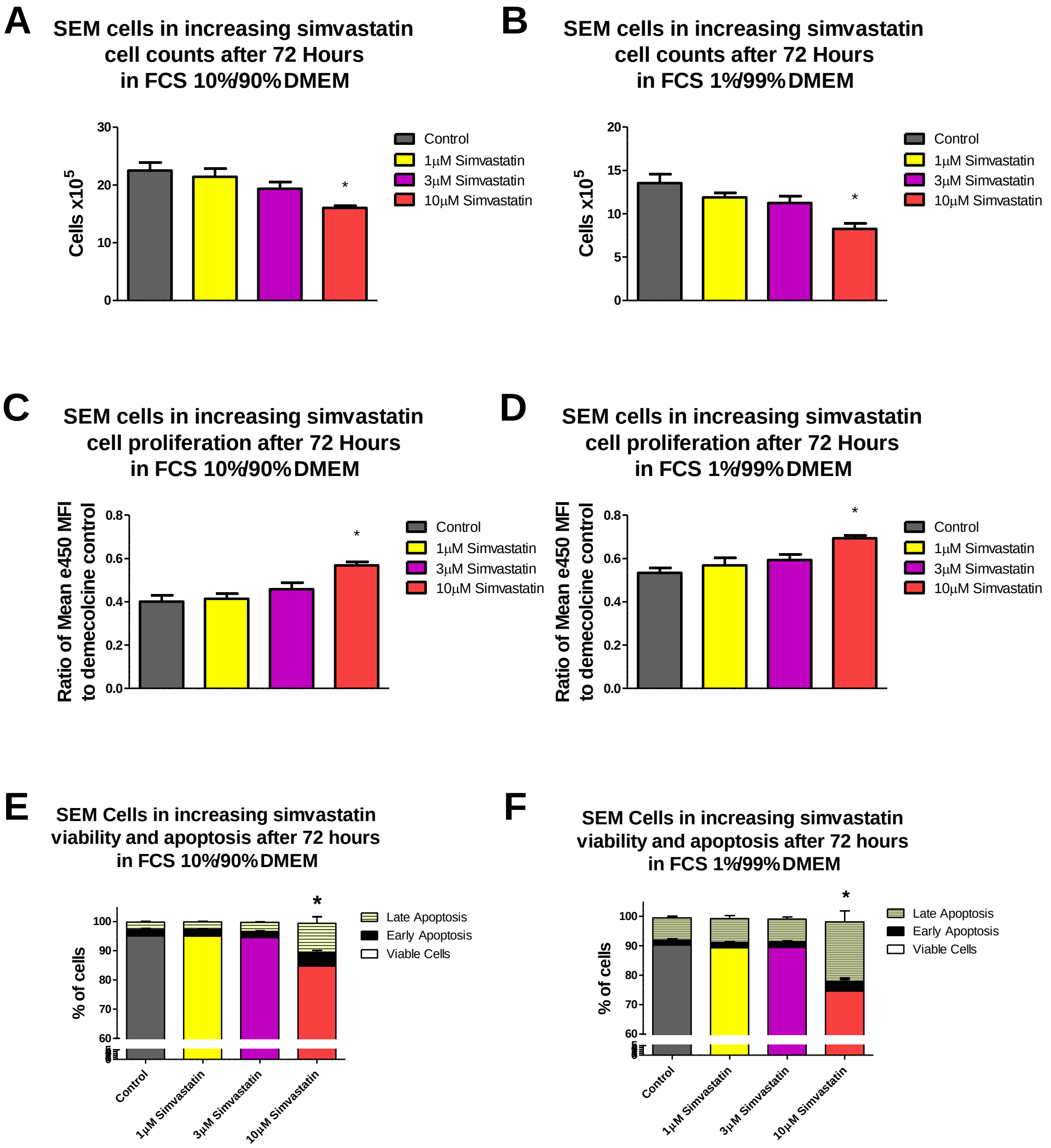


Figure 5-5 SEM cell counts, cell proliferation and cell viability with increasing concentrations of simvastatin *in vitro* in 10% FCS (A, C, E) and 1%FCS (B, D, F). A, B – cell counts; C, D - proliferation; E, F – cell viability and apoptosis. n=4, Mean + SEM shown. Line on A, B indicates cell count at the start of the experiment. *denotes p-value (at highest/least significant) 0.01-0.05 for 10µM simvastatin vs all other conditions.

5.2.2.3 Effect of rescue of simvastatin toxicity with mevalonate or cholesterol

The next step was to confirm this effect of simvastatin was mediated through effects on HMGCR rather than non-specific “off-target” effects. To do this, cells were exposed to an effective dose of simvastatin (10µM, based on the above experiments) with and without either mevalonate or cholesterol prepared as described previously (section 2.2.7.4).

Firstly mevalonate, the direct product of the reaction simvastatin inhibits, was used. A high dose of mevalonate was chosen arbitrarily at 20μM as there was no evidence of mevalonate cytotoxicity at high doses. Addition of mevalonate showed clear rescue of simvastatin cytotoxicity in SEM cells both in terms of cell growth and apoptosis in 10% FCS using mevalonate. The rescue of simvastatin toxicity was less marked in 1% FCS compared with 10% FCS with persistent reduced cell growth in simvastatin-treated SEM cells despite mevalonate rescue (Figure 5-6).

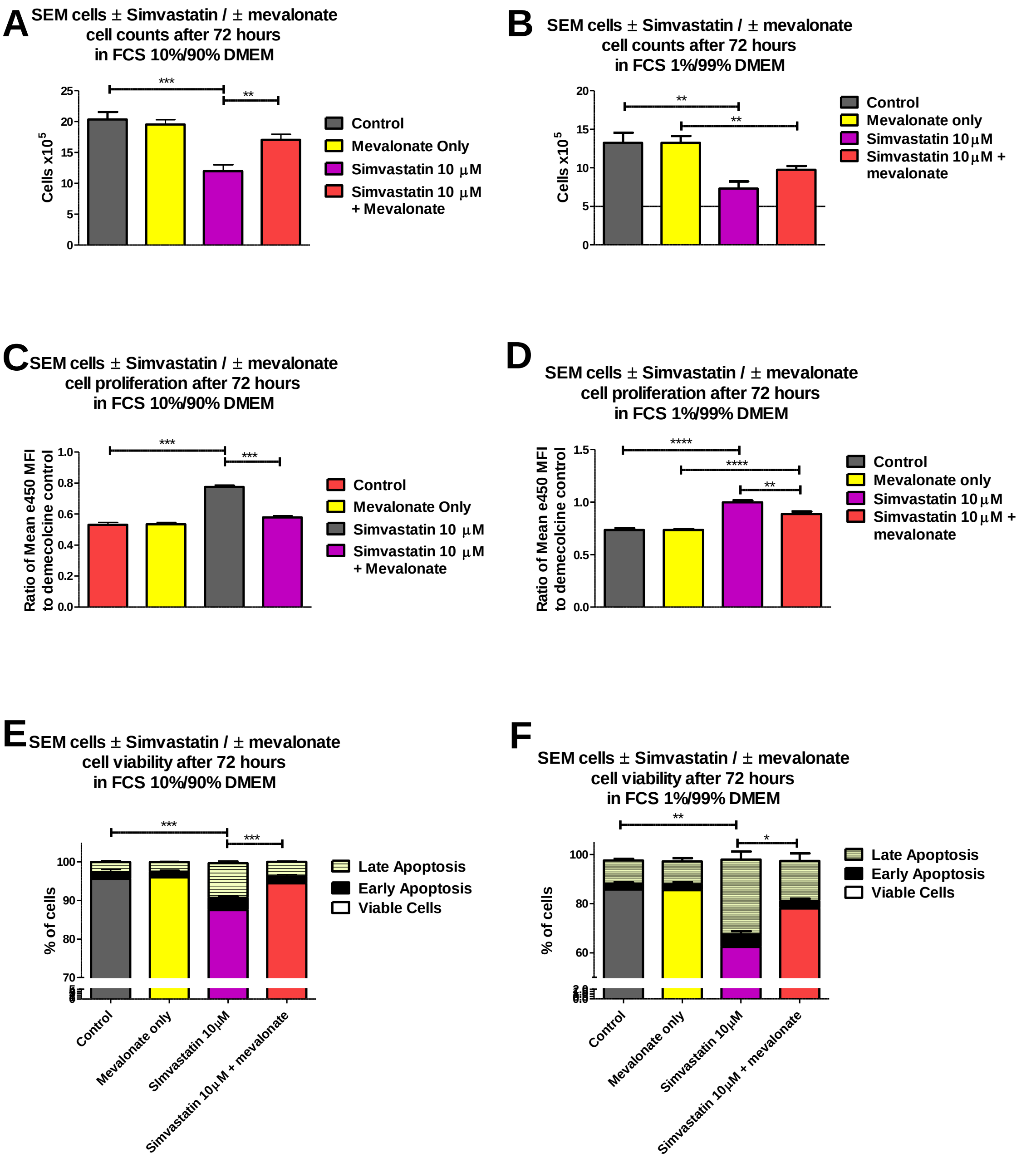


Figure 5-6 SEM cell counts and cell proliferation with simvastatin treatment ± mevalonate *in vitro* in (A, C, E) 10% and (B, D, F) 1% FCS A, B - cell counts; C, D – cell proliferation; E, F – cell viability and apoptosis. n=5 (cell counts) or 3 (other experiments), Mean + SEM shown. Line on cell counts indicates cell count at the start of the experiment. P-values denoted: *<0.05, **<0.01, *<0.001, **** <0.0001.**

The next phase of the project was attempting to rescue simvastatin toxicity with supplementing cholesterol, to assess whether reduced cholesterol itself, rather than an intermediate between mevalonate and cholesterol, was the cause of simvastatin cytotoxicity. High doses of cholesterol are known to be toxic, so an initial dose titration of cholesterol was performed, using effect of cholesterol on cell counts to measure toxicity. Little effect was seen at concentrations below 100μM, and if anything cells in increased cholesterol had a small growth advantage (Figure 5-7).

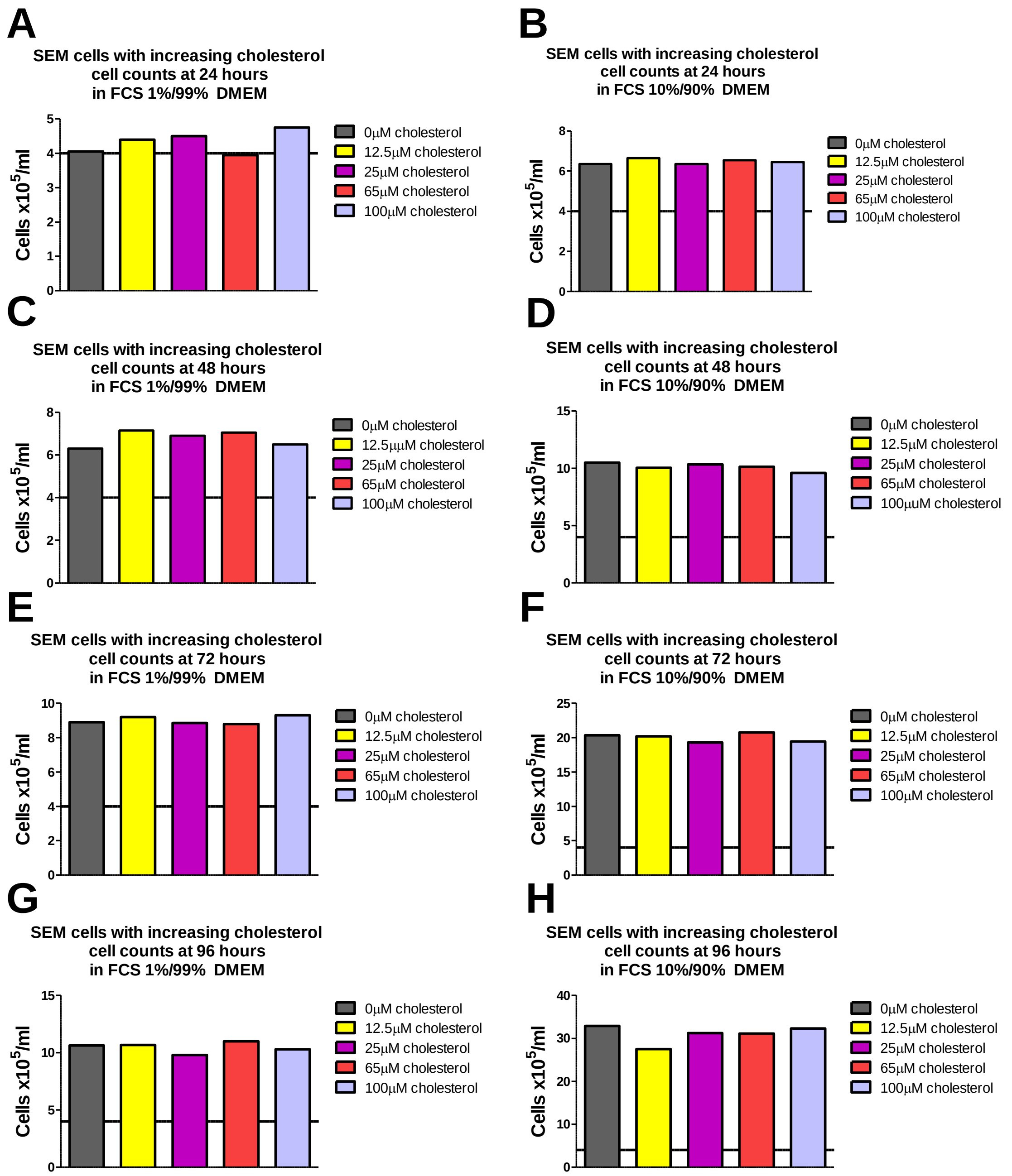


Figure 5-7 SEM cell counts with increasing concentration of cholesterol *in vitro* in 1% FCS (A, C, E, G) and 10% FCS (B, D, F, H); A,B – 24 hours; C,D – 48 hours; E,F – 72 hours; G,H – 96 hours. Experiment performed once - no statistical analysis carried out.

Given the lack of evidence of toxicity at these doses of cholesterol, a further experiment was carried out with higher doses, focusing on 72 and 96 hour timepoints(Figure 5-8). This showed a reduction in cell counts with 250μM cholesterol in 1% FCS, and with 500μM cholesterol in 10% FCS.

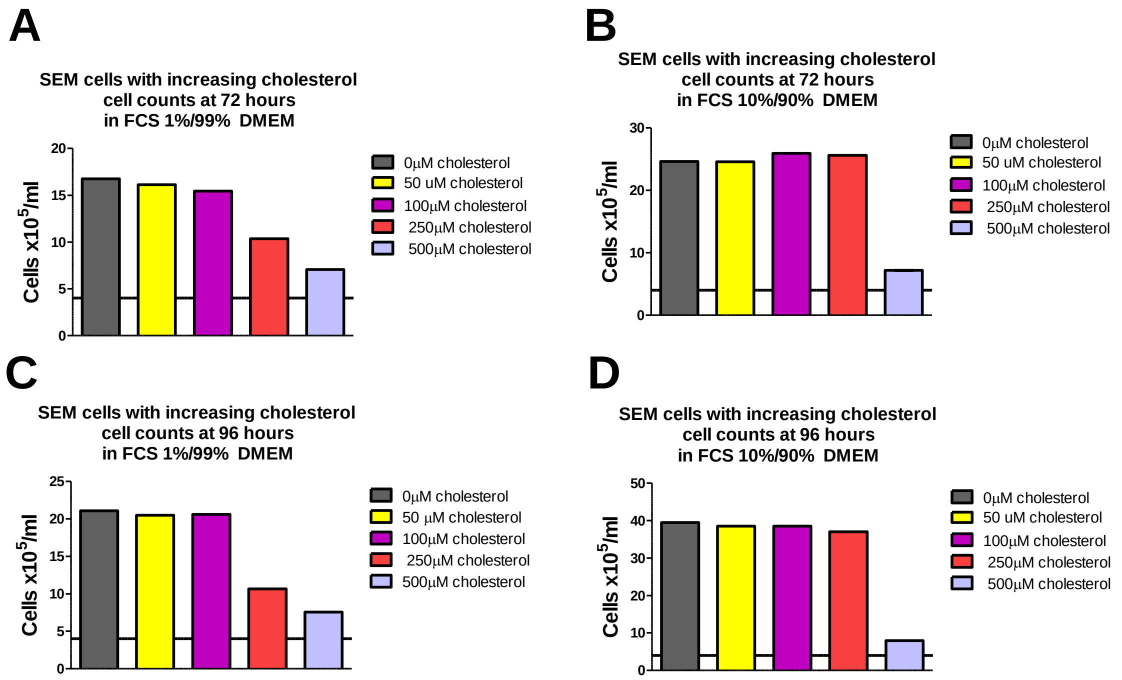


Figure 5-8 SEM cell counts with high and increasing concentration of cholesterol *in vitro* in 1% and 10% FCS A, C – 1% FCS; B, D –10% FCS; A, B – 72 hours; C, D – 96 hours. Experiment performed once only, therefore not suitable for statistical analysis.

As there was no evidence of cholesterol toxicity below 250μM, a concentration of 100μM cholesterol was initially chosen to attempt to rescue simvastatin cytotoxicity. This experiment showed gross cell death after 24 hours cholesterol "rescue" of simvastatin treatment, and for subsequent experiments the concentration of cholesterol was reduced 50% to 50μM. A repeat experiment using 50μM cholesterol rescue of simvastatin was then performed, showing no evidence of rescue of simvastatin cytotoxicity with cholesterol – there is instead a small reduction in cell viability with cholesterol "rescue" of simvastatin treatment in 10% FCS after 72 hours. In 1% FCS the increased toxicity in cells exposed to both simvastatin and cholesterol was more pronounced with a clear decrease in cell viability compared with simvastatin treatment alone (Figure 5-9).

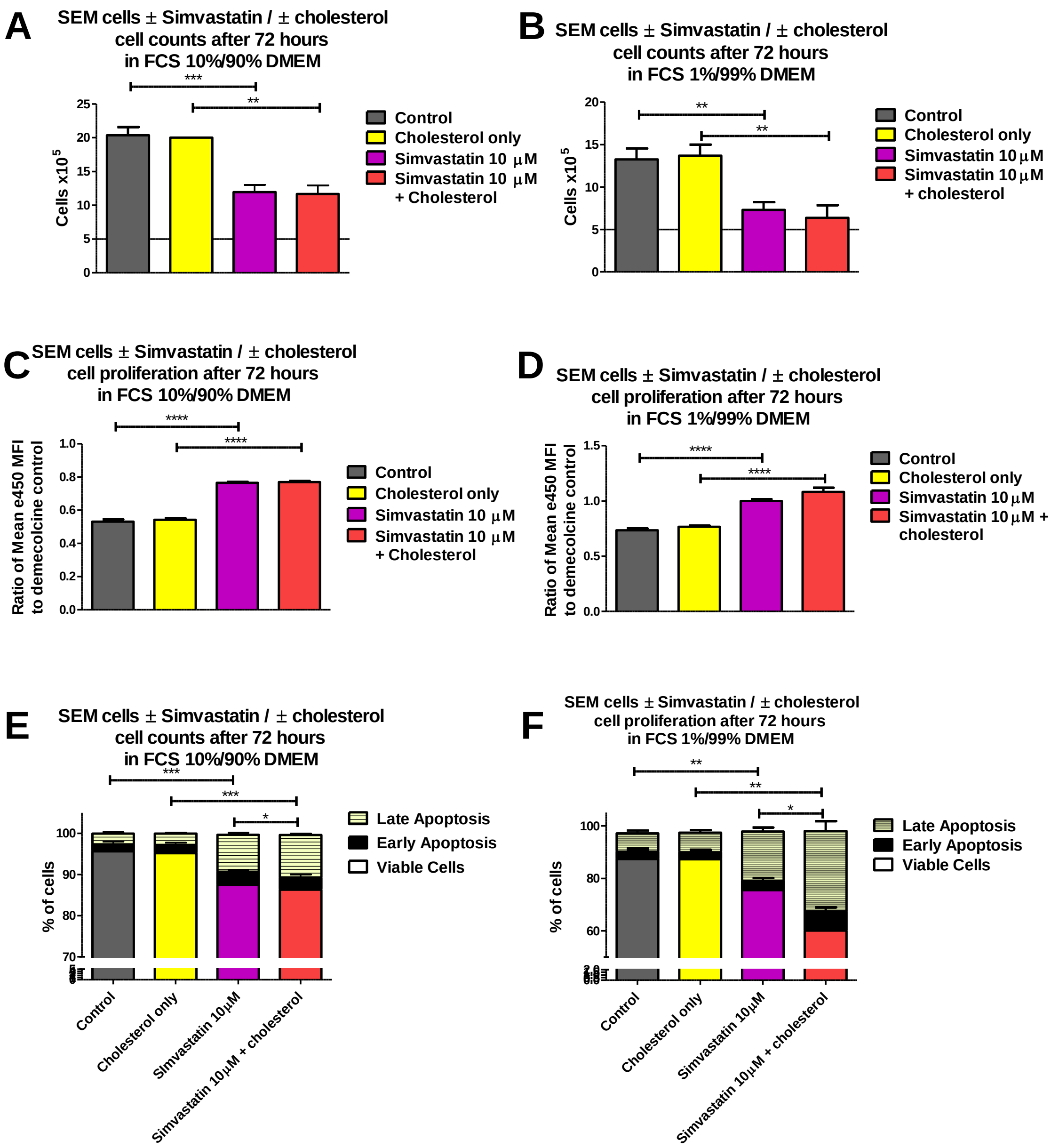


Figure 5-9 SEM cell counts, cell proliferation and cell viability with simvastatin treatment ± cholesterol *in vitro* in 10% FCS (A, C, E) or 1% FCS (B, D, F) A, B - cell counts; C, D – cell proliferation; E, F – cell viability and apoptosis. n=5 (cell counts) or 3 (other experiments).Mean + SEM shown. Line on cell counts indicates cell count at the start of the experiment. P-values denoted: *<0.05, **<0.01, *<0.001, **** <0.0001.**

5.2.2.4 Summary

Simvastatin is toxic to SEM cells *in vitro* at a concentration of 10µM. This toxicity is effectively rescued by the addition of mevalonate, the product of HMGCR (the enzyme inhibited by simvastatin), suggesting that this toxicity is via an on-target effect. The addition of cholesterol enhances the toxicity of simvastatin, particularly in low-serum conditions. This increased toxicity in 1% FCS may represent a specific effect of cholesterol depletion in a cholesterol-poor environment, but may simply reflect increased metabolic stress. One interesting hypothesis to explain the combined toxicity of simvastatin and cholesterol would be that addition of cholesterol suppresses HMGCR transcription at a genome level via SREBP, and simvastatin suppression at a protein level provides a double-hit to this pathway. This would be particularly effective if a metabolic intermediate between mevalonate and cholesterol was the essential component the lack of which was causing the toxicity seen. This was investigated in the following section.

5.2.3 Effect of simvastatin on gene expression

The next series of experiments were designed to see if the addition of exogenous cholesterol to the SEM cells *in vitro* reduced gene expression of HMGCR (potentially via the SREBP system, which is regulated by the intracellular abundance of sterols as noted above). To do this the experiments comparing SEM cells grown in 1% FCS-DMEM with 0.1% DMSO control ± cholesterol 50μM to cells grown with simvastatin 10μM ± cholesterol 50μM were repeated and RNA extracted from the cells after 72 hours. This was then converted to cDNA and analysed for HMGCR and HMGCS1 expression using qPCR as described above (section 2.2.5).

From these data, it appears that HMGCR is upregulated on treatment with simvastatin. There does not appear to be any impact of cholesterol supplementation, suggesting repression of gene transcription may not be the method of increased cell toxicity. It is worth noting the high rates of apoptosis in these cells may skew these results (Figure 5-10).

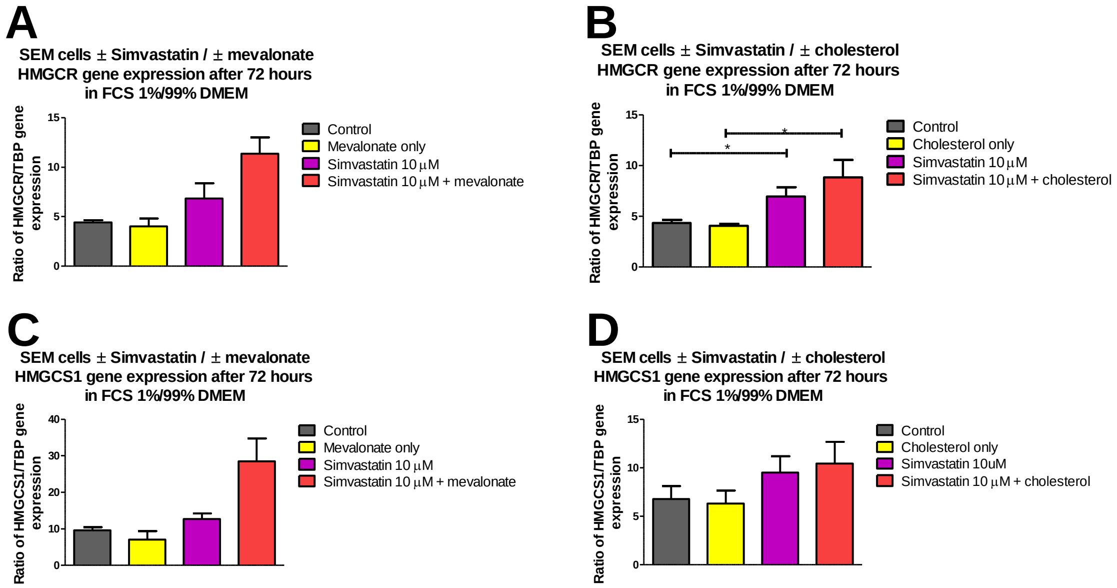


Figure 5-10 SEM cell qPCR for HMGCR and HMGCS1 with 72 hours simvastatin treatment ± cholesterol or mevalonate in 1% FCS. A – HMGCR gene expression with simvastatin ± mevalonate; B – HMGCR gene expression with simvastatin ± cholesterol; C – HMGCS1 gene expression with simvastatin ± mevalonate; D – HMGCS1 gene expression with simvastatin ± cholesterol. n= 3 (simvastatin ± mevalonate) or 6 (simvastatin ± cholesterol). Mean + SEM shown. P-values denoted: *<0.05.

5.2.4 Genetic manipulation of cholesterol synthesis *in vitro*

In order to further explore the impact of impaired cholesterol synthesis on SEM cells, genetic inhibition of HMGCR was attempted. This has some key advantages over drug treatment – it is intrinsic to cells, so the level of inhibition does not vary with e.g. reduced CSF drug distribution compared with plasma *in vivo*; the level of inhibition within a cell does not vary over time compared with drug levels peaking after dosing then decreasing as the drug is metabolised; genetic inhibition can be used to track effects of reduced gene expression in a single cell population within complex *in vitro* co-culture models or *in vivo* models.

5.2.4.1 CRISPr-Cas9

In the first instance, CRISPr-Cas9 knockout of HMGCR was attempted. Gene knockout was felt to provide a “cleaner” form of genetic inhibition – it is irreversible, and the level of inhibition does not vary between affected cells, or over time. This comes at the cost of increased cell toxicity. There was an attempt to ameliorate cell toxicity with supplemental mevalonate as with the simvastatin *in vitro* experiments described above (Chapter 5.2.2.3). Following on from the reduction in simvastatin toxicity in 10% FCS compared with 1% FCS in these previous experiments, the cells were grown in 40% FCS to minimise potential toxicity.

The chosen method was a “double-nickase” which requires two CRISPR guides complementary to two different parts of the target gene (HMGCR in this case) in order to provide increased specificity for the target gene, though at the cost of reduced efficiency of genetic manipulation. The chosen system had one guide linked to a puromycin-resistance gene and one guide linked to GFP so successfully transfected cells would be both GFP-positive and puromycin resistant.

In order to determine the correct dose of puromycin, a dose-finding experiment (in 20μM mevalonate and FCS 40%/60% DMEM) was performed to discover the lowest dose that would kill 100% of cells in these culture conditions at 48 hours, and a dose of 15ug/ml puromycin was chosen to select successfully transfected cells. Cells were transfected using a proprietary lipofectamine-based method described above (section 2.2.11.1). These were carried out using a range of amounts of plasmid DNA and volumes of transfection solution volumes.

Though there was some evidence of successful transfection with GFP positivity after puromycin selection in a small numbers of cells, there was clear disruption of cell morphology and there was no outgrowth of a viable clone from these experiments (Figure 5-11). This led to the hypothesis that full knock-out of HMGCR is lethal for SEM cells *in vitro*.

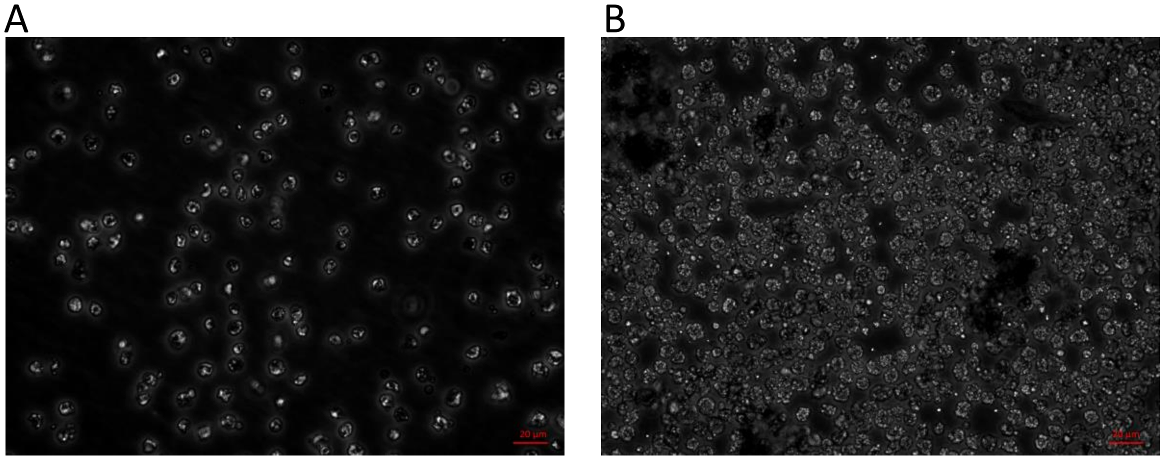


Figure 5-11 Photograph at 10x magnification showing morphology of SEM cells after puromycin selection; A – scramble control; B – HMGCR knockout.

5.2.4.2 Lentiviral transfection of shRNA

Following the failure to produce a stable knockout of HMGCR using CRISPr-Cas9 technology, it was decided to attempt an shRNA knockdown instead. This had two main advantages: lentiviral transfection of shRNA is a mature technology with consistent transfection efficiency; in contrast to the complete loss of gene function with CRISPr knockout, gene silencing with shRNA provides only partial loss of function, potentially ameliorating toxicity of loss of HMGCR.

The methods are detailed above (section 2.2.11.2), but in essence the plan was to use a PLKO-puromycin resistance plasmid together with packaging plasmids to produce lentiviral particles capable of selectively silencing HMGCR RNA. Supplemental mevalonate was used to attempt to reduce the toxicity of HMGCR knockdown.

Puromycin dosing was again determined experimentally (this time in 10% FCS), and a dose of 5ug/ml chosen for selection (note that the CRISPR experiments were performed in 40% FCS, and these in 10% FCS, which is the most likely reason for reduced dose requirement of puromycin), and serial cell counts were used to track the emergence of any puromycin-resistant clones. These results show emergence of successfully transfected puromycin-resistant clones from scramble-shRNA constructs in 3 of the 4 experiments, but none from HMGCR-shRNA tranfesction(Figure 5-12).

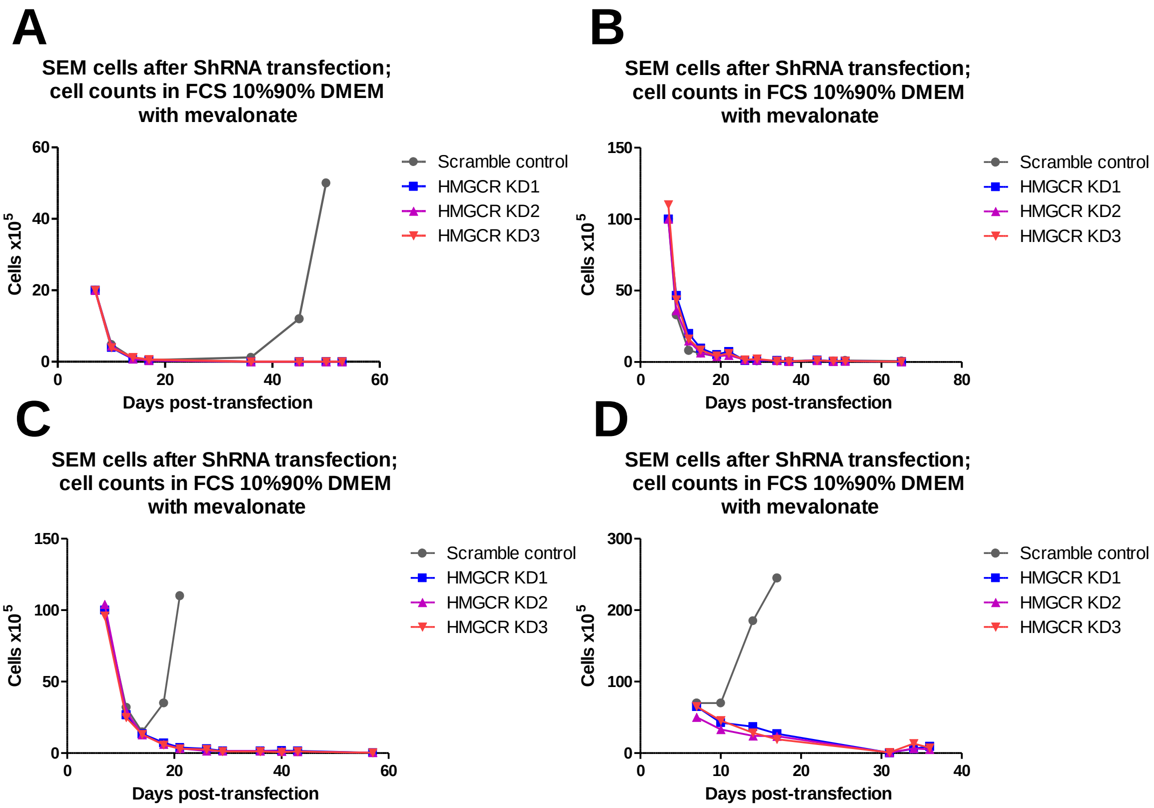


Figure 5-12 SEM cell counts following lentiviral transfection with HMGCR-shRNA; A – experiment 1; B – experiment 2; C – experiment 3; D – experiment 4.

5.2.4.3 Summary

In 12 attempts at transfection HMGCR-KD, and 12 attempts with HMGCR-shRNA, there was no evidence of a viable puromycin-resistant clone. It seems likely that genetic inhibition of HMGCR is too toxic to support viable clones despite mevalonate supplementation.

5.2.5 Effect of stromal cell co-culture on cell expansion and cell death

The final series of *in vitro* experiments were an examination of the effect of co-culture of SEM cells and REH cells with either human bone-marrow derived stromal cells (HS5 cell line) or human meningeal cells (primary cells). Whilst not directly metabolic, the effect on meningeal stroma on cell survival in the CNS *in vivo* was of interest to this project. These experiments were carried out similarly to the experiments above with cell proliferation and viability assessed by flow cytometry. In these experiments, the absence of cell-proliferation dye e450 was used to exclude stromal cells from flow cytometry analysis. Clearly this is not possible to do with cell counts, which were not recorded for these experiments. These experiments showed that meningeal cell co-culture was associated with increased cell proliferation in both cell lines tested compared to HS5 co-culture, and that HMen co-culture but not HS5 co-culture was associated with increased proliferation compared to suspension cells in SEM. In addition, HS5 co-culture but not HMen co-culture was associated with reduced cell viability, though this only reached statistical significance in SEM cells (Figure 5-13).

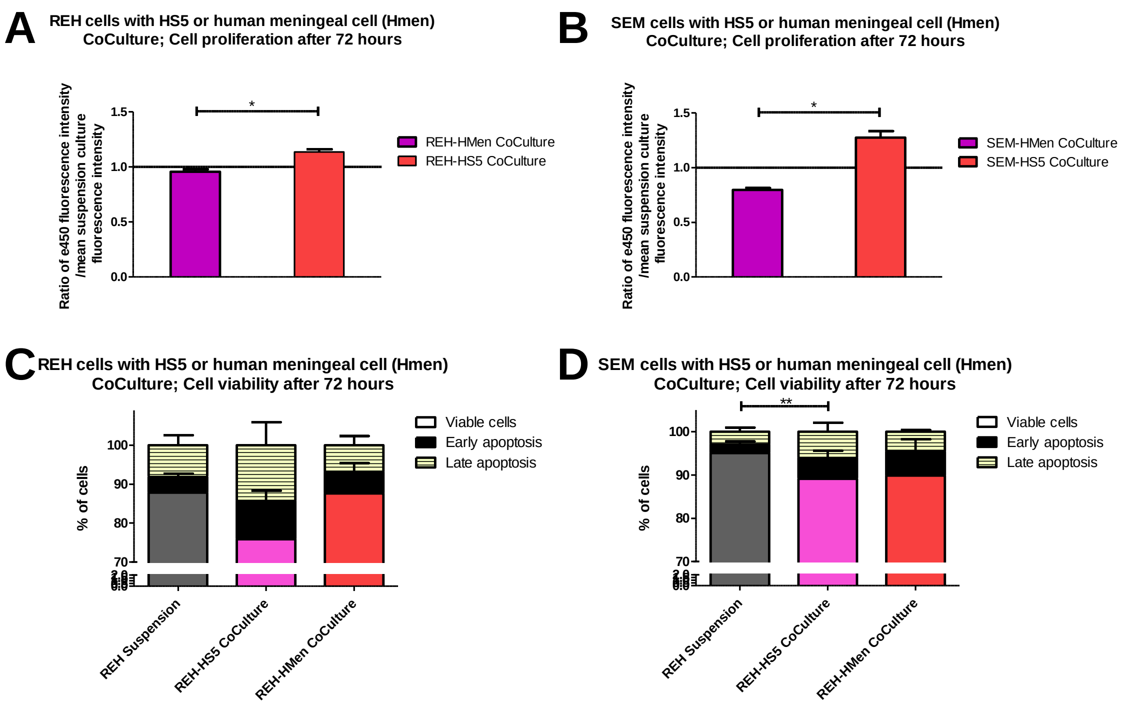


Figure 5-13 REH and SEM cell proliferation viability after co-culture with human bone marrow stromal cells (HS5, cell line) or human meningeal stromal cells (HMen, primary cells)for 72 hours. A, C – REH cells; B, D – SEM cells; A, B – proliferation (compared to cells in suspension, MFI of suspension cells noted by vertical line); C, D – cell viability and apoptosis.n=3, Mean + SEM shown. P-values denoted: *<0.05, **<0.01.

5.3 *In vivo* drug treatment

Returning to the theme of cholesterol metabolism, the next step in was to investigate of the effect of simvastatin therapy on CNS leukaemia burden of NSG mice injected with SEM cells.

In previous experiments it had been demonstrated that SEM cells upregulate expression of cholesterol biosynthesis genes – including HMGCR - in the CNS compared to systemic leukaemia (section 3.2). In addition it has been shown that interfering with HMGCR using simvastatin is toxic to SEM cell *in vitro*, particularly under low-serum conditions (section 5.2.2.2).

There has been previous experiments published attempting to achieve in-vivo synergy with statin and chemotherapy in T-ALL xenograft models, which showed no effect of statins, potentially because of inadequate dose levels(Samuels et al. 2014).

5.3.1 Simvastatin *in vivo* dose escalation

The therapeutic impact of simvastatin specifically on CNS leukaemia *in vivo* was investigated. To do this, first the highest tolerated dose of simvastatin was determined experimentally. This was done by a dose-escalation experiment carried out in 4 NSG mice over 37 days. Simvastatin was suspended in 0.5% methylcellulose/5% DMSO and the mixture sonicated until a fine slurry was formed as described previously (section 2.2.10).

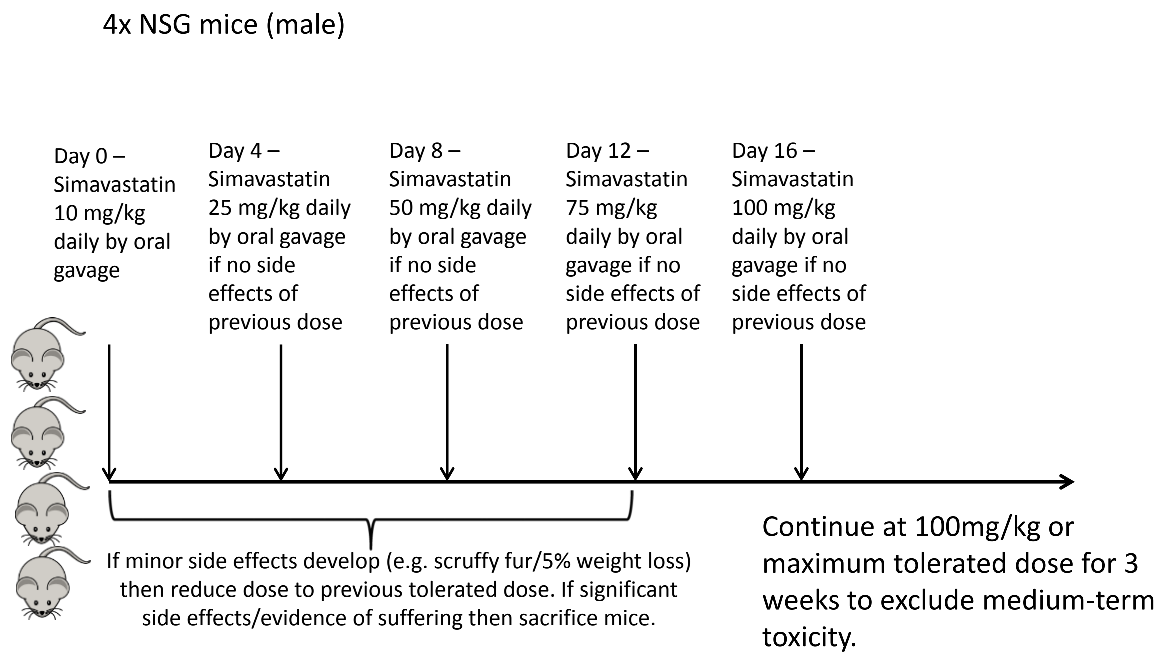


Figure 5-14 Experimental plan of simvastatin *in vivo* dose finding experiment

There was no evidence of any behavioural changes in the mice on simvastatin, and no evidence of weight loss. At post-mortem there were no gross abnormalities of any major organs.

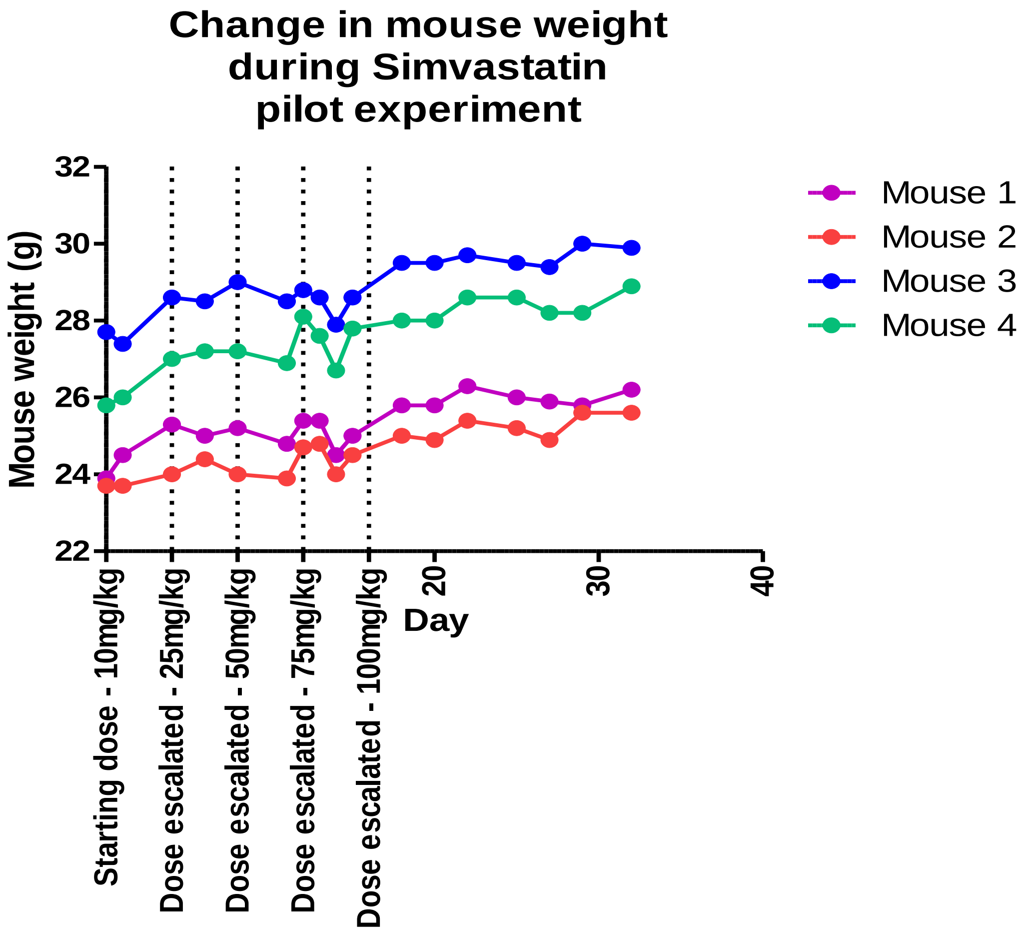


Figure 5-15 Chart of mouse weights during Simvastatin pilot experiment.

With these results, a dose of 100mg/kg simvastatin (approximately 10x the equivalent maximum adult dose in humans) was chosen. This was to ensure the maximum penetrance into the CNS compartment. Simvastatin was initially selected as the best statin for this work due to its relatively high CNS penetrance, though this is still only a fraction of the plasma level. There is evidence that high doses of statins are required to have anti-leukaemia impact(Samuels et al. 2014).

5.3.2 Experimental plan

To assess the impact of simvastatin therapy on CNS infiltration with leukaemia, 12 mice (6 male/6 female) were injected with SEM cells as described above (section 2.2.2.1). After 10 days, to allow engraftment of the leukaemic cells, simvastatin 100mg/kg or vehicle control were given daily by oral gavage until day 28 or mice developed clinical signs of leukaemia (e.g. weight loss/hindlimb paralysis). The mice were culled, spleen weights recorded, femurs and skulls harvested for histology, and spleens recovered for molecular analysis of the SEM cells.

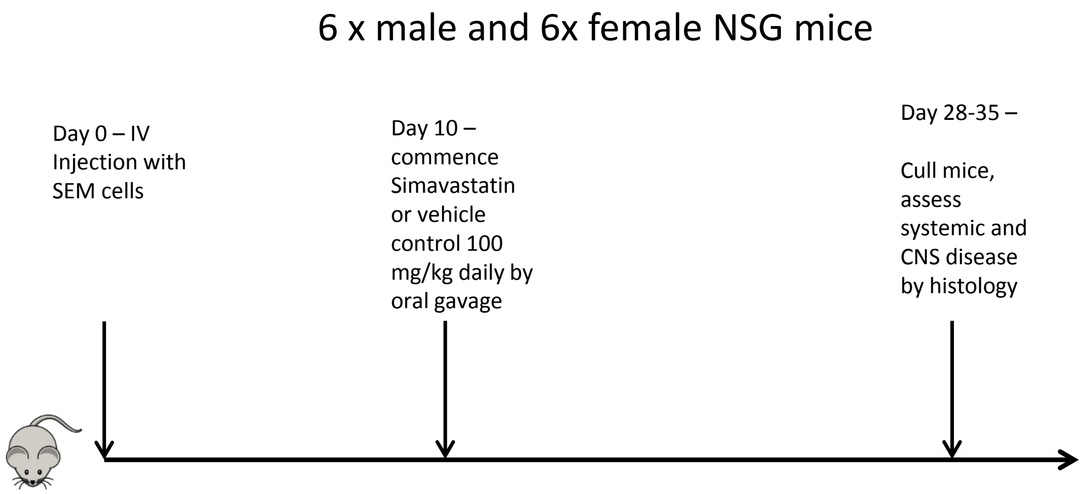


Figure 5-16 Experimental plan of simvastatin *in vivo* dose finding experiment

5.3.3 Results

The amount of CNS infiltration of the leptomeninges was estimated by quantifying the area of disease on histology sections fixed and stained as described above (section 2.2.2.6) in 5-6 sections of the skull per mouse, taken at standard points: forebrain, early mid-brain, midbrain, late-midbrain and hindbrain. Slides were scanned and the area of CNS involvement quantified using HALO software (Indica Labs Inc.) to measure the area of CNS ALL on each slide. For each mouse a mean area/section was calculated and used for analysis. Alternative methods of quantification e.g. a total area (corrected for number of sections), or excluding forebrain sections (which have a smaller cross-sectional area) could be used, but provided very similar results (Figure 5-17). It appears from these results that, contrary to expectation, simvastatin treatment in male mice increased the amount of CNS leukaemia by histology.

There were no significant differences in spleen weights between simvastatin-treated and control males or females, and the increase in CNS disease burden remained statistically significant when corrected for spleen weight (a surrogate marker for overall leukaemic burden in the mouse), suggesting that this may be a CNS-specific finding.

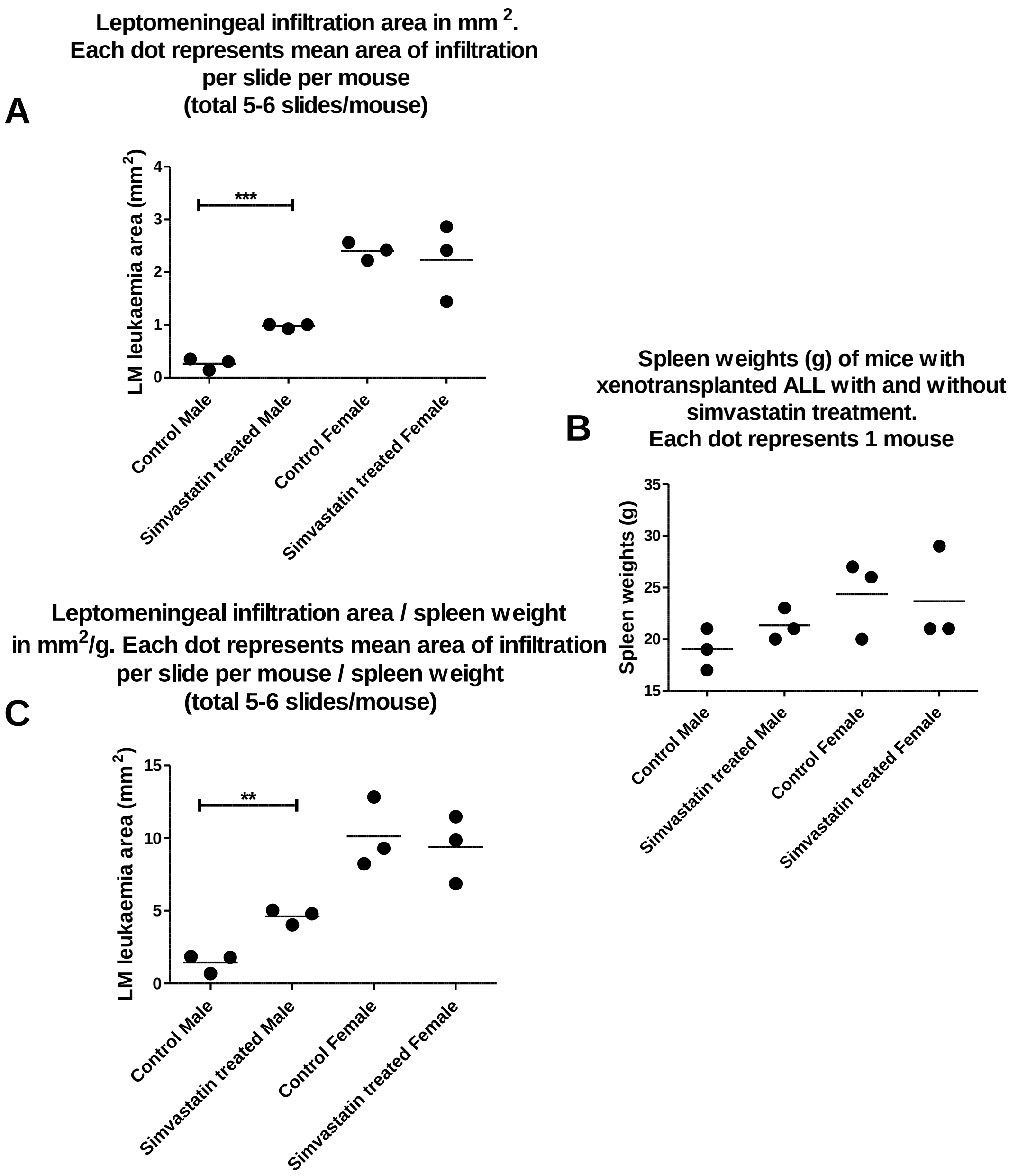


Figure 5-17 Impact of Simvastatin treatment on leukaemic infiltration of the CNS and spleen of NSG mice, A – CNS infiltration as quantified by mean area of leukaemic infiltration in 5-6 slides per mouse; B – spleen weights at end of experiment; C – CNS infiltration corrected for spleen weight.

5.4 Lipidomic Analysis

5.4.1 Pilot analysis

In addition to the LC-MS analysis above (Chapter 4:), an attempt was made to analyse the lipid metabolome of CSF from children with and without ALL. This was based on the findings detailed previously (section 3.2) that at a transcriptional level there was significant differences in lipid metabolism in ALL cells from the CNS and from the spleen. To do this two paired techniques were used: GC-MS for detection and quantification of cholesterol, and lipid-LC-MS for analysis of fatty acid:cholesterol esters (described previously in section 2.2.6).

A major challenge for this analysis was the extremely low abundance of lipids in normal CSF, though clearly this increases the potential for lipid metabolism to be a critical factor for CNS ALL. After some pilot experiments, 100µL of CSF was found to contain around 2 nmoles of cholesterol - sufficient to be detected by GC-MS.

In these pilot experiments there were some indications that lipidomic analysis of CSF may be a useful tool for analysing differences between leukaemic and non-leukaemic CSF. In particular there a particular reduction in the abundance of monounsaturated cholesterol ester cholesterol:oleic acid (Figure 5-18) in CSF taken at diagnosis from children with ALL compared with unpaired controls. This may be particularly relevant given the previous findings of changes in gene expression of SCD – which codes for the enzyme that converts saturated fatty acids to monounsaturated fatty acids (e.g. stearic acid (C18:0) to oleic acid (C18:1)) - described in Chapter 3:.

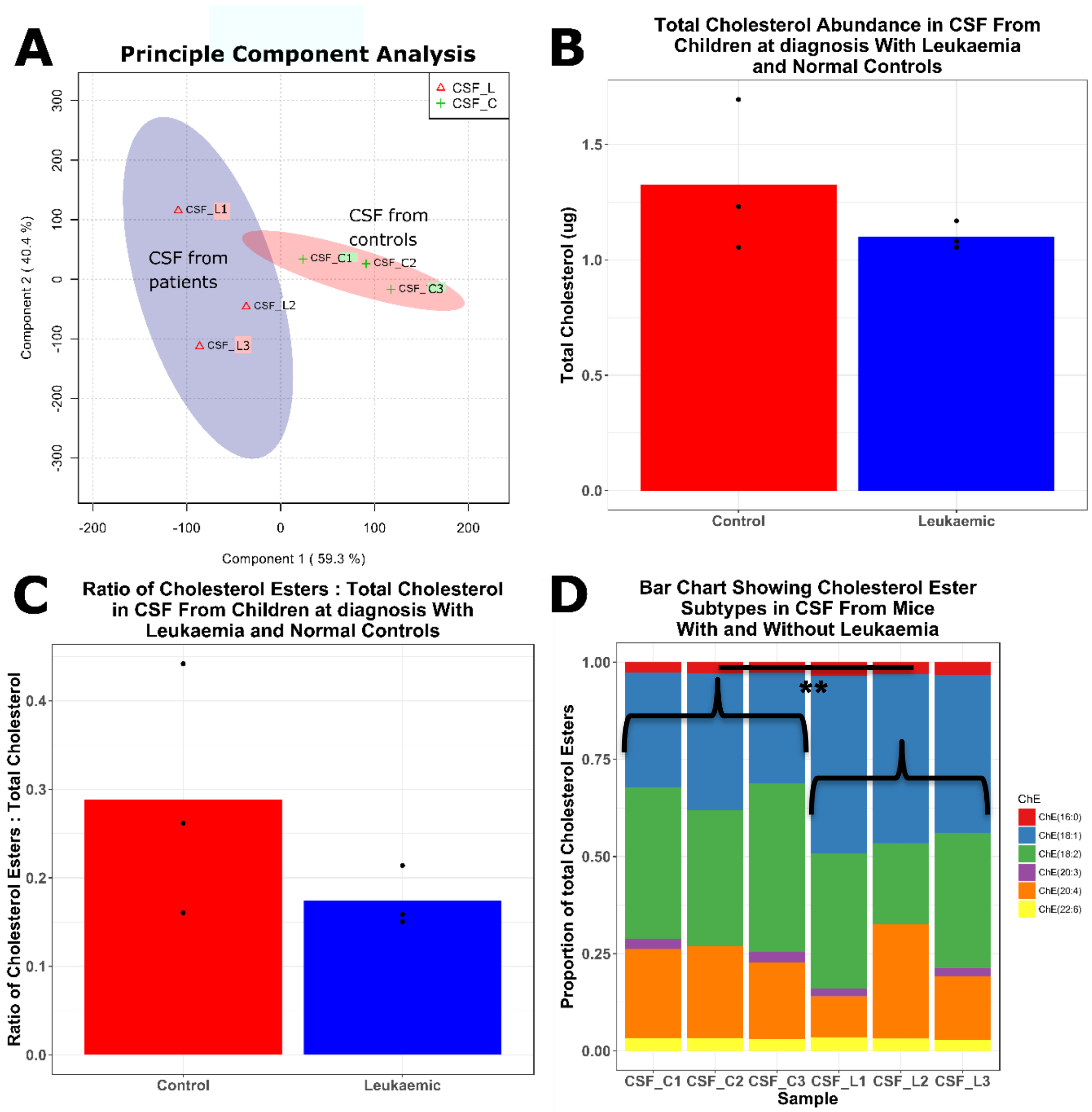


Figure 5-18 Pilot GC-MS analysis of CSF from children with ALL and normal controls. A – Principle component analysis; B – total cholesterol abundance in CSF; C – Ratio of cholesterol esters to total cholesterol; D – barcharts showing the cholesterol ester subtypes making up the cholesterol ester pool in each sample. n=3 in each group. "CSF_C#" denotes control samples and "CSF-L#" samples from leukaemic children

5.4.2CSF lipidomic analysis

To further investigate these findings three approaches were taken. Firstly, as the abundance of cholesterol were approaching the sensitivity of the assay at the amount of CSF that was available, a different technique using an enzymatic fluorometric technique to quantify free and total cholesterol was used. This allowed better quantification of cholesterol, but did not allow for the more subtle measurements of cholesterol ester subtypes. Secondly, the samples left over from the untargeted LC-MS analysis above (section 4.3) – i.e. 20 samples for children at diagnosis with ALL, 19 from the same children on maintenance chemotherapy, and 18 unmatched controls – were prepared for GC-MS and lipid-LC-MS analysis as above. Finally, small polar metabolites associated with lipid metabolism in the CSF and plasma of mice with and without leukaemia were analysed using standard LC-MS (as above).

5.4.2.1 Enzymatic analysis of cholesterol

The enzymatic analysis was carried out on CSF and plasma from NSG mice with (n=6) and without (n=3) ALL (xenografted with SEM cells as before, section 2.2.2.1), and on CSF from patients at diagnosis (n=5), the same patients on maintenance therapy (n=5) and on paired CSF and plasma from normal controls (n=5). The technique is described in detail previously (section 2.2.6.6), but in brief, cholesterol was oxidised, and the resulting H_2O_2 caused a detectible fluorescence of a marker compound. Unfortunately due to the tiny volumes involved, CSF results were only available for 3 of the 6 leukaemic mice, and there was enough sample for measurement of only total cholesterol and not cholesterol esters in mouse CSF. There was sufficient plasma for measurement in 5 of the 6 leukaemic mice.

Results showed a clear reduction in cholesterol abundance in plasma from mice with leukaemia compared with controls, and a trend towards reduced cholesterol in the CSF of leukaemic mice (p-value=0.083). In CSF from children with leukaemia there was a reduction in total cholesterol in diagnostic samples compared with samples later in treatment. Looking in more depth, there was no difference in free cholesterol, but a decrease in cholesterol ester abundance in diagnostic CSF from children with leukaemia compared with the same children on maintenance chemotherapy, and a trend towards the same finding comparing diagnostic CSF with normal controls (p-value=0.054) (Figure 5-19).

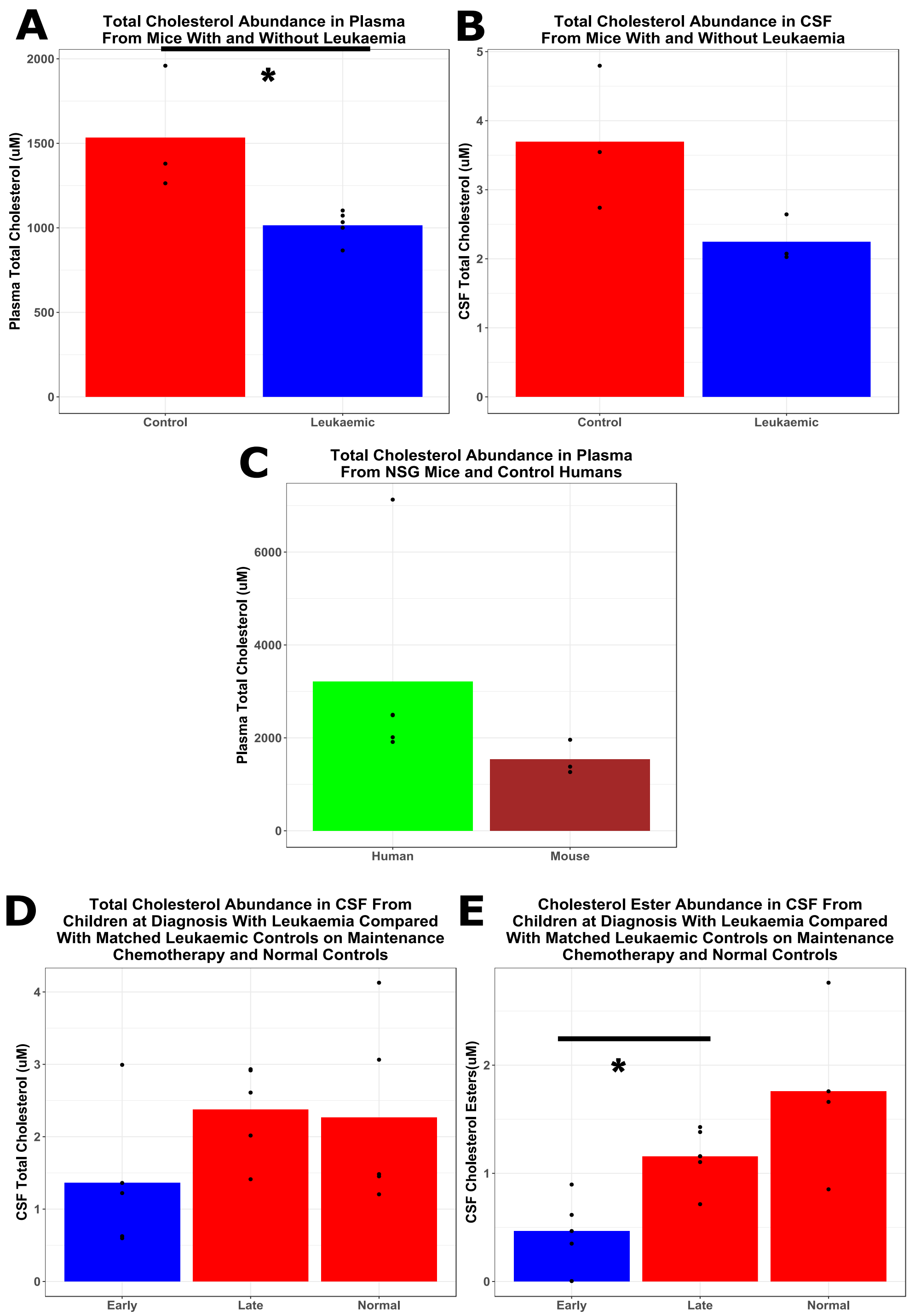


Figure 5-19 Cholesterol abundance in CSF and plasma of mice with and without ALL, and children with and without leukaemia determined using Amplex™ Red enzymatic cholesterol quantification. A – total cholesterol abundance in blood plasma from mice with and without ALL, n=3(control) and 5(leukaemic). p=0.020; B – total cholesterol abundance in CSF from mice with and without ALL, n=3. p=0.083; C – total cholesterol abundance in non-leukaemic NSG mice and control human plasma n=3 (murine) and 4 (human). ; D – total cholesterol abundance in CSF from children at diagnosis "Early", later in ALL therapy "Late", and normal controls "Normal" n=5. Differences did not reach statistical significance; E – cholesterol ester abundance using the same groups as D. p-value 0.016 Early vs Late, 0.054 Early vs Normal.

5.4.2.2 LC-MS and GC-MS analysis

For the second part of this experiment, CSF samples from children with ALL at diagnosis, the same children on maintenance chemotherapy, and from unmatched control were prepared and analysed using GC-MS and lipidomic LC-MS as above. Unfortunately the abundance of lipids in this analysis was insufficient to allow detailed comparisons. The overall abundance of cholesterol appears from this analysis to be largely unchanged between groups. Interestingly, in contrast with the results above, there appears to be a very modest increase in cholesterol abundance in the CSF from children at ALL diagnosis – this reached statistical significance in comparison with the unmatched control CSF, but is of uncertain biological significance given the issues with low lipid abundance in samples. Analysis of cholesterol esters unfortunately was hampered by low abundance with several results below the limits of detection, making comparisons between groups difficult.

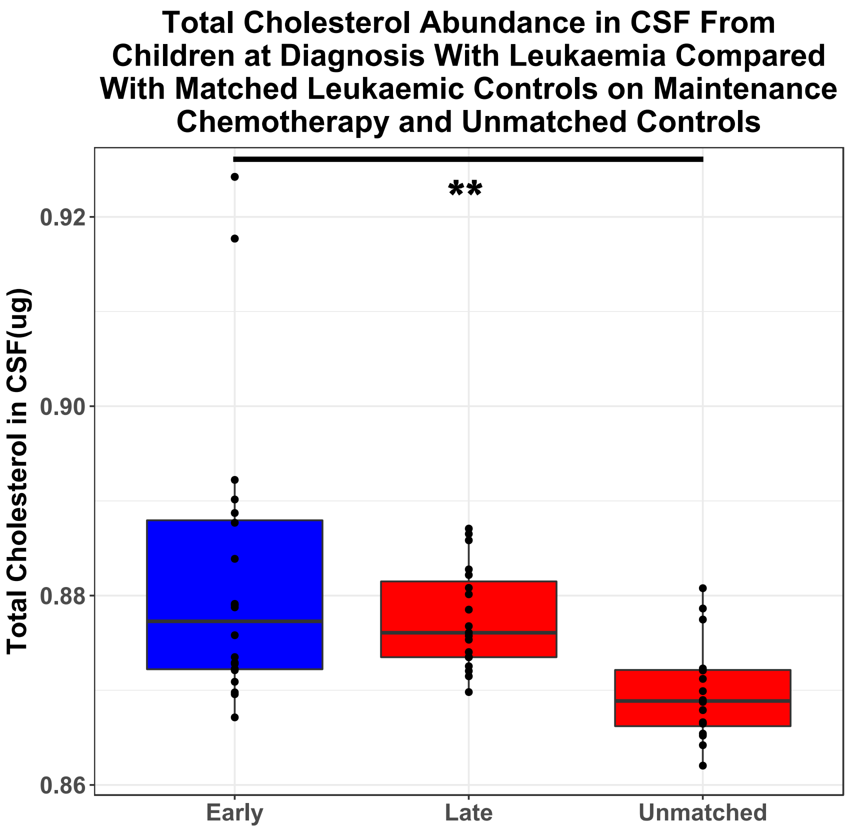


Figure 5-20 Total cholesterol in CSF from children at diagnosis with ALL compared with the same children later in therapy and normal controls. Cholesterol measured using GC-MS. n=20 (Early), n=19 (Late), n=17 (Normal). p=0.004 Early vs Normal

Finally, CSF and plasma from mice with and without xenografted SEM cell leukaemia was analysed using “targeted” LC-MS. Mice were xenotransplanted with SEM human ALL cell line cells as described previously (section 2.2.1.1), and metabolic analysis carried out using LC-MS as described previously (section 2.2.6.3).To assess cholesterol metabolism, the only metabolite it was possible to measure was mevalonate (described previously, section 5.2.2.2), the first committed precursor to cholesterol, and a small polar molecule amenable to LC-MS analysis. No difference was detected in mevalonate abundance in the CSF between mice with and without leukaemia in CSF or plasma.

To assess fatty acid saturation, the ratio of saturated to monounsaturated C16 (palmitate/palmitoleate) and C18 (stearate/oleate) fatty acids were assessed in CSF and plasma in mice with and without leukaemia. Interestingly, there was a shift in the ratio towards saturated fatty acids in the CSF compared to plasma in mice with or without leukaemia, and in mice with leukaemia this ratio was further shifted for C16 but not C18 fatty acids. There were no statistically significant changes in fatty acids in the CSF or plasma, though there was a trend to increased palmitate and reduced palmitoleate in the CSF (p-value=0.072, Figure 5-21).

Looking more closely at the ratio of saturated:unsaturated fatty acids in the CSF, there was an increased ratio (i.e. a relatively high proportion of saturated fatty acids) of palmitate:palmitoleate in CSF but not plasma from mice with leukaemia. There were no significant differences in stearate:oleate. When this ratio is compared between CSF and plasma, there is a CSF-specific skewing of the ratio towards increased saturated fatty acids in leukaemic but not control mice (Figure 5-22).

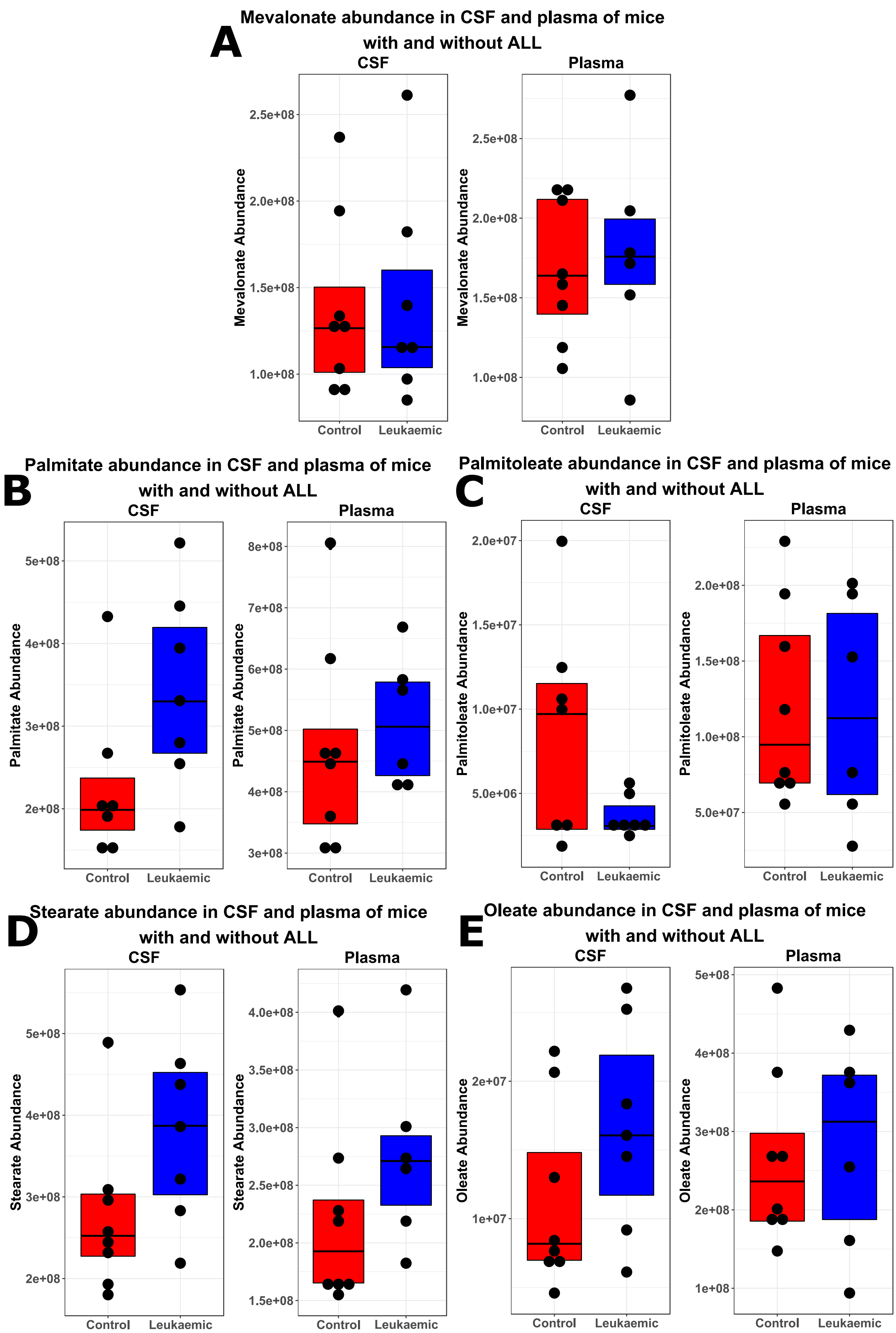


Figure 5-21 Comparison of mevalonate and fatty acid abundance in CSF and plasma of mice with and without leukaemia assessed using LC-MS. A – mevalonate; B – palmitate (p-value leukaemic vs control 0.075); C – palmitoleate (p-value leukaemic vs control 0.072); D – stearate; E - oleate

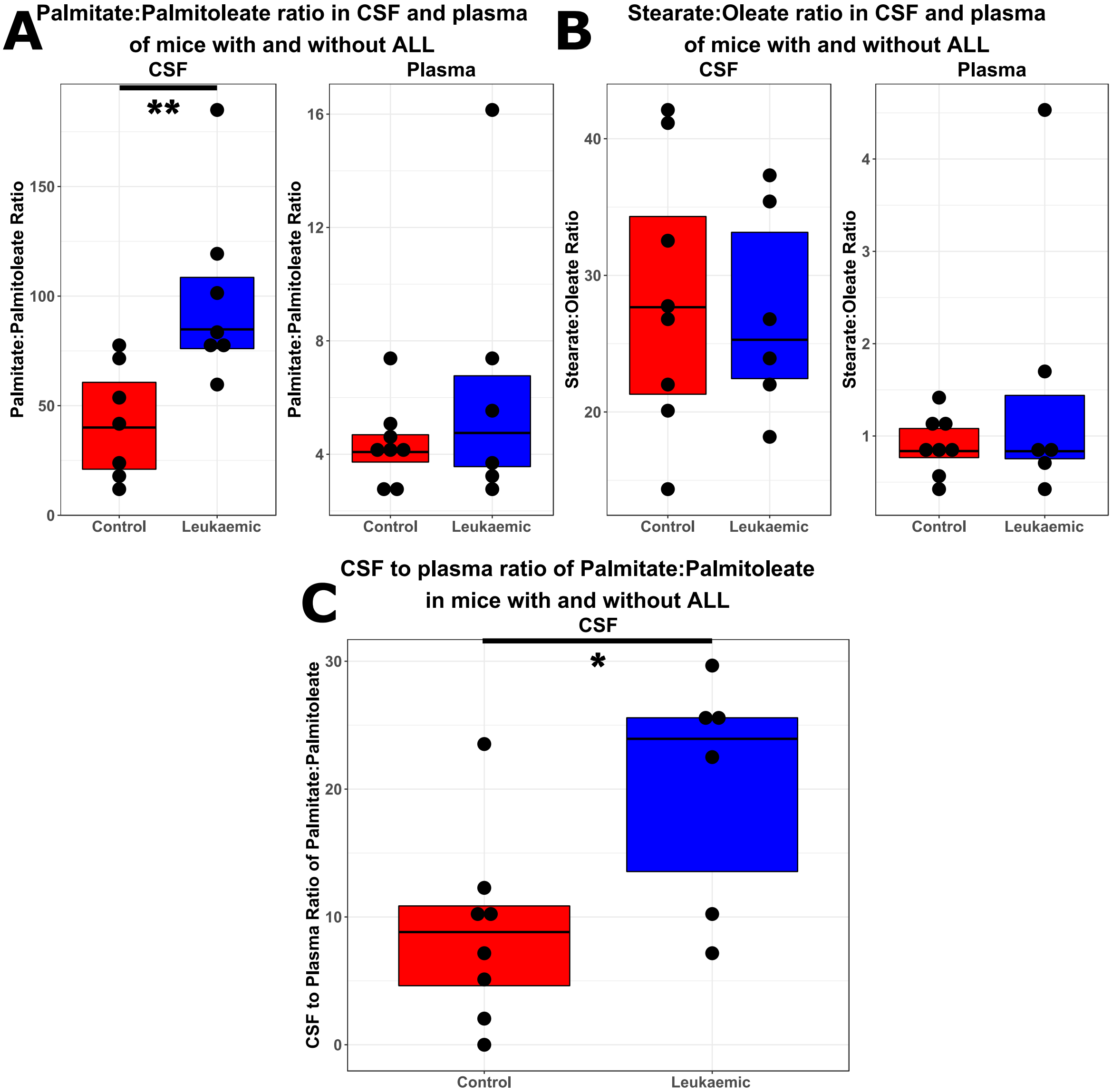


Figure 5-22 Comparison of saturated:monounsaturated C16 and C18 fatty acid abundance in CSF and plasma of mice with and without leukaemia assessed using LC-MS. A: palmitate:palmitoleate (p=0.008); B – stearate:oleate; C – CSF to plasma ratio of palmitate:palmitoleate (p=0.024)

With this clear skewing of the saturated:monounsaturated C16 fatty acid ratio in the CSF of mice with CNS leukaemia, the next step was to look in human CSF. Data from same CSF samples from children with CNS relapse with ALL, and children at diagnosis and later on in therapy was used as previously using LC-MS analysis (chapter 4.5). When a similar palmitate:palmitoleate ratio was assessed in this human data, however, there was no clear evidence of a skew in the ratio either at diagnosis or at the time of CNS relapse (Figure 5-23).

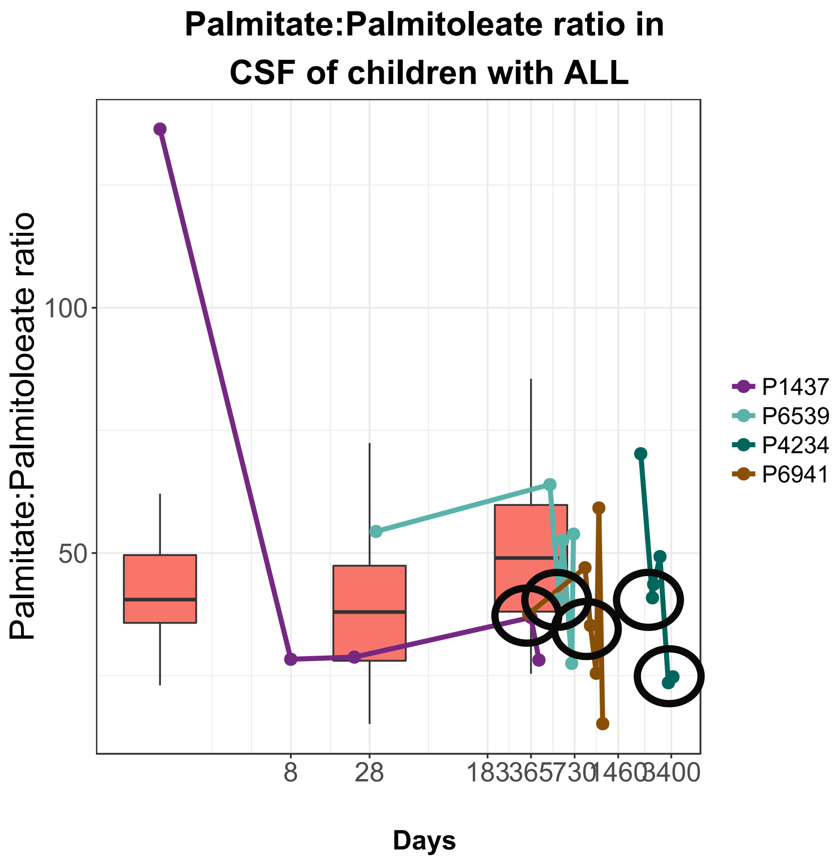


Figure 5-23 CSF Abundance of palmitate:palmitoleate ratio as a timeline in 4 children who suffered CNS relapse of ALL. Boxplots show median abundance (line), quartiles (box) and range (whiskers, outlier as dot) in CSF from patients with ALL who did not suffer CNS relapse at days 1, 28 and 365 post-diagnosis. Sample taken closest to isolated CNS relapse for each child denoted by circles. Note time is on a log axis.

5.4.3 Cellular lipid/fatty acid analysis

To follow up this work on CSF lipid analysis, the lipid and fatty acid profiles of SEM human leukaemia cells taken from the CNS of mice were compared with SEM cells taken from spleens of mice. This was to better understanding of the changes in cell metabolism between the CNS and the spleen, and how they related to the changes in the transcriptome described previously (section 3.2) and changes in the metabolome described above.

The experiment was designed as described previously (section 4.4). Briefly, 9 mice were injected with SEM cells, and culled at 28 days post-injection. Cells were retrieved, quickly pelleted and snap frozen.

Analysis was carried out using standard LC-MS, as well as GC-MS analysis of total cholesterol abundance (GC-MS analysis performed by colleague Grace McGregor).

From the LC-MS analysis, cells from the CNS had significantly higher abundance of mevalonate than cells from the spleen, in keeping with the transcriptomic changes described in Chapter 3:). On analysis of total cellular cholesterol abundance, interestingly there appears to be an increase in abundance in cells retrieved from the CNS.

In terms of free fatty acid abundance, there was an increase in saturated fatty acids palmitate (C16) and stearate (C18) in the CNS, and a corresponding shift in the ratio of saturated:monounsaturated fatty acids (Figure 5-24).

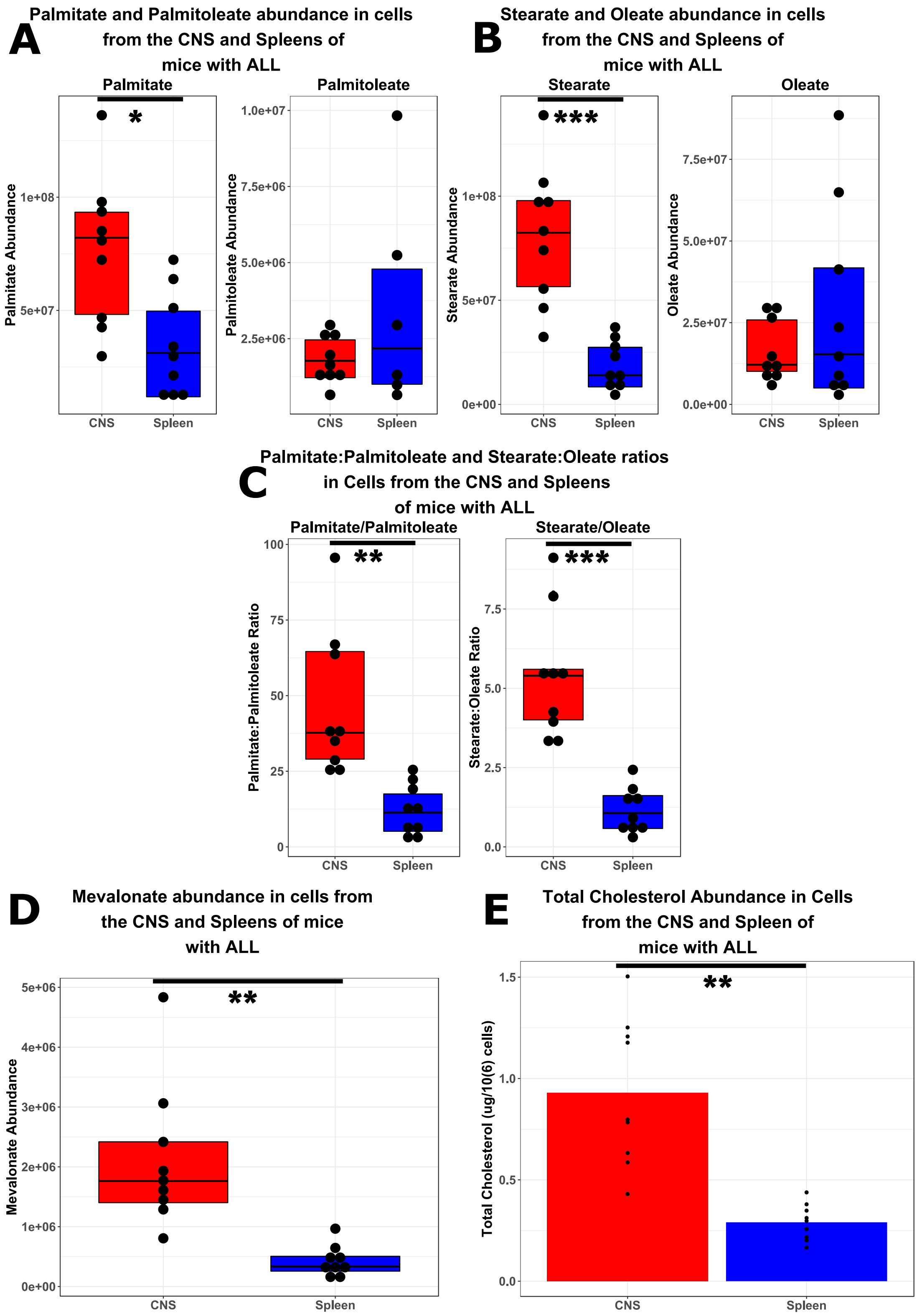


Figure 5-24 Intracellular abundance of free fatty acids, cholesterol and mevalonate of ALL cells from the CNS and spleen using LC-MS or (cholesterol) GC-MS. A – palmitate (p=0.016) and palmitoleate; B – stearate (p=0.0009) and oleate; C – C16 and C18 saturated:monounsaturated ratios (p=0.003 and p=0.0003 respectively); D – mevalonate (p=0.0028); E – total cholesterol (p=0.0013). Statistical analysis carried out using paired student's t-test.

5.4.4 Summary

Lipidomic analysis of CSF has shown reduced cholesterol esters in the CSF of mice with leukaemia and of children at diagnosis with ALL using an enzymatic technique. Conversely there was an increase in cholesterol in CSF samples from children with ALL at diagnostic compared to normal controls (in different children's samples to the previous experiment) by GC-MS analysis. Particularly there is a consistent increase in the saturated:unsaturated fatty acid and cholesterol ester ratios in the CSF of children at diagnosis. This did not translate into an increase in the palmitate:palmitoleate ratio in children with CNS relapse.

Lipid/fatty acid analysis of cellular metabolism again showed a skewing of the saturated:unsaturated fatty acid ratio. In addition there was a significant increase in mevalonate and an increase in cellular cholesterol – in contrast to what might be expected given the low abundance of cholesterol in the microenvironment.

5.5 *In vivo* metabolite tracing

5.5.1 Rationale and experimental design

Finally, to further explore the changes in cellular metabolism between systemic and CNS leukaemia, isotope-labelled metabolites were used. In brief, mice were injected with SEM human ALL cell line cells. After 28 days (i.e. after systemic and engraftment of disease), the mice were injected with isotope labelled glucose (n=9) or acetate (n=9). These labelled molecules are identical to normal metabolites except that each carbon atom in the metabolite had been replaced with C13-carbon (i.e carbon with 1 extra neutron). This has no biological/biochemical effect, but the original compound and its metabolic products can be detected using LC-MS, allowing "tracing" of the metabolism of the molecule *in vivo*.

Glucose was chosen as it is a key metabolic fuel and biomass source. As discussed previously (section 1.3.3), altered glucose metabolism was the first major discovery in cancer metabolomics. Acetate was chosen as the major alternative small molecule lipid biomass source, and recently described as a key molecule in several different tumour types.

As the kinetics of diffusion of glucose and acetate into the CSF are unknown, each of the experiments were carried out in two cohorts – in the first, 5 mice were culled 20 minutes after IV injection of glucose or acetate, then 4 mice culled after 40 minutes. Cells were retrieved from the CNS and spleen, and snap-frozen as above (section 2.2.2).

An important question at the start of the experiment design was how to control for the different amounts of red cell contamination between the two sites of interest –typically the CNS had <5% RBC contamination and the spleen 10%-30% by cell count (less by cell mass) by flow cytometry analysis (section 2.2.9.3)). The typical method used in this project was density centrifugation (section 2.2.2.4), but this takes a significant amount of time. An alternative method to achieve good red cell clearance is using rapid chemical red cell lysis (section 0) but the effect of this on the leukaemic cell metabolome is unknown. A final alternative – using systemic leukaemic cells from the bone marrow – was not felt to be useful as there was a comparable amount of murine cell contamination in samples collected from the bone marrow (Figure 5-25).

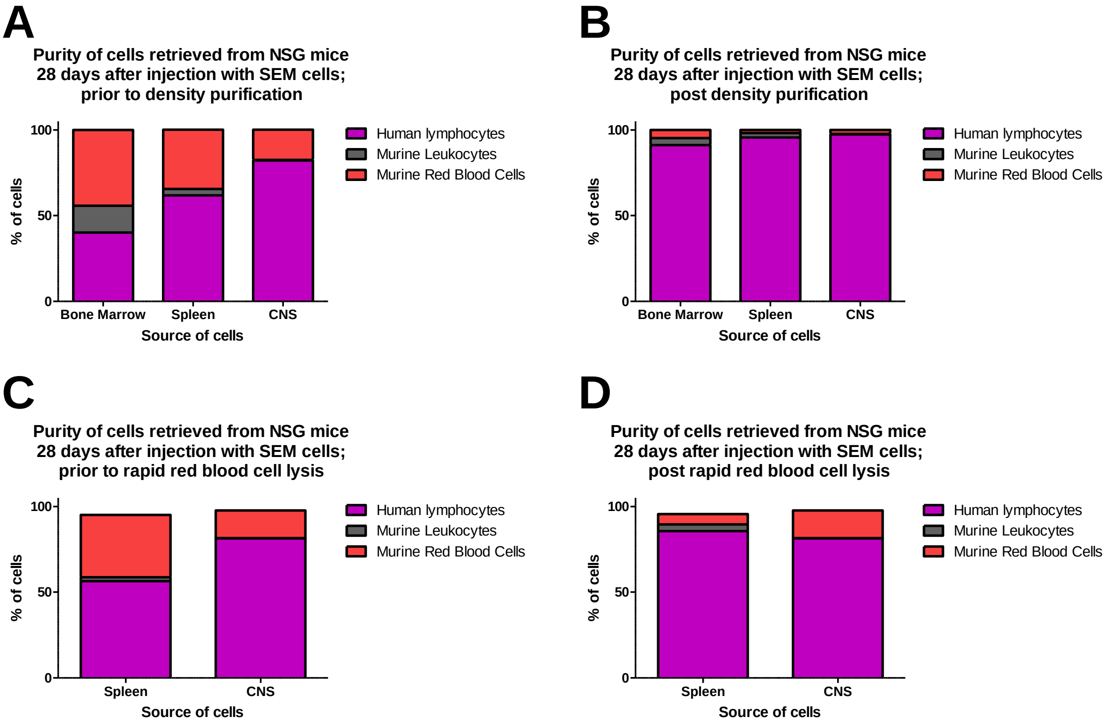


Figure 5-25 SEM cell purity after *in vivo* harvest pre- and post-purification A,B – Cell purity after harvest from bone marrow, spleen and CNS pre-(A) and post-(B) density centrifugation (n=2); C,D – Cell purity after harvest from spleen and CNS pre-(C) and post(D) rapid (3mins) red blood cell lysis (n=2). Cell types determined by flow cytometry after staining with human CD19 (SEM cells), murine CD45 (murine leukocytes), and murine Ter-119 (murine red cells).

The first part of this experiment therefore compared the metabolic profile of cells from the spleen and CNS either (1): centrifuged for 30 seconds and snap-frozen, (2) suspended for 3 minutes in RBC lysis buffer at RT then centrifuged and snap-frozen (3 minutes had been found to be sufficient to lyse >99% of RBCs, data not shown), or (3) suspended in 0.9% saline at RT then centrifuged and snap frozen in 3 mice. The results showed a dramatic shift in the metabolic profile of cells with even 3 minutes sitting in 0.9% saline at RT compared with snap-freezing, with less shift between samples incubated in RBC lysis buffer compared with samples incubated in PBS. In addition, incubation in RBC lysis buffer appears to more significantly impact the metabolic profile of cells in CNS samples despite the significantly lower RBC contamination, presumably due to a lower ALL cell number in the incubation solution and a therefore higher relative toxicity to ALL cells. Based on these results, in subsequent experiments cells retrieved for metabolic analysis were centrifuged and snap-frozen prior to extraction (Figure 5-26).

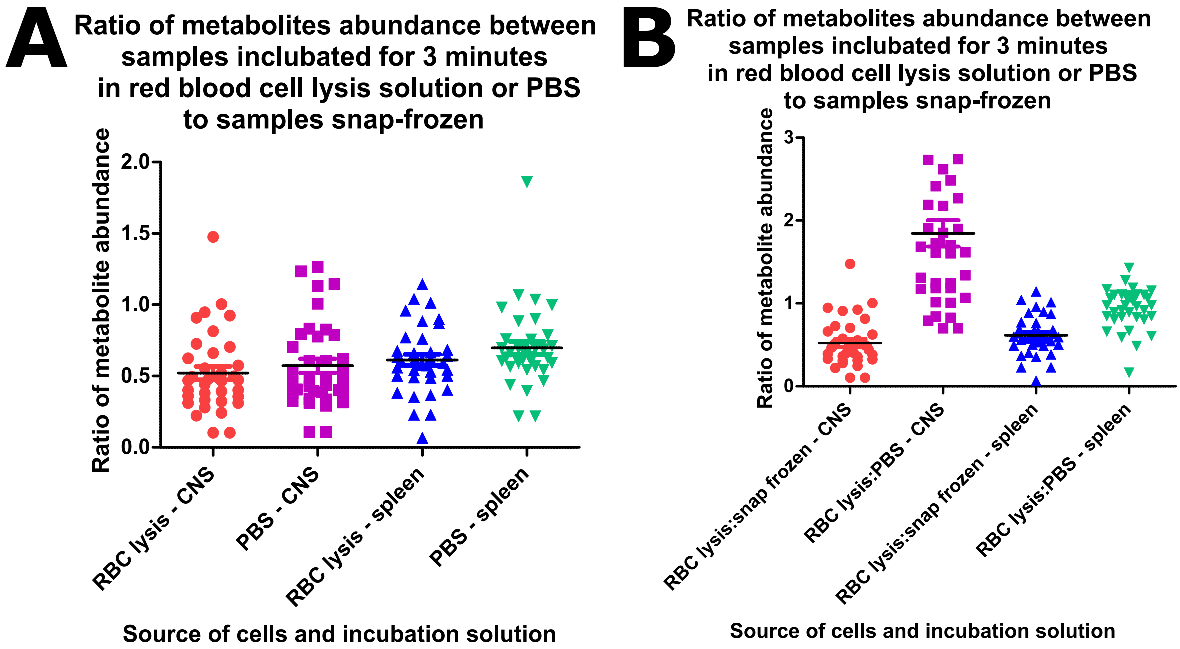


Figure 5-26 Changes in metabolite detection between identical samples snap-frozen in dry ice, and either incubated in red cell lysis buffer or PBS. Note the graphs show ratio of individual metabolites in (A) samples snap-frozen to samples either incubated in red cell lysis solution or PBS (i.e. metabolites with the same abundance in snap-frozen and incubated samples would have a ratio of 1), or (B) samples RBC lysed to samples snap-frozen or incubated in PBS.

5.5.2 Glucose tracing

From this experiment, it is clear that there is substantial glucose metabolism *in vivo* in ALL cells retrieved from both the CNS and the spleen. Unfortunately it was not possible to fully assess the proportion of glucose in CSF and plasma that was isotope-labelled due to difficulty obtaining sufficient CSF (CSF was obtained from 4 mice in the 20 minute cohort, and 2 in the 40 minute cohort), and high variability between samples of CSF, and therefore it is difficult to draw authoritative conclusions (Figure 5-27). On analysing the data however, there were some interesting findings explored below.

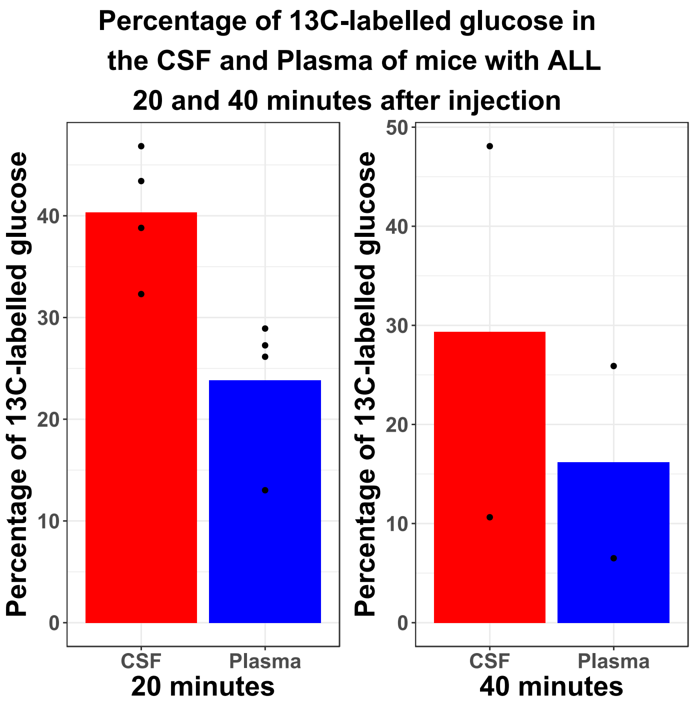


Figure 5-27 Proportion of ^{13}C-glucose vs ^{12}C-glucose labelling in the CSF and plasma of mice 20 minutes and 40 minutes after intravenous injection of ^{13}C-glucose.

5.5.2.1 Glycolysis

When the metabolites involved in glycolysis were analysed there was, perhaps surprisingly given the hypoxia signalling seen on transcriptomic analysis (Chapter 3:), little difference in ^{13}C-labelling in the CNS and the spleen, with a modest increase in the fraction of fully-^{13}C labelled pyruvate at 20 minutes, and lactate at 20 and 40 minutes the only statistically significant changes (Figure 5-28). Whilst increased glycolysis would fit with the hypoxia signalling seen on transcriptomics, these findings may simply reflect the increased fraction of labelled glucose in the CSF shown above (Figure 5-27).

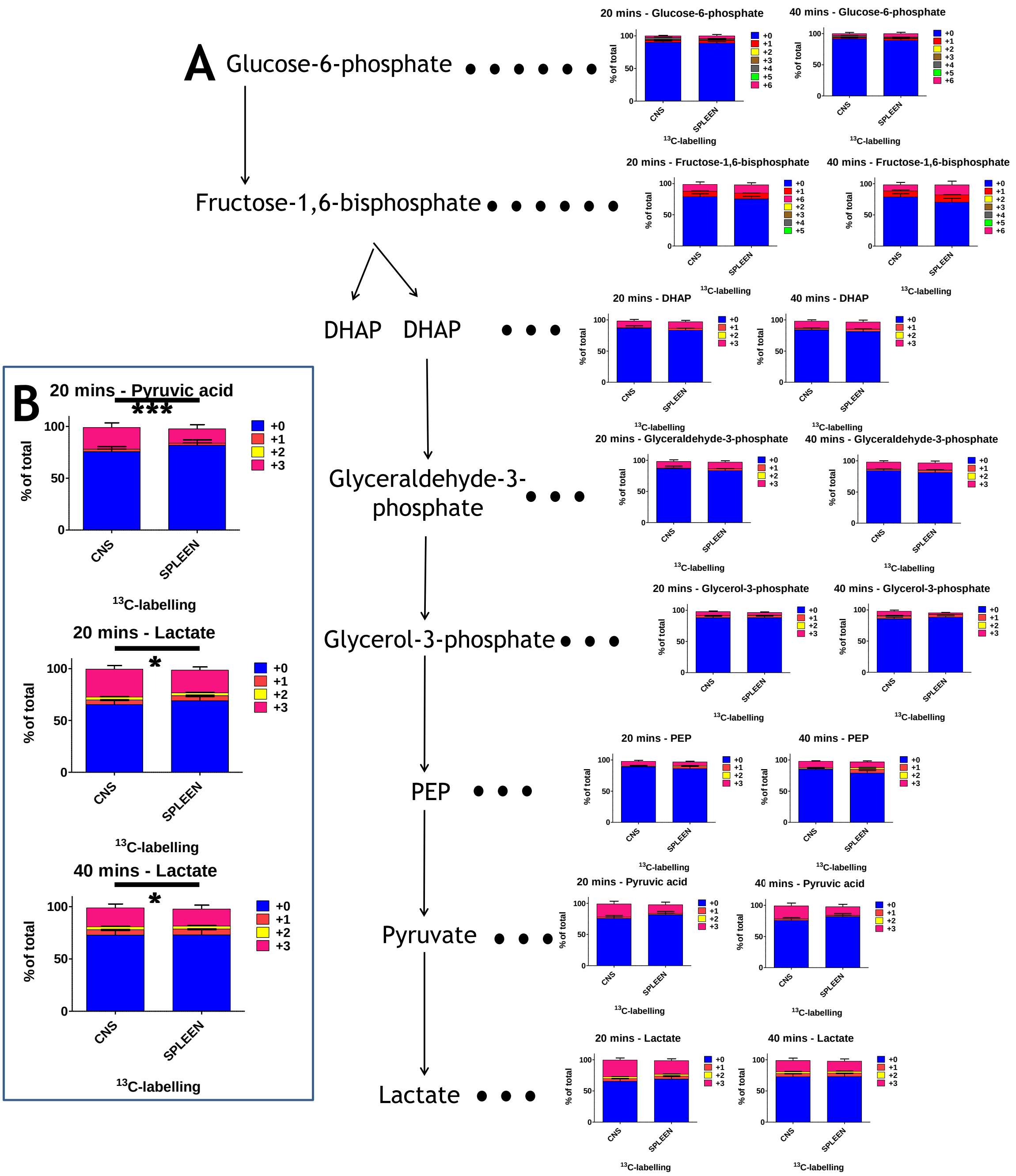


Figure 5-28 Overview of ^{13}C labelling of the glycolytic pathway metabolites in ALL cells in the CNS or spleen 20 or 40 minutes after injection with ^{13}C-glucose; colours indicate number of ^{13}Carbon atoms in the metabolites. A – overview; B – pyruvic acid labelling at 20 minutes (p=0.0005), and lactate at 20 minutes (p=0.016) and 40 minutes p=0.027) shown in more detail

5.5.2.2 TCA Cycle

Looking at the TCA cycle, there is again increased labelling of metabolites in the CNS – not in keeping with a hypoxic phenotype – particularly 40 minutes after injection with glucose (Figure 5-29). Interestingly there is an increase in ^{13}C labelling of CSF glutamine, and many of the metabolites in the TCA with significant incorporation of ^{13}C labelling are “downstream” of glutamine via AKG, and could conceivably have received labelling this way. It is worth noting that meningeal stromal cells are known to express the enzyme glutamine synthase and could be involved in maintaining glutamine homeostasis in the CSF, and this could potentially be an alternative fuel source for CNS leukaemia cells (Figure 5-30).

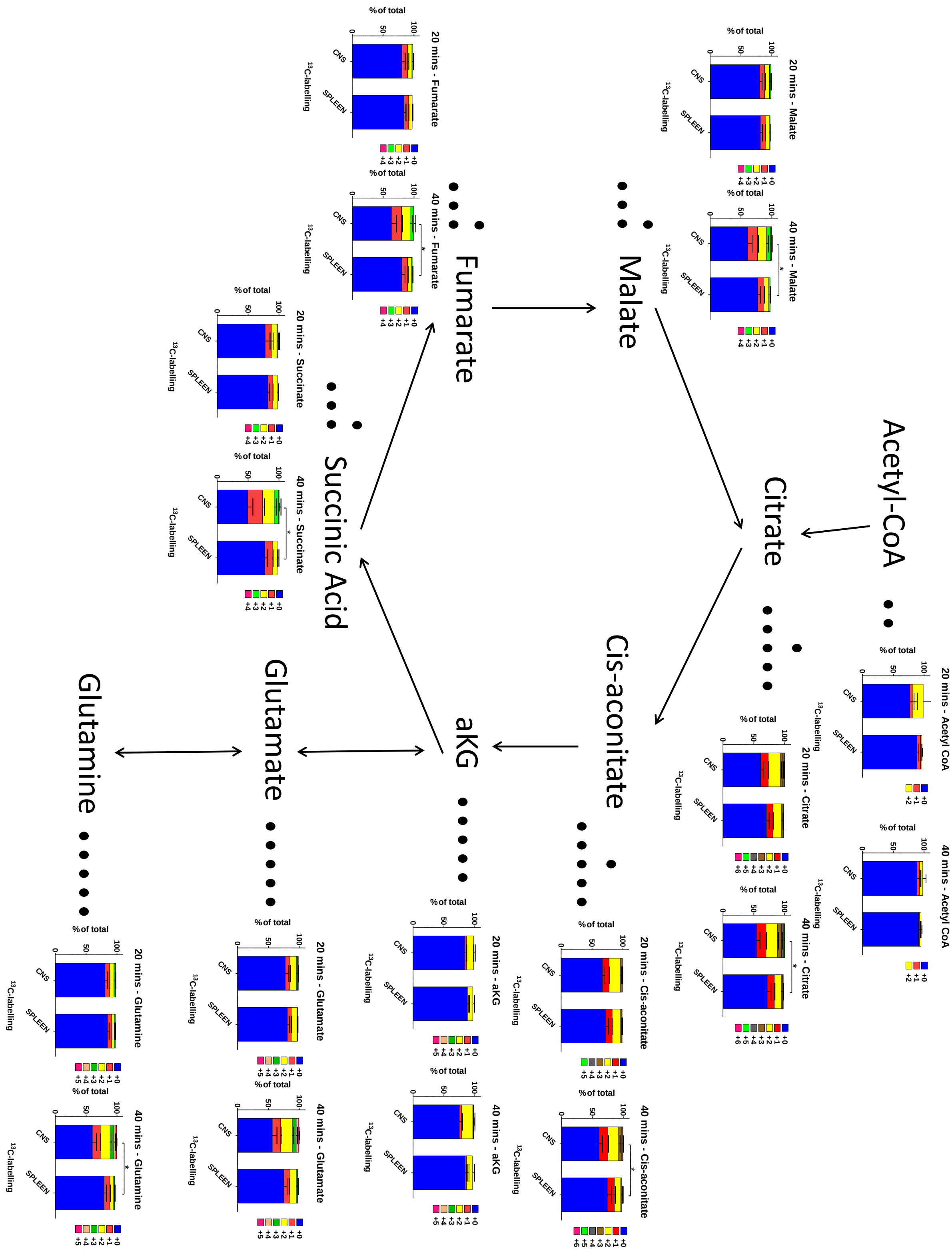


Figure 5-29 Overview of ^{13}C labelling of the TCA pathway metabolites in ALL cells in the CNS or spleen 20 or 40 minutes after injection with ^{13}C-glucose schematic dots represent carbon atoms in metabolites. P-values denoted: *<0.05.

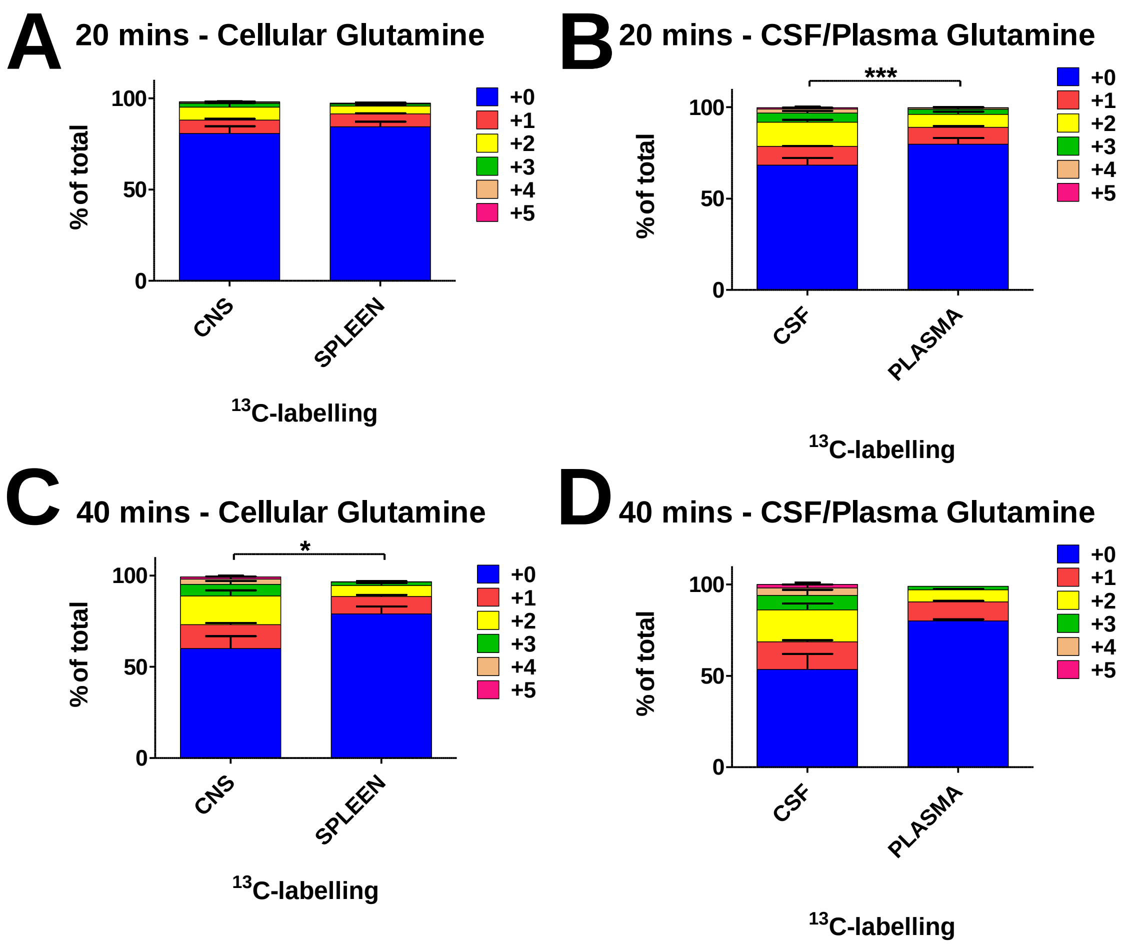


Figure 5-30 Glutamine labelling with ^{13}C in ALL cells from the CNS and spleen and from CSF and plasma 20 (glutamate) or 40 (glutamine) minutes after injection with ^{13}C-glucose. A – intracellular glutamine 20 minutes after ^{13}C injection; B – CSF/plasma glutamine 20 minutes after ^{13}C injection (p=0.0004); C - intracellular glutamine 40 minutes after ^{13}C injection (p=0.036); D – CSF/plasma glutamine 40 minutes after ^{13}C injection (n=2 so no statistical analysis performed).

5.5.2.3 Summary

These data have shown evidence that ALL cells in the CNS have increased glycolytic flux, and evidence of conversion of glucose to glutamine in this microenvironment. There is evidence of ongoing TCA activity despite the low-glucose, low-lipid, and low-oxygen microenvironment.

5.5.3 Acetate tracing

From analysing data after C13-acetate injection, there was very little tracing found. There was no evidence of labelled acetate incorporated into fatty acids, and none into TCA metabolites (Figure 5-31), and none into cholesterol (analysis by colleague Grace MacGregor, data not shown).

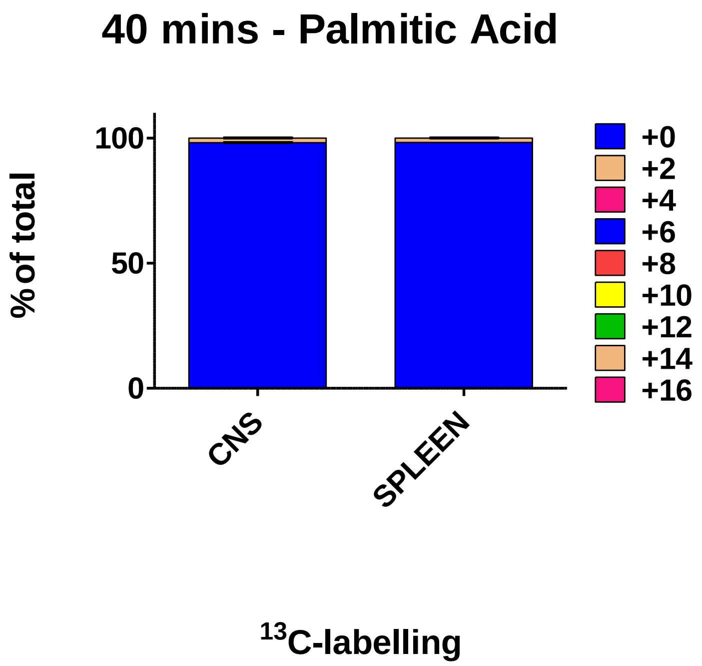


Figure 5-31 Fraction of ^{13}C-incorporated into palmitic acid 40 minutes after injection of IV fully ^{13}C-labelled acetate *in vivo*

5.6 Conclusions

In this chapter it has been demonstrated inhibiting cholesterol synthesis with simvastatin has a clear impact *in vitro* and *in vivo*. Simvastatin treatment is effective in killing ALL cells *in vitro*, particularly in a low-nutrient environment. *In vivo*, high-dose simvastatin treatment seems to potentiate CNS disease in the male mice tested. It is possible this impact is mediated through induction of cholesterol synthesis pathways in systemic ALL cells in the presence of simvastatin; and that this allows better survival in the CNS where the effective dose of simvastatin is markedly reduced due to incomplete penetration of the drug through the blood:brain and blood:CSF barriers.

The impact of simvastatin on cell proliferation and viability *in vitro* is profound, and the exacerbation of cytotoxicity with the addition of cholesterol interesting. The initial idea that the addition of cholesterol may cause downregulation of the cholesterol biosynthesis pathway was not borne out by PCR analysis – but it is difficult to draw firm conclusions from RNA retrieved from cells that are under significant stress. An additional complicating factor could be that cholesterol was added to the cells in a water-soluble form bound to cyclodextran – it could be the cyclodextran rather than the cholesterol itself that is at the root of this combined toxicity with simvastatin. The failure to form a viable cell knockdown for cholesterol biosynthesis was unfortunate. It may be that targeting the pathway less directly (e.g. by targeting the SREBF system) may allow the production of more viable clones though this does come at the expense of increased potential for off-target effects.

Metabolomic analysis of ALL cells harvested from the CNS and spleen shows significant changes in cell metabolism consistent with the low-nutrient microenvironment, and consistent with our RNAseq results. In particular the intracellular levels of mevalonate are higher in cells retrieved from the CNS compared to the spleen, which supports the hypothesis that cholesterol biosynthesis is increased in ALL cells in the CNS. There is a consistent pattern of increased saturated:unsaturated fatty acid ratios in the CSF of children and mice with leukaemia and ALL cells from the CNS vs spleens of mice. This fits with previously described changes in the transcription of genes involved in fatty acid metabolism(section 3.3.3). The saturation/desaturation ratio is has many important roles in cell biology, including maintaining membrane flexibility, cell signalling and even chemotherapy sensitivity(Zeng et al. 2008; Lee et al. 2018). We have shown in these experiments that CSF in non-leukaemic mice has a much higher palmitate (saturated):palmitoleate(unsaturated) ratio than plasma. It may be that upregulation of SCD and the partial reversal of this difference in saturated:unsaturated ratio in the CNS ALL is a way for the ALL cells to partially compensate for this change in environment.

Finally the data shown suggests that there may be increased glycolytic flux in the ALL cells in the CNS vs spleen, though the lack of clear data of CSF ^{13}C-glucose labelling makes drawing firm conclusions more challenging. In addition the intravenous introduction of glucose may change the CSF glucose concentration skewing these results – there is no evidence to confirm or refute this as yet. The increase in ^{13}C-glutamine in the CSF compared with plasma is very interesting – further studies are required to determine if this is the result of metabolism in ALL cells or in the meningeal stroma.

Chapter 6: Discussion and Future Directions

This project set out with 3 aims:

1) To identify metabolic adaptations of ALL cells to the CNS microenvironment that could be targets for CNS-directed therapy

2) To identify new metabolic markers of CNS involvement in the CSF

3) To investigate directly metabolic changes of ALL cells in the CNS identified in the course of this investigation

These aims have been addressed in the previous 3 chapters with findings summarised below. These findings have resulted in many more questions and some of the implications are explored as future directions.

6.1 Summary of findings

6.1.1 ALL cells undergo metabolic adaptations to the CNS niche

It has been shown with transcriptomic and metabolic analysis of cells from the CNS and spleen of mice that ALL cells undergo metabolic adaptations in the CNS niche. These most strikingly take the form of changes in lipid metabolism with evidence of cholesterol biosynthesis upregulation and changes in fatty acid metabolism.

This link between transcriptomic and metabolomic results is reassuring that these are true findings. Work is ongoing to more systematically integrate and analyse these transcriptomic and metabolomic datasets to provide more evidence for these findings and to look for novel changes in the CNS environment.

While there is now some evidence of transcriptional changes in CNS ALL from other studies(van der Velden et al. 2015; Münch et al. 2017), this study has some key strengths:

1) The use of a mouse model rather than primary cells leads to a trade off. We have excellent quality data from high numbers of viable CNS ALL cells, though at the cost of less direct clinical relevance. In addition to the difficulty in

obtaining cells from patient CSF for transcriptomic analysis, there is concern that the cells that are free in CSF may not be truly representative to the bulk CNS ALL population which is bound to the meninges – it may be that these cells have detached due to early apoptosis (certainly there is evidence of altered morphology or apoptosis on microscopy of many CSF leukaemia blasts). The key findings from the mouse model used in this project were validated in primary human cells, giving this project the best of both worlds in many respects.

2) The removal of RNA reads from the mouse genome allows us to be highly confident that the findings we have are a true reflection of CNS ALL and not contaminated with microenvironmental cells, though at the cost of excluding some highly-conserved genes from analysis.

3) The use of ALL cell lines allows for consistent high quality data to be obtained and also makes it possible to directly compare the metabolome and the transcriptome of CNS ALL to allow validation of results and potentially help pick out true signals from background noise.

The clinical implications of the importance of cholesterol biosynthesis in CNS ALL are discussed below (section 6.2.1).

6.1.2 There are novel candidate markers for CNS ALL

There is evidence for changes in the metabolome between CSF from children with ALL at diagnosis (most of whom we presume have CNS ALL) vs the same children later in treatment (presumably without CNS ALL), and between children with ALL at diagnosis and normal controls (i.e. children/adolescents who underwent lumbar puncture for investigation of CNS pathology, but were not found to have any). Of perhaps more relevance was the finding of creatine as a possible marker of CNS ALL that was separated in our diagnostic group from both controls.

One of the strengths of this project was the comparison of the CSF metabolome from children with ALL at diagnosis vs control CSF, a mouse model of CNS ALL, and children with CNS relapse of ALL. From these 3 analyses creatine and appears to be the most promising candidate biomarkers for further evaluation in different centres, with different techniques and potentially for prospective validation.

A key weakness in this approach however was the lack of true lipidomic analysis of CSF for lipid biomarkers. There was some evidence for alteration of the CSF saturated:monounsaturated fatty acid ratio in the presence of CNS ALL, but this was analysed via limited examination of CSF cholesterol esters and free fatty acids. Unfortunately experiments looking for the true abundance of saturated:monounsaturated fatty acids in diacyl and triacylgrycerols were not able to provide meaningful data, likely due to inadequate sample size.

Further analyses of these data for more integrated networks is discussed below (section 6.2.2).

6.1.3 Targeting critical metabolic weaknesses of CNS ALL is a viable strategy, but more work is needed to find successful therapeutics

As part of the detailed investigation into the importance of cholesterol biosynthesis *in vitro* and in CNS ALL *in vivo*, an attempt was made to target CNS ALL by disrupting cholesterol biosynthesis with HMGCR inhibitor simvastatin, and by directly knocking out or knocking down HMGCR. Genetic manipulation was not successful, presumably due to lethality of gene knockdown for ALL cells despite mevalonate supplementation – it was not possible to confirm this by assessing HMGCR transcription with PCR or western blot as there were not enough viable cells. Simvastatin treatment was successful *in vitro* – there was some evidence for reduced cell proliferation with lower doses (3μL), and clear cell toxicity with higher doses (10μL).

The *in vivo* findings were particularly interesting. Despite a dose-finding experiment allowing the use of high-dose oral simvastatin (100mg/kg), there was an increase in CNS ALL in male mice treated with simvastatin with no change in systemic disease burden. The specificity of this effect to male mice is particularly interesting. Two possibilities for this could be either: differences in the levels of steroid hormones (e.g. oestrogens, progestogens and testosterone) between male and female mice influence ALL cell biology *in vivo*; or that there are differences in immune function between male and female NSG mice –the IL2rg mutation, which results in removal of ILγ chains (and subsequently NK cells) in NSG mice, is X-linked. Regardless, the effect of simvastatin on CNS leukaemia in mice provides additional evidence for the importance of cholesterol biosynthesis for CNS ALL. A potential explanation for this finding is the upregulation of cholesterol biosynthesis in the systemic ALL cells in response to reduced availability of cholesterol in the plasma, leading to a cell phenotype better adapted for the CNS niche. In order to test this hypothesis, the plasma or CSF cholesterol abundance, and gene expression levels of cholesterol synthesis genes in CNS and spleen ALL cells, should be tested in ALL xenograft models with and without simvastatin treatment.

These data provide evidence that effective targeting of cholesterol synthesis in ALL cells in the CNS niche may be an attractive strategy to prevent or treat CNS relapse. This would be best shown in the first instance with a genetic knock-down/knock-out model of cholesterol biosynthesis to provide proof of concept, then strategies to target CNS ALL (e.g. more lipophilic drugs to optimise penetration of the CNS, or direct injection of therapeutics into the intrathecal space) could be explored.

6.2 Future directions

6.2.1 Cholesterol synthesis upregulation as a risk factor for CNS relapse

The most exciting finding from this project was the detection of a cholesterol biosynthesis signature at diagnosis that appears to carry a significantly increased risk of CNS relapse – a higher increase in risk than any currently used "high-risk" features. If validated, this could have implications for targeting CNS-directed therapy to children at particularly high risk to prevent CNS relapse, and eventually may allow for reduction in CNS-directed therapy for children at lower risk.

There are some key caveats to these data:

1) These are retrospective analyses of publicly available data.

2) Thiese are data from one study, using one treatment protocol

3) These are data from children with “high-risk” ALL, and may not be applicable to standard risk disease. In particular the use of a z-score analysis uses the mean level of expression of each gene to discriminate high-expressers from low-expressers, therefore changes in the distribution of gene expression (e.g. if high-risk children have a higher or lower overall level of cholesterol biosynthesis than standard risk children then the population of children identified by z-score analysis may be different).

Despite these caveats this date is extremely exciting. A power calculation based on a 3-fold increase in risk of CNS relapse, and a baseline CNS relapse rate of 2.5% of children suggests that transcriptomic data from around 1280 children would be required to validate these data. Several groups have been approached for supporting data and this analysis is ongoing.

6.2.2 Network analysis of metabolomic data

There are several powerful tools available for network or pathway analysis of transcriptomic data (GSEA and GeneMANIA are two used in this project). There are fewer tools available for metabolomic data, and very few for untargeted metabolomic data which is complicated by the difficulty in reliably identifying peaks. Work is currently ongoing with the Gottlieb group in Tel Aviv to analyse our data using machine-learning algorithms to find networks of metabolites that could indicate the presence of CNS leukaemia.

This is clearly a very challenging endeavour, particularly given the nature of primary CSF data which will have many confounding factors (e.g. the use of antibiotics and tumour lysis prophylaxis, high systemic disease burden, different levels of CNS disease between children) some of which (e.g. drugs) can be partially accounted for manually but may prove difficult to exclude algorithmically.

6.2.3 Metabolomic analysis of the CNS microenvironment

The metabolomic data presented in this thesis are extremely interesting and together with the transcriptional data provide a robust analysis of metabolic adaptations of CNS ALL cells. A wider question of metabolomic changes in the CNS microenvironmental niche has not been specifically addressed; cancer cell-stromal cell interactions are important for a wide variety of cancers

The data presented here have provided evidence of glutamine in the CSF derived (within 20 minutes) from glucose. Glutamate (a glutamine derivative) is a key neurotransmitter in the CNS, so it makes sense for the abundance of glutamine and glutamate in the CSF to be highly regulated. It is possible that this regulation is exploited by CNS ALL cells – the data above (section 5.5.2.2) shows evidence of labelled glutamine in ALL cells in the CNS which may derive from TCA metabolism within cells, but which may derive from the labelled glutamate in the CSF. Further investigation of this environment may provide additional insights into the metabolomic adaptations of ALL to the CNS niche.

6.2.4 Applicability of findings to CNS lymphoma

This project focusses on BCP-ALL, but the findings may have implications for similar diseases such as T-cell ALL, or T- or B-cell lymphoma. CNS relapse in B cell lymphoma is a significant clinical issues, and unlike with ALL, there is no clear consensus regarding the use of intrathecal and/or systemic chemotherapy for CNS prophylaxis. The BCSH guideline on the prevention of secondary CNS lymphoma pragmatically recommends intrathecal methotrexate while discussing the limitations of the evidence base for this, and cautiously endorses the use of systemic methotrexate with caveats(McMillan et al. 2013). There are significant overlaps in the biology of B-cell lymphoma and leukaemia including evidence for the role of the HIF pathway and VEGFα (Kim et al. 2011), osteopontin ((Tun et al. 2008; Yuan et al. 2013), and IRF4((Tun et al. 2008).

Lymphoma is a more common disease than childhood ALL (UK incidence of Non-Hodgkin lymphoma (NHL) was 23 per 100,000/year in 2015 for NHL, vs 0.168 per 100,000/year in 2013-2015 for ALL, figures from Cancer Research UK), and CNS relapse a more common issues for high-risk lymphoma than childhood ALL (CNS relapse rate ranges from <1%-25% depending on the presence of risk factors, and a typical overall rate of 2%-10% (Hollender et al. 2002; Boehme et al. 2009; Schmitz et al. 2012)). If the findings of this project (in particular the assessment of risk by cholesterol biosynthesis upregulation and the identification of better biomarkers of CNS disease) could be translated to CNS lymphoma there is potential for significant clinical benefit. From a different perspective, the larger clinical cohort of patients with lymphoma may allow the collection of a higher number of samples for analysis prospectively, and the faster validation of any biomarkers or therapeutic interventions than in childhood ALL.

6.3 Overall conclusions

This thesis presents data supporting the original hypothesis that ALL cells undergo metabolic adaptation to the CNS niche. It provides evidence for the targeting of CNS ALL cholesterol metabolism as a novel therapeutic strategy, and has identified a potential novel metabolic marker for CNS ALL. The most immediately exciting finding of this project is the identification of upregulation of cholesterol biosynthesis in the bone marrow at diagnosis as a highly significant risk factor for CNS ALL and work is in progress to validate this in independent cohorts.

Chapter 7: References

Akers, S.M. et al., 2010. VE-cadherin and PECAM-1 enhance ALL migration across brain microvascular endothelial cell monolayers. *Experimental hematology*, 38(9), pp.733–43.

Azzout-Marniche, D. et al., 2000. Insulin effects on sterol regulatory-element-binding protein-1c (SREBP-1c) transcriptional activity in rat hepatocytes. *The Biochemical journal*, 350 Pt 2(Pt 2), pp.389–93.

Bartram, J. et al., 2018. High throughput sequencing in acute lymphoblastic leukemia reveals clonal architecture of central nervous system and bone marrow compartments. *Haematologica*, 103(3), pp.e110–e114.

Bashford-Rogers, R.J.M. et al., 2016. Eye on the B-ALL: B-cell receptor repertoires reveal persistence of numerous B-lymphoblastic leukemia subclones from diagnosis to relapse. *Leukemia*, 30(12), pp.2312–2321.

Bleyer, W.A., 1989. Biology and pathogenesis of CNS leukemia. *The American journal of pediatric hematology/oncology*, 11(1), pp.57–63.

Bleyer, W.A. et al., 1991. Monthly pulses of vincristine and prednisone prevent bone marrow and testicular relapse in low-risk childhood acute lymphoblastic leukemia: a report of the CCG-161 study by the Childrens Cancer Study Group. *Journal of Clinical Oncology*, 9(6), pp.1012–1021.

Boehme, V. et al., 2009. CNS events in elderly patients with aggressive lymphoma treated with modern chemotherapy (CHOP-14) with or without rituximab: an analysis of patients treated in the RICOVER-60 trial of the German High-Grade Non-Hodgkin Lymphoma Study Group (DSHNHL). *Blood*, 113(17), pp.3896–3902.

Borowitz, M.J. et al., 2003. Minimal residual disease detection in childhood precursor–B-cell acute lymphoblastic leukemia: relation to other risk factors. A Children’s Oncology Group study. *Leukemia*, 17(8), pp.1566–1572.

Bowman, W.P. et al., 2011. Augmented therapy improves outcome for pediatric high risk acute lymphocytic leukemia: results of Children's Oncology Group trial P9906. *Pediatric blood & cancer*, 57(4), pp.569–77.

Bromberg, J.E.C. et al., 2007. CSF flow cytometry greatly improves diagnostic accuracy in CNS hematologic malignancies. *Neurology*, 68(20), pp.1674–9.

Buonamici, S. et al., 2009. CCR7 signalling as an essential regulator of CNS infiltration in T-cell leukaemia. *Nature*, 459(7249), pp.1000–4.

BURCHENAL, J.H. et al., 1953. Clinical evaluation of a new antimetabolite, 6-mercaptopurine, in the treatment of leukemia and allied diseases. *Blood*, 8(11), pp.965–99.

Burger, B., 2003. Diagnostic Cerebrospinal Fluid Examination in Children With Acute Lymphoblastic Leukemia: Significance of Low Leukocyte Counts With Blasts or Traumatic Lumbar Puncture. *Journal of Clinical Oncology*, 21(2), pp.184–188.

Cario, G. et al., 2007a. High interleukin-15 expression characterizes childhood acute lymphoblastic leukemia with involvement of the CNS. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 25(30), pp.4813–20.

Cario, G. et al., 2007b. High interleukin-15 expression characterizes childhood acute lymphoblastic leukemia with involvement of the CNS. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 25(30), pp.4813–20.

Chamberlain, M.C. et al., 2009. Diagnostic tools for neoplastic meningitis: detecting disease, identifying patient risk, and determining benefit of treatment. *Seminars in oncology*, 36(4 Suppl 2), pp.S35-45.

CLAUSEN, N. & IBSEN, K.K., 1984. Central Nervous System Relapse Surveillance by Serial Beta 2 -Microglobulin Measurements in Childhood Acute Lymphoblastic Leukemia. *Acta Paediatrica*, 73(6), pp.848–854.

Clayton, P.E. et al., 1988. Growth in children treated for acute lymphoblastic leukaemia. *Lancet (London, England)*, 1(8583), pp.460–2.

Consortium, T.G.O., 2015. Gene Ontology Consortium: going forward. *Nucleic Acids Research*, 43(D1), pp.D1049–D1056.

Cousens, P. et al., 1988. Cognitive effects of cranial irradiation in leukaemia: a survey and meta-analysis. *Journal of child psychology and psychiatry, and allied disciplines*, 29(6), pp.839–52.

DeBerardinis, R.J. & Chandel, N.S., 2016. Fundamentals of cancer metabolism. *Science Advances*, 2(5), pp.e1600200–e1600200.

Domaniewski, J. et al., 1987. Carcinoembryonic antigen (CEA), alphafetoprotein (AFP) alpha and beta subunits of human chorionic gonadotropin (hCG) in cerebrospinal fluid of children with acute lymphoblastic leukaemia. *Acta paediatrica Hungarica*, 28(2), pp.113-7.

Dutch Childhood Oncology Group, D.M.W.M. et al., 2006. Prognostic significance of blasts in the cerebrospinal fluid without pleiocytosis or a traumatic lumbar puncture in children with acute lymphoblastic leukemia: experience of the Dutch Childhood Oncology Group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 24(15), pp.2332–6.

Eiser, C., 1978. Intellectual abilities among survivors of childhood leukaemia as a function of CNS irradiation. *Archives of disease in childhood*, 53(5), pp.391–5.

Eiser, C. & Lansdown, R., 1977. Retrospective study of intellectual development in children treated for acute lymphoblastic leukaemia. *Archives of disease in childhood*, 52(7), pp.525–9.

Elder, A. et al., 2017. Abundant and equipotent founder cells establish and maintain acute lymphoblastic leukaemia. *Leukemia*, 31(12), pp.2577–2586.

Eskazan, A.E., Ar, M.C. & Baslar, Z., 2012. Intracranial extramedullary hematopoiesis in patients with thalassemia: a case report and review of the literature. *Transfusion*, 52(8), pp.1715–1720.

Evans, D.I., O'Rouke, C. & Jones, P.M., 1974. The cerebrospinal fluid in acute leukaemia of childhood: studies with the Cytocentrifuge. *Journal of clinical pathology*, 27(3), pp.226–30.

Farber, S. et al., 1948. Temporary Remissions in Acute Leukemia in Children Produced by Folic Acid Antagonist, 4-Aminopteroyl-Glutamic Acid (Aminopterin). *New England Journal of Medicine*, 238(23), pp.787–793.

Feng, S. et al., 2011. Matrix metalloproteinase-2 and -9 secreted by leukemic cells increase the permeability of blood-brain barrier by disrupting tight junction proteins. H. K. Avraham, ed. *PloS one*, 6(8), p.e20599.

Fleisherl, J.H. et al., 1979. CHOLESTEROL IN CEREBROSPINAL FLUID OF BRAIN TUMOR PATIENTS. *Life Sciences*, 19, pp.1517–1526.

Ganeshan, K. & Chawla, A., 2014. Metabolic Regulation of Immune Responses. *Annual Review of Immunology*, 32(1), pp.609–634.

Gaynes, J.S. et al., 2017. The central nervous system microenvironment influences the leukemia transcriptome and enhances leukemia chemo-resistance. *Haematologica*, 102(4), pp.e136–e139.

George, S.L. et al., 1973. Factors influencing survival in pediatric acute leukemia. The SWCCSG experience, 1958-1970. *Cancer*, 32(6), pp.1542–53.

Gilchrist, G.S. et al., 1994. Low numbers of CSF blasts at diagnosis do not predict for the development of CNS leukemia in children with intermediate-risk acute lymphoblastic leukemia: a Childrens Cancer Group report. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 12(12), pp.2594–600.

Glass, J.P. et al., 1979. Malignant cells in cerebrospinal fluid (CSF): The meaning of a positive CSF cytology. *Neurology*, 29(10), pp.1369–1375.

Goldstein, J.L. & Brown, M.S., 1990. Regulation of the mevalonate pathway. *Nature*, 343(6257), pp.425–30.

Hagedorn, N. et al., 2007. Submicroscopic bone marrow involvement in isolated extramedullary relapses in childhood acute lymphoblastic leukemia: a more precise definition of “isolated” and its possible clinical implications, a collaborative study of the Resistant Disease Committee . *Blood*, 110(12), pp.4022–9.

Halsey, C. et al., 2011. The impact of therapy for childhood acute lymphoblastic leukaemia on intelligence quotients; results of the risk-stratified randomized central nervous system treatment trial MRC UKALL XI. *Journal of hematology & oncology*, 4, p.42.

Hansen, P.B. et al., 1991. Cerebrospinal fluid beta-2-microglobulin in adult patients with acute leukemia or lymphoma: a useful marker in early diagnosis and monitoring of CNS-involvement. *Acta Neurologica Scandinavica*, 85(3), pp.224–227.

Hardisty, R.M. & Norman, P.M., 1967. Meningeal leukaemia. *Archives of disease in childhood*, 42(224), pp.441–7.

HERSCHKOWITZ, N. & CUMINGS, J.N., 1964. CREATINE KINASE IN CEREBROSPINAL FLUID. *Journal of neurology, neurosurgery, and psychiatry*, 27(3), pp.247–50.

Hill, F.G.H. et al., 2004. Successful treatment without cranial radiotherapy of children receiving intensified chemotherapy for acute lymphoblastic leukaemia: results of the risk-stratified randomized central nervous system treatment trial MRC UKALL XI (ISRC TN 16757172). *British journal of haematology*, 124(1), pp.33–46.

Holland, M. et al., 2011. RAC2, AEP, and ICAM1 expression are associated with CNS disease in a mouse model of pre-B childhood acute lymphoblastic leukemia. *Blood*, 118(3), pp.638–49.

Hollender, A. et al., 2002. Central nervous system involvement following diagnosis of non-Hodgkin's lymphoma: a risk model. *Annals of oncology : official journal of the European Society for Medical Oncology*, 13(7), pp.1099–107.

Horton, J.D., Goldstein, J.L. & Brown, M.S., 2002. SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. *The Journal of clinical investigation*, 109(9), pp.1125–31.

Hua, X. et al., 1993. SREBP-2, a second basic-helix-loop-helix-leucine zipper protein that stimulates transcription by binding to a sterol regulatory element (cDNA cloning/cholesterol/low density lipoprotein receptor/3-hydroxy-3-methylglutaryl-coenzyme A synthase). *Biochemistry*, 90, pp.11603–11607.

Hua, X. et al., 1995. Structure of the Human Gene Encoding Sterol Binding Protein-l (SREBFI) and Localization of Regulatory Element SREBFI and SREBFZ to Chromosomes 17~11.2 and 22q13. *GENOMICS*, 25, pp.667–673.

Huei-Mei Kuo, A. et al., 1975. Proliferative kinetics of central nervous system (CNS) leukemia. *Cancer*, 36(1), pp.232–239.

Ikonen, E., 2008. Cellular cholesterol trafficking and compartmentalization. *Nature Reviews Molecular Cell Biology*, 9(2), pp.125–138.

Illingworth, D.R. & Glover, J., 1971. THE COMPOSITION OF LIPIDS IN CEREBROSPINAL FLUID OF CHILDREN AND ADULTS. *Journal of Neurochemistry*, 18(5), pp.769–776.

Incesoy-Özdemir, S. et al., The relationship between cerebrospinal fluid osteopontin level and central nervous system involvement in childhood acute leukemia. *The Turkish journal of pediatrics*, 55(1), pp.42–9.

Irving, J. et al., 2014. Ras pathway mutations are prevalent in relapsed childhood acute lymphoblastic leukemia and confer sensitivity to MEK inhibition. *Blood*, 124(23), pp.3420–30.

Isoda, K. et al., 2003. Osteopontin transgenic mice fed a high-cholesterol diet develop early fatty-streak lesions. *Circulation*, 107(5), pp.679–81.

Jankovic, M. et al., 1994. Association of 1800 cGy cranial irradiation with intellectual function in children with acute lymphoblastic leukaemia. *The Lancet*, 344(8917), pp.224–227.

Janssen, S.F. et al., 2013. Gene Expression and Functional Annotation of the Human and Mouse Choroid Plexus Epithelium S. D. Ginsberg, ed. *PLoS ONE*, 8(12), p.e83345.

Jeha, S. et al., 2009. Increased risk for CNS relapse in pre-B cell leukemia with the t(1;19)/TCF3-PBX1. *Leukemia*, 23(8), pp.1406–9.

Kanehisa, M. et al., 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, 45(D1), pp.D353–D361.

Kanehisa, M. et al., 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 44(D1), pp.D457–D462.

Kanehisa, M. & Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1), pp.27–30.

Kato, I. et al., 2017. Hypoxic adaptation of leukemic cells infiltrating the CNS affords a therapeutic strategy targeting VEGF. *Blood*.

Kim, J.A. et al., 2011. Hypoxia-associated protein expression in primary central nervous system diffuse large B-cell lymphoma: does it predict prognosis? *Leukemia & lymphoma*, 52(2), pp.205–13.

Kivisäkk, P. et al., 2003. Human cerebrospinal fluid central memory CD4+ T cells: evidence for trafficking through choroid plexus and meninges via P-selectin. *Proceedings of the National Academy of Sciences of the United*

States of America, 100(14), pp.8389–94.

Koch, S. et al., 2001. Characterization of four lipoprotein classes in human cerebrospinal fluid. *J. Lipid Res.*, 42(7), pp.1143–1151.

Kondoh, T. et al., 2014. CD7 promotes extramedullary involvement of the B-cell acute lymphoblastic leukemia line Tanoue by enhancing integrin β2-dependent cell adhesiveness. *International Journal of Oncology*, 45(3), pp.1073–1081.

Kovesi, T.A. & Hsu, E., Changes in lactate dehydrogenase isoenzymes associated with relapse of childhood acute lymphocytic leukemia. *Pediatric hematology and oncology*, 11(5), pp.527–33.

Krause, S. et al., 2015. Mer tyrosine kinase promotes the survival of t(1;19)-positive acute lymphoblastic leukemia (ALL) in the central nervous system (CNS). *Blood*, 125(5), pp.820–830.

Krishnan, S. et al., 2010. Temporal changes in the incidence and pattern of central nervous system relapses in children with acute lymphoblastic leukaemia treated on four consecutive Medical Research Council trials, 1985-2001. *Leukemia*, 24(2), pp.450–9.

Lee, J.Y. et al., 2001. Saturated fatty acids, but not unsaturated fatty acids, induce the expression of cyclooxygenase-2 mediated through Toll-like receptor 4. *J Biol Chem*. 276(20), pp.16683-9

Levine, G.A., Winkelstein, A. & Shadduck, R.K., 1973. CNS involvement as the initial manifestation of acute leukemia. *Cancer*, 31(4), pp.959–962.

Levinsen, M. et al., 2014. Flow Cytometric Leukemic Blasts Detection in Cerebrospinal Fluid of Children with Acute Lymphoblastic Leukemia. *Blood*, 124(21).

Liberzon, A. et al., 2015. The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Systems*, 1(6), pp.417–425.

Louveau, A. et al., 2015. Structural and functional features of central nervous system lymphatic vessels. *Nature*, 523(7560), pp.337–341.

Love, M.I., Huber, W. & Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), p.550.

Luomala, M. et al., 2007. Osteopontin levels are associated with cholesterol synthesis markers in mildly hypercholesterolaemic patients. *Acta cardiologica*, 62(2), pp.177–81.

M?nch, V. et al., 2017. Central Nervous System Involvement in Acute Lymphoblastic Leukemia Is Mediated by Vascular Endothelial Growth Factor. *Blood*, p.blood-2017-03-769315.

Mackay, G.M. et al., 2015. Analysis of Cell Metabolism Using LC-MS and Isotope Tracers. In *Methods in enzymology*. pp. 171–196.

Mahmoud, H.H. et al., 1993. Low leukocyte counts with blast cells in cerebrospinal fluid of children with newly diagnosed acute lymphoblastic leukemia. *The New England journal of medicine*, 329(5), pp.314–9.

McMillan, A. et al., 2013. Guideline on the prevention of secondary central nervous system lymphoma: British Committee for Standards in Haematology. *British Journal of Haematology*, 163(2).

MIELCAREK, M. et al., 1997. Expression of intercellular adhesion molecule 1 (ICAM-1) in childhood acute lymphoblastic leukaemia: correlation with clinical features and outcome. *British Journal of Haematology*, 96(2), pp.301–307.

Mitchell, C.D. et al., 2005. Benefit of dexamethasone compared with prednisolone for childhood acute lymphoblastic leukaemia: results of the UK Medical Research Council ALL97 randomized trial. *British Journal of Haematology*, 129(6), pp.734–745.

Moleski, M., 2000. Neuropsychological, neuroanatomical, and neurophysiological consequences of CNS chemotherapy for acute lymphoblastic leukemia. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists*, 15(7), pp.603–30.

MOORE, E.W. et al., 1960. The central nervous system in acute leukemia: a postmortem study of 117 consecutive cases, with particular reference to hemorrhages, leukemic infiltrations, and the syndrome of meningeal leukemia. *Archives of internal medicine*, 105, pp.451–68.

Moschovi, M. et al., 2004. Serum Lipid Alterations in Acute Lymphoblastic Leukemia of Childhood. *Journal of Pediatric Hematology Oncology*, 26(5), pp.289-93

Mostafavi, S. et al., 2008. GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome biology*, 9 Suppl 1(Suppl 1), p.S4.

Neglia, J.P. et al., 1991. Second Neoplasms after Acute Lymphoblastic Leukemia in Childhood. *New England Journal of Medicine*, 325(19), pp.1330–1336.

Nesbit, M.E. et al., 1982. Sanctuary therapy: a randomized trial of 724 children with previously untreated acute lymphoblastic leukemia: A Report from Children’s Cancer Study Group. *Cancer research*, 42(2), pp.674–80.

Nygaard, R. et al., 1991. Second malignant neoplasms in patients treated for childhood leukemia. A population-based cohort study from the Nordic countries. The Nordic Society of Pediatric Oncology and Hematology (NOPHO). *Acta paediatrica Scandinavica*, 80(12), pp.1220–8.

Oliff, A. et al., 1979. HypothalamiePituitary Dysfunction Following CNS Prophylaxis in Acute Lymphocytic Leukemia: Correlation With CT Scan Abnormalities. *Medical and Pediatric Oncology*, 7, pp.141–151.

Paoletti, P. et al., 1969. The sterol test for the diagnosis of human brain tumors. *Neurology*, 19(2), pp.190–7.

Parker, T.S. et al., 1984. Plasma mevalonate as a measure of cholesterol synthesis in man. *The Journal of clinical investigation*, 74(3), pp.795–804.

Pine, S.R. et al., 2005. Detection of central nervous system leukemia in children with acute lymphoblastic leukemia by real-time polymerase chain reaction. *The Journal of molecular diagnostics : JMD*, 7(1), pp.127–32.

Piovezani Ramos, G. et al., 2016. Initial presentation of CNS-restricted acute lymphoblastic B cell leukaemia as peripheral polyneuropathy. *BMJ Case Reports*, 2016, p.bcr2016214645.

Pitas, R.E. et al., 1987. Lipoproteins and their receptors in the central nervous system. Characterization of the lipoproteins in cerebrospinal fluid and identification of apolipoprotein B,E(LDL) receptors in the brain. *The Journal of biological chemistry*, 262(29), pp.14352–60.

Price, R.A. & Johnson, W.W., 1973. The central nervous system in childhood leukemia: I. The arachnoid. *Cancer*, 31(3), pp.520–533.

Pudek, M.R. et al., 1985. Beta2-microglobulin levels in cerebrospinal fluid of children with leukemia and lymphoma. *Clinical Biochemistry*, 18(3), pp.180–183.

Pui, C.-H., 2006. Central nervous system disease in acute lymphoblastic leukemia: prophylaxis and treatment. *Hematology. American Society of Hematology. Education Program*, 2006(1), pp.142–6.

Pui, C.-H. et al., 2003. Extended Follow-up of Long-Term Survivors of Childhood Acute Lymphoblastic Leukemia. *New England Journal of Medicine*, 349(7), pp.640–649.

Ransohoff, R.M., Kivisäkk, P. & Kidd, G., 2003. Three or more routes for leukocyte migration into the central nervous system. *Nature reviews. Immunology*, 3(7), pp.569–81.

Rao, D. et al., 2012. Case series: CSF LDH, proteins and electrolyte levels in patients of acute lymphocytic leukemia. *Clinica chimica acta; international journal of clinical chemistry*, 413(13–14), pp.1045–8.

Repa, J.J. et al., 2000. Regulation of mouse sterol regulatory element-binding protein-1c gene (SREBP-1c) by oxysterol receptors, LXRalpha and LXRbeta. *Genes & development*, 14(22), pp.2819–30.

Rimm, I.J. et al., 1987. Brain tumors after cranial irradiation for childhood acute lymphoblastic leukemia. A 13-year experience from the Dana-Farber Cancer Institute and the Children's Hospital. *Cancer*, 59(8), pp.1506–8.

Roy, A. et al., 2005. Outcome after first relapse in childhood acute lymphoblastic leukaemia - lessons from the United Kingdom R2 trial. *British Journal of Haematology*, 130(1), pp.67–75.

Sakka, L., Coll, G. & Chazal, J., 2011. Anatomy and physiology of cerebrospinal fluid. *European Annals of Otorhinolaryngology, Head and Neck Diseases*, 128(6), pp.309–316.

Samuels, A.L. et al., 2014. A pre-clinical model of resistance to induction therapy in pediatric acute lymphoblastic leukemia. *Blood cancer journal*, 4, p.e232.

Samuels, A.L. et al., 2010. Validation of a mouse xenograft model system for gene expression analysis of human acute lymphoblastic leukaemia. *BMC Genomics*, 11(1), p.256.

Schmitz, N. et al., 2012. CNS disease in younger patients with aggressive B-cell lymphoma: an analysis of patients treated on the Mabthera International Trial and trials of the German High-Grade Non-Hodgkin Lymphoma Study Group. *Annals of Oncology*, 23(5), pp.1267–1273.

Schug, Z. et al., 2015. Acetyl-CoA Synthetase 2 Promotes Acetate Utilization and Maintains Cancer Cell Growth under Metabolic Stress. *Cancer Cell*, 27(1), pp.57–71.

Seidenfeld, J. & Marton, L.J., 1979. Biochemical Markers of Central Nervous System Tumors Measured in Cerebrospinal Fluid and Their Potential Use in Diagnosis and Patient Management: A Review. *J Natl Cancer Inst*, 63(4), pp.919–931.

Sheen, C. et al., 2011. Statins are active in acute lymphoblastic leukaemia (ALL): a therapy that may treat ALL and prevent avascular necrosis. *British journal of haematology*, 155(3), pp.403–7.

Simone, J. et al., 1972. "Total therapy" studies of acute lymphocytic leukemia in children. Current results and prospects for cure. *Cancer*, 30(6), pp.1488–94.

Smith, C.A. et al., 2005. METLIN: a metabolite mass spectral database. *Therapeutic drug monitoring*, 27(6), pp.747–51.

Stone, J.B. & DeAngelis, L.M., 2016. Cancer-treatment-induced neurotoxicity—focus on newer treatments. *Nature Reviews Clinical Oncology*, 13(2), pp.92–105.

Subramanian, A. et al., 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, 102(43), pp.15545–50.

SULLIVAN, M.P., 1957. Intracranial complications of leukemia in children. *Pediatrics*, 20(5, Part 1), pp.757–81.

Sutow, W.W. et al., 1971. L-asparaginase therapy in children with advanced leukemiaThe Southwest cancer chemotherapy study group. *Cancer*, 28(4), pp.819–824.

Tabesh, H. et al., 2011. An intracranial extramedullary hematopoiesis in a 34-year-old man with beta thalassemia: a case report. *Journal of Medical Case Reports*, 5(1), p.580.

Di Terlizzi, R. & Platt, S., 2006. The function, composition and analysis of cerebrospinal fluid in companion animals: Part I – Function and composition. *The Veterinary Journal*, 172(3), pp.422–431.

THOMAS, L.B. et al., 1964. DEVELOPMENT OF MENINGEAL LEUKEMIA (L1210) DURING TREATMENT OF SUBCUTANEOUSLY INOCULATED MICE WITH METHOTREXATE. *Cancer*, 17, pp.352–60.

Tsuchiya, J. et al., 1978. Proliferative kinetics of the leukemic cells in meningeal leukemia. *Cancer*, 42(3), pp.1255–62.

Tun, H.W. et al., 2008. Pathway analysis of primary central nervous system lymphoma. *Blood*, 111(6), pp.3200–10.

Uruena, M. et al., 1991. Impaired pubertal growth in acute lymphoblastic leukaemia. *Archives of disease in childhood*, 66(12), pp.1403–7.

Vance, J.E., Hayashi, H. & Karten, B., 2005. Cholesterol homeostasis in neurons and glial cells. *Seminars in cell & developmental biology*, 16(2), pp.193–212.

Vazquez, A. et al., 2016. Cancer metabolism at a glance. *Journal of cell science*, 129(18), pp.3367–73.

van der Velden, V.H.J. et al., 2015. New cellular markers at diagnosis are associated with isolated central nervous system relapse in paediatric B-cell precursor acute lymphoblastic leukaemia. *British Journal of Haematology*

Venkatesh, B. et al., 1999. The continuous measurement of cerebrospinal fluid gas tensions in critically ill neurosurgical patients: a prospective observational study. *Intensive care medicine*, 25(6), pp.599–605.

Vicente, V., González, M. & Borrasca, A.L., 1982. CEREBROSPINAL FLUID LEVELS BETA 2 MICROGLOBULIN AND FERRITIN IN LYMPHOPROLIVERATIVE DISORDERS. *Acta Paediatrica*, 71(2), pp.325–326.

Vora, A. et al., 2013. Treatment reduction for children and young adults with low-risk acute lymphoblastic leukaemia defined by minimal residual disease (UKALL 2003): a randomised controlled trial. *The Lancet. Oncology*, 14(3), pp.199–209.

Wang, Z., Gerstein, M. & Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*, 10(1), pp.57–63.

Warburg, O., 1956. On the Origin of Cancer Cells. *Science*, 123(3191), pp.309–314.

Warde-Farley, D. et al., 2010. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic acids research*, 38(Web Server issue), pp.W214-20.

Warrier, I.A. et al., 1981. 818 SIGNIFICANCE OF ELEVATED B2 MICROGLOBULIN (B2m) IN THE CEREBROSPINAL FLUID (CSF) OF CHILDREN WITH ACUTE LEUKEMIA (AL). *Pediatric Research*, 15(S4), pp.579–579.

Wasserstrom, W.R., Glass, J.P. & Posner, J.B., 1982. Diagnosis and treatment of leptomeningeal metastases from solid tumors: Experience with 90 patients. *Cancer*, 49(4), pp.759–772.

Williams, M.T.S. et al., 2014. Interleukin-15 enhances cellular proliferation and upregulates CNS homing molecules in pre-B acute lymphoblastic leukemia. *Blood*, 123(20), pp.3116–27.

Williams, M.T.S. et al., 2016. The ability to cross the blood-cerebrospinal fluid barrier is a generic property of acute lymphoblastic leukaemia blasts. *Blood*.

Wishart, D.S. et al., 2007. HMDB: the Human Metabolome Database. *Nucleic Acids Research*, 35(Database), pp.D521–D526.

Xia, J. & Wishart, D.S., 2011. Metabolomic Data Processing, Analysis, and Interpretation Using MetaboAnalyst. *Current Protocols in Bioinformatics*, 34(1), p.14.10.1-14.10.48.

Yang, L. et al., 2012. Osteopontin and integrin are involved in cholesterol gallstone formation. *Medical science monitor : international medical journal of experimental and clinical research*, 18(1), pp.BR16-23.

Yousafzai, Y.M., 2015. Mechanisms of central nervous system disease in childhood acute lymphoblastic leukaemia.

Yuan, J. et al., 2013. Preferential up-regulation of osteopontin in primary central nervous system lymphoma does not correlate with putative receptor CD44v6 or CD44H expression. *Human pathology*, 44(4), pp.606–11.

Zaharchuk, G. et al., 2005. Measurement of cerebrospinal fluid oxygen partial pressure in humans using MRI. *Magnetic Resonance in Medicine*, 54(1), pp.113–121.

Van Zanten, A.P. et al., 1986. Cerebrospinal fluid lactate dehydrogenase activities in patients with central nervous system metastases. *Clinica Chimica Acta*, 161(3), pp.259–268.

Zeng, L. et al., 2008. Saturated Fatty Acids Modulate Cell Response to DNA Damage: Implication for Their Role in Tumorigenesis D.-Y. Jin, ed. *PLoS ONE*, 3(6), p.e2329.

Appendix

Visual Basic Scripts:

```
Sub Copy+paste_for_early_vs_late_pools_May15()

Dim metabolite(1 To 33) As String

metabolite(1) = "Adenine"
metabolite(2) = "aKG"
metabolite(3) = "Alanine"
metabolite(4) = "Arginine"
metabolite(5) = "Asparagine"
metabolite(6) = "Citrate"
metabolite(7) = "Cytidine"
metabolite(8) = "GLN"
metabolite(9) = "Glucose"
metabolite(10) = "Glycine"
metabolite(11) = "Lactate"
metabolite(12) = "Leucine"
metabolite(13) = "Lysine"
metabolite(14) = "methionine"
metabolite(15) = "Phenylalanine"
metabolite(16) = "proline"
metabolite(17) = "Pyruvate"
metabolite(18) = "Serine"
metabolite(19) = "threonine"
metabolite(20) = "tryptophan"
metabolite(21) = "Tyrosine"
metabolite(22) = "Ornithine"
metabolite(23) = "GLU"
metabolite(24) = "Succinic acid"
metabolite(25) = "Fumarate"
metabolite(27) = "homocysteine"
metabolite(26) = "IsoLeucine"
metabolite(28) = "urate"
metabolite(29) = "Pyruvate +3"
metabolite(30) = "Lactate +3"
metabolite(31) = "Glucose +6"
metabolite(32) = "Alanine +1"
metabolite(33) = "Arginine +6"


Dim a As Integer
Dim b As Integer
Dim mycell As Range

Worksheets(metabolite(1)).Select
Range("A39", "C45").Copy
```

```
For a = 2 To 25

Worksheets(metabolite(a)).Select
Range("A39").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False

    Next a

Worksheets(metabolite(1)).Select
Range("E39", "G45").Copy

For a = 26 To 28

Worksheets(metabolite(a)).Select
Range("A39").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False

 Next a

End Sub
*************************************************************************************
```

```
Sub T_Test_early_vs_late_May_15()

Dim metabolite(1 To 33) As String

metabolite(1) = "Adenine"
metabolite(2) = "aKG"
metabolite(3) = "Alanine"
metabolite(4) = "Arginine"
metabolite(5) = "Asparagine"
metabolite(6) = "Citrate"
metabolite(7) = "Cytidine"
metabolite(8) = "GLN"
metabolite(9) = "Glucose"
metabolite(10) = "Glycine"
metabolite(11) = "Lactate"
metabolite(12) = "Leucine"
metabolite(13) = "Lysine"
metabolite(14) = "methionine"
metabolite(15) = "Phenylalanine"
metabolite(16) = "proline"
metabolite(17) = "Pyruvate"
metabolite(18) = "Serine"
metabolite(19) = "threonine"
metabolite(20) = "tryptophan"
metabolite(21) = "Tyrosine"
metabolite(22) = "IsoLeucine"
metabolite(23) = "GLU"
metabolite(24) = "homocysteine"
metabolite(25) = "Fumarate"
metabolite(26) = "Ornithine"
metabolite(27) = "Succinic acid"
metabolite(28) = "urate"
metabolite(29) = "Pyruvate +3"
metabolite(30) = "Lactate +3"
metabolite(31) = "Glucose +6"
metabolite(32) = "Alanine +1"
metabolite(33) = "Arginine +6"

Dim a As Integer
Dim b As Integer

Worksheets("Stats").Select
Range("h1").Select
Selection = "T-test p-value for combined early vs. late"

For b = 2 To 29
  ActiveSheet.Cells(b, 8).Select
```

```
    Selection = (metabolite(b - 1))
Next b



For a = 1 To 28

    Worksheets(metabolite(a)).Select
        Range("B46").Select
        Selection = "T-Test for difference between combined early and late values"
        Range("B47").Select
        Selection = "=T.TEST(B40:B44,C40:C44,2,1)"
        Range("B47").Copy

    Worksheets("stats").Activate
    ActiveSheet.Cells(a + 1, 9).Select
    Selection.PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks:=False, Transpose:=False

    Next a

Dim mycell As Range
Worksheets("stats").Select
    For Each mycell In Range("I2", "I29").Cells
        If mycell.Value <= 0.05 Then
        mycell.Interior.Color = vbGreen
        End If
        If mycell.Value <= 0.001 Then
        mycell.Interior.Color = RGB(0, 208, 0)
        End If
    Next mycell
End Sub
**********************************************************************************************
```

```
Sub collate_early_only_values_May15()

Dim metabolite(1 To 33) As String

metabolite(1) = "Adenine"
metabolite(2) = "aKG"
metabolite(3) = "Alanine"
metabolite(4) = "Arginine"
metabolite(5) = "Asparagine"
metabolite(6) = "Citrate"
metabolite(7) = "Cytidine"
metabolite(8) = "GLN"
metabolite(9) = "Glucose"
metabolite(10) = "Glycine"
metabolite(11) = "Lactate"
metabolite(12) = "Leucine"
metabolite(13) = "Lysine"
metabolite(14) = "methionine"
metabolite(15) = "Phenylalanine"
metabolite(16) = "proline"
metabolite(17) = "Pyruvate"
metabolite(18) = "Serine"
metabolite(19) = "threonine"
metabolite(20) = "tryptophan"
metabolite(21) = "Tyrosine"
metabolite(22) = "Ornithine"
metabolite(23) = "GLU"
metabolite(24) = "Succinic acid"
metabolite(25) = "Fumarate"
metabolite(26) = "IsoLeucine"
metabolite(27) = "homocysteine"
metabolite(28) = "urate"
metabolite(29) = "Pyruvate +3"
metabolite(30) = "Lactate +3"
metabolite(31) = "Glucose +6"
metabolite(32) = "Alanine +1"
metabolite(33) = "Arginine +6"

Dim a As Integer
Dim b As Integer

For a = 1 To 25

Worksheets(metabolite(a)).Select

Range("B49").Select
Selection = "1"
```

```
Range("C49").Select
Selection = "2"
Range("D49").Select
Selection = "3"
'Range("E1").Select
'Selection = "4"
Range("e49").Select
Selection = "5"
Range("f49").Select
Selection = "6"

Range("B20", "B22").Copy
Range("B50").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("D20", "D22").Copy
Range("C50").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("F20", "F22").Copy
Range("D50").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("H20", "H22").Copy
Range("E50").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("J20", "J22").Copy
Range("F50").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False

    Next a

For a = 26 To 28

Worksheets(metabolite(a)).Select

Range("B49").Select
Selection = "1"
Range("C49").Select
Selection = "2"
Range("D49").Select
Selection = "3"
'Range("E1").Select
'Selection = "4"
Range("e49").Select
Selection = "5"
Range("f49").Select
Selection = "6"

Range("B8", "B10").Copy
```

```
Range("B50").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("D8", "D10").Copy
Range("C50").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("F8", "F10").Copy
Range("D50").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("H8", "H10").Copy
Range("E50").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("J8", "J10").Copy
Range("F50").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False

Next a

End Sub
******************************************************************************************
```

```
Sub delete_irrelevent_rows_from_RNA_data()

Dim a As Long
Application.Calculation = xlCalculationManual
Application.ScreenUpdating = False
    For a = 66000 To 2 Step -1

        With Worksheets("Data1").Cells(a, 21)
            If Not .Value = "protein_coding" Then Worksheets("Data1").Rows(a).EntireRow.delete
        End With
    Next a

    For a = 66000 To 2 Step -1

        With Worksheets("Data1").Cells(a, 22)
          If Not .Value > 100 Then Worksheets("Data1").Rows(a).EntireRow.delete
       End With

    Next a
Application.Calculation = xlCalculationAutomatic
Application.ScreenUpdating = True

End Sub
************************************************************************************
```

```
Sub Met_Arrange_September15_fatty_acids()
'declare array of variable 'metabolite' as string
Dim metabolite(1 To 6) As String
metabolite(1) = "Linoleic acid"
metabolite(2) = "Linolenic acid"
metabolite(3) = "Oleic acid"
metabolite(4) = "Palmitic acid"
metabolite(5) = "Palmitoleic acid"
metabolite(6) = "Stearic acid"

' set sheet1 as "All samples"
If ActiveSheet.Name <> "All Samples" Then ActiveSheet.Name = "All Samples"

Dim i As Integer                    ' Integers used in 'For' loops
Dim j As Integer
Dim k As Integer
Dim m As Integer
Dim LastCol As Integer

Sheets.Add After:=Sheets("All Samples")
ActiveSheet.Name = metabolite(1) 'creates first worksheet

  For k = 2 To 6                 'Creates subsequent worksheets for each metabolite
   Sheets.Add After:=Sheets(metabolite(k - 1))
   'Sheets.Add
   ActiveSheet.Name = metabolite(k)
   Next k

                                 'Back to All Samples worksheet
Worksheets("All Samples").Select

   ' Loop through cells A1-A1200, metabolites 1-30 until 'Metabolite' is found

   For i = 1 To 1500
      For j = 1 To 6
         If Cells(i, 2).Value = metabolite(j) Then
            Rows(i).Copy                ' Copy entire row

           Worksheets(metabolite(j)).Select 'select appropriate workbook
                             'paste transpose entire row to appropriate workbook

              With ActiveSheet
              LastCol = .Cells(1, .Columns.Count).End(xlToLeft).Column
              Sheets(metabolite(j)).Cells(1, LastCol + 1).Select
              Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=True
              Worksheets("All Samples").Select
              End With
```

```
        Application.CutCopyMode = False 'clear clipboard
       End If
    Next j

  Next i

'Paste titles onto each worksheet

Range("A1:T1").Select
Selection.Copy

  For m = 1 To 6
  Sheets(metabolite(m)).Select
  Columns(1).Select
  Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=True
    Next m

 Application.CutCopyMode = False         'clear clipboard

End Sub
************************************************************************************
```

```
Sub Copy_Into_Table_Sept2015_Fatty_acids()

Dim metabolite(1 To 6) As String
metabolite(1) = "Linoleic acid"
metabolite(2) = "Linolenic acid"
metabolite(3) = "Oleic acid"
metabolite(4) = "Palmitic acid"
metabolite(5) = "Palmitoleic acid"
metabolite(6) = "Stearic acid"

Dim a As Integer

For a = 1 To 6

    Worksheets(metabolite(a)).Select
    'write out table backbone

    Range("B7").Select
    Selection = "1-Early"
    Range("C7").Select
    Selection = "1-Late"
    Range("D7").Select
    Selection = "2-Early"
    Range("E7").Select
    Selection = "2-Late"

    Range("F7").Select
    Selection = "3-Early"
    Range("G7").Select
    Selection = "3-Late"

    Range("H7").Select
    Selection = "4-Early"
    Range("I7").Select
    Selection = "4-Late"

    Range("J7").Select
    Selection = "5-Early"
    Range("K7").Select
    Selection = "5-Late"

    Range("A8").Select
    Selection = "Peak Area"


    Range("B7", "M7").Copy
    Range("B19").Select
```

```
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False

' copy peak area

Range("B3").Copy
Range("B8").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("C3").Copy
Range("B9").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("D3").Copy
Range("B10").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("E3").Copy
Range("C8").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("F3").Copy
Range("C9").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("G3").Copy
Range("C10").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("H3").Copy
Range("D8").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("I3").Copy
Range("D9").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("J3").Copy
Range("D10").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("K3").Copy
Range("E8").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("L3").Copy
Range("E9").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("M3").Copy
Range("E10").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("N3").Copy
Range("F8").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("O3").Copy
Range("F9").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("P3").Copy
```

```
Range("F10").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("Q3").Copy
Range("G8").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("R3").Copy
Range("G9").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("S3").Copy
Range("G10").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("T3").Copy
Range("H8").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("U3").Copy
Range("H9").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("V3").Copy
Range("H10").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("W3").Copy
Range("I8").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("X3").Copy
Range("I9").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("Y3").Copy
Range("I10").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("Z3").Copy
Range("J8").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("AA3").Copy
Range("J9").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("AB3").Copy
Range("J10").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("AC3").Copy
Range("k8").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("AD3").Copy
Range("k9").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
Range("AE3").Copy
Range("k10").Select
Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
```

```
    Next a

End Sub
```

****

```
Sub Select_cholesterol_synthesis_genes_from_array_updatedMay16()

Dim mycell As Range
Dim a As Long
Dim lastrow As Integer

Sheets.Add After:=Sheets("GSE11392_series_matrix")
ActiveSheet.Name = "Cholesterol pathway analysis" 'creates first worksheet

For a = 1 To 54720
Worksheets("GSE11392_series_matrix").Select

    If ActiveSheet.Cells(a, 3).Value2 = "CYP51A1" Or ActiveSheet.Cells(a, 3).Value2 = "DHCR7" Or ActiveSheet.Cells(a, 3).Value2 = "DHCR24" Or ActiveSheet.Cells(a, 3).Value2 = "FDFT1" Or ActiveSheet.Cells(a, 3).Value2 = "HMGCR" Or ActiveSheet.Cells(a, 3).Value2 = "HMGCS1" Or ActiveSheet.Cells(a, 3).Value2 = "LSS" Or ActiveSheet.Cells(a, 3).Value2 = "MSMO1" Or ActiveSheet.Cells(a, 3).Value2 = "MVK" Or ActiveSheet.Cells(a, 3).Value2 = "LSS" Or ActiveSheet.Cells(a, 3).Value2 = "SQLE" Or ActiveSheet.Cells(a, 3).Value2 = "TM7SF2" Or ActiveSheet.Cells(a, 3).Value2 = "PMVK" Or ActiveSheet.Cells(a, 3).Value2 = "MVD" Or ActiveSheet.Cells(a, 3).Value2 = "IDI1" Or ActiveSheet.Cells(a, 3).Value2 = "FDPS" Or ActiveSheet.Cells(a, 3).Value2 = "NSDHL" Or ActiveSheet.Cells(a, 3).Value2 = "SC5D" Or ActiveSheet.Cells(a, 3).Value2 = "SC5DL" Or ActiveSheet.Cells(a, 3).Value2 = "SC4MOL" Or ActiveSheet.Cells(a, 3).Value2 = "EBP" Or ActiveSheet.Cells(a, 3).Value2 = "HSD17B7" Or ActiveSheet.Cells(a, 3).Value2 = "IDI2" Then
    'Row(a).Copy
    Range(Cells(a, 1), Cells(a, 424)).Copy
    Sheets("Cholesterol pathway analysis").Select
    With ActiveSheet
        lastrow = .Cells(.Rows.Count, 3).End(xlUp).Row
        Sheets("Cholesterol pathway analysis").Cells(lastrow + 1, 1).Select
        Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
        Application.CutCopyMode = False
        End With
    End If

Next a

End Sub
```

****

```
Sub Select_cholesterol_pathway_perturbation_from_array_May16()


Dim b As Long
Dim a As Long
Dim number As Integer


Worksheets("Pathway perturbation analysis").Select
For b = 5 To 52

   For a = 1 To 54615

      If ActiveSheet.Cells(a, 2).Value2 = "CYP51A1" Or ActiveSheet.Cells(a, 2).Value2 = "DHCR7" Or ActiveSheet.Cells(a, 2).Value2 = "DHCR24" Or ActiveSheet.Cells(a, 2).Value2 = "FDFT1" Or ActiveSheet.Cells(a, 2).Value2 = "HMGCR" Or ActiveSheet.Cells(a, 2).Value2 = "HMGCS1" Or ActiveSheet.Cells(a, 2).Value2 = "LSS" Or ActiveSheet.Cells(a, 2).Value2 = "MSMO1" Or ActiveSheet.Cells(a, 2).Value2 = "MVK" Or ActiveSheet.Cells(a, 2).Value2 = "LSS" Or ActiveSheet.Cells(a, 2).Value2 = "SQLE" Or ActiveSheet.Cells(a, 2).Value2 = "TM7SF2" Or ActiveSheet.Cells(a, 2).Value2 = "PMVK" Or ActiveSheet.Cells(a, 2).Value2 = "MVD" Or ActiveSheet.Cells(a, 2).Value2 = "IDI1" Or ActiveSheet.Cells(a, 2).Value2 = "FDPS" Or ActiveSheet.Cells(a, 2).Value2 = "NSDHL" Or ActiveSheet.Cells(a, 2).Value2 = "SC5D" Or ActiveSheet.Cells(a, 2).Value2 = "SC5DL" Or ActiveSheet.Cells(a, 2).Value2 = "SC4MOL" Or ActiveSheet.Cells(a, 2).Value2 = "EBP" Then

         If (ActiveSheet.Cells(a, b).Value / ActiveSheet.Cells(a, 55).Value >= 1.5) Then
         number = ActiveSheet.Cells(54617, b).Value2 + 1
         ActiveSheet.Cells(54617, b).Value2 = number

         Else

         If (ActiveSheet.Cells(a, b).Value / ActiveSheet.Cells(a, 55).Value <= 0.6666) Then
         number = ActiveSheet.Cells(54618, b).Value2 - 1
         ActiveSheet.Cells(54618, b).Value2 = number

         End If
         End If
      End If

Next a

Next b

End Sub
```

****

```
Sub Modified_Select_cholesterol_pathway_perturbation_from_array_May16()



Dim b As Integer
Dim a As Integer
Dim number As Integer
Dim number2 As Single


Worksheets("Cholesterol analysis").Select
For b = 4 To 220


  For a = 2 To 38
        number2 = 0
      If ActiveSheet.Cells(a, 3).Value2 = "CYP51A1" Or ActiveSheet.Cells(a, 3).Value2 = "DHCR7" Or ActiveSheet.Cells(a, 3).Value2 = "DHCR24" Or ActiveSheet.Cells(a, 3).Value2 = "FDFT1" Or ActiveSheet.Cells(a, 3).Value2 = "HMGCR" Or ActiveSheet.Cells(a, 3).Value2 = "HMGCS1" Or ActiveSheet.Cells(a, 3).Value2 = "LSS" Or ActiveSheet.Cells(a, 3).Value2 = "MSMO1" Or ActiveSheet.Cells(a, 3).Value2 = "MVK" Or ActiveSheet.Cells(a, 3).Value2 = "LSS" Or ActiveSheet.Cells(a, 3).Value2 = "SQLE" Or ActiveSheet.Cells(a, 3).Value2 = "TM7SF2" Or ActiveSheet.Cells(a, 3).Value2 = "PMVK" Or ActiveSheet.Cells(a, 3).Value2 = "MVD" Or ActiveSheet.Cells(a, 3).Value2 = "IDI1" Or ActiveSheet.Cells(a, 3).Value2 = "FDPS" Or ActiveSheet.Cells(a, 3).Value2 = "NSDHL" Or ActiveSheet.Cells(a, 3).Value2 = "SC5D" Then
        number2 = ActiveSheet.Cells(a, b).Value / ActiveSheet.Cells(a, 234).Value
        If (number2 >= 2) Then
        number = ActiveSheet.Cells(95, b).Value2 + 1
        ActiveSheet.Cells(95, b).Value2 = number


        Else


        If (number2 <= 0.5) Then
        number = ActiveSheet.Cells(96, b).Value2 - 1
        ActiveSheet.Cells(96, b).Value2 = number


        End If
        End If
      End If

Next a

Next b

End Sub
```

****

```
Sub Metabolites_Array_Aug16()


'declare array of variable 'metabolite' as string

Dim metabolite(1 To 191) As String
Dim sample(1 To 19) As String

metabolite(1) = "Adenine"
metabolite(2) = "aKG"
metabolite(3) = "Alanine"
metabolite(4) = "Alanine +1"
metabolite(5) = "Alanine +2"
metabolite(6) = "Alanine +3"
metabolite(7) = "Arginine"
metabolite(8) = "Arginine +1"
metabolite(9) = "Arginine +2"
metabolite(10) = "Arginine +3"
metabolite(11) = "Arginine +4"
metabolite(12) = "Arginine +5"
metabolite(13) = "Arginine +6"
metabolite(14) = "Asparagine"
metabolite(15) = "Asparagine +1"
metabolite(16) = "Asparagine +2"
metabolite(17) = "Asparagine +3"
metabolite(18) = "Asparagine +4"
metabolite(19) = "Aspartate"
metabolite(20) = "Aspartate +1"
metabolite(21) = "Aspartate +2"
metabolite(22) = "Aspartate +3"
metabolite(23) = "Aspartate+4"
metabolite(24) = "Cytidine"
metabolite(25) = "cis-aconitate"
metabolite(26) = "cis-aconitate +1"
metabolite(27) = "cis-aconitate +2"
metabolite(28) = "cis-aconitate +3"
metabolite(29) = "cis-aconitate +4"
metabolite(30) = "cis-aconitate +5"
metabolite(31) = "cis-aconitate +6"
metabolite(32) = "Glucose"
metabolite(33) = "Glycine"
metabolite(34) = "Glycine +1"
metabolite(35) = "Glycine +2"
metabolite(36) = "Lactate"
metabolite(37) = "homocysteine"
metabolite(38) = "Leucine"
metabolite(39) = "Leucine +1"
```

metabolite(40) = "Leucine +2"
metabolite(41) = "Leucine +3"
metabolite(42) = "Leucine +4"
metabolite(43) = "Leucine +5"
metabolite(44) = "Leucine +6"
metabolite(45) = "Lysine"
metabolite(46) = "Lysine+1"
metabolite(47) = "Lysine+2"
metabolite(48) = "Lysine+3"
metabolite(49) = "Lysine+4"
metabolite(50) = "Lysine+5"
metabolite(51) = "Lysine+6"
metabolite(52) = "methionine"
metabolite(53) = "Phenylalanine"
metabolite(54) = "proline"
metabolite(55) = "proline+1"
metabolite(56) = "Proline+2"
metabolite(57) = "Proline+3"
metabolite(58) = "Proline+4"
metabolite(59) = "Proline+5"
metabolite(60) = "Serine"
metabolite(61) = "Serine +1"
metabolite(62) = "Serine +2"
metabolite(63) = "Serine +3"
metabolite(64) = "threonine"
metabolite(65) = "tryptophan"
metabolite(66) = "tryptophan+1"
metabolite(67) = "tryptophan+2"
metabolite(68) = "tryptophan+3"
metabolite(69) = "tryptophan+4"
metabolite(70) = "tryptophan+5"
metabolite(71) = "tryptophan+6"
metabolite(72) = "tryptophan+7"
metabolite(73) = "tryptophan+8"
metabolite(74) = "tryptophan+9"
metabolite(75) = "tryptophan+10"
metabolite(76) = "tryptophan+11"
metabolite(77) = "Tyrosine"
metabolite(78) = "IsoLeucine"
metabolite(79) = "urate"
metabolite(80) = "Valine"
metabolite(81) = "Valine+1"
metabolite(82) = "Valine+2"
metabolite(83) = "Valine+3"
metabolite(84) = "Valine+4"
metabolite(85) = "Valine+5"
metabolite(86) = "xanthine"

metabolite(87) = "homocysteine"

metabolite(88) = "Fumarate"

metabolite(89) = "Fumarate +1"

metabolite(90) = "Fumarate +2"

metabolite(91) = "Fumarate +3"

metabolite(92) = "Fumarate +4"

metabolite(93) = "Ornithine"

metabolite(94) = "Ornithine+1"

metabolite(95) = "Ornithine+2"

metabolite(96) = "Ornithine+3"

metabolite(97) = "Ornithine+4"

metabolite(98) = "Ornithine+5"

metabolite(99) = "Lactate +3"

metabolite(100) = "Glucose +6"

metabolite(101) = "Alanine +1"

metabolite(102) = "Arginine +6"

metabolite(103) = "3-hydroxybutyric acid"

metabolite(104) = "Acetyl CoA"

metabolite(105) = "Acetyl CoA +1"

metabolite(106) = "Acetyl CoA +2"

metabolite(107) = "aKG +1"

metabolite(108) = "aKG +2"

metabolite(109) = "aKG +3"

metabolite(110) = "aKG +4"

metabolite(111) = "aKG +5"

metabolite(112) = "Arachidonic acid +1"

metabolite(113) = "Arachidonic acid +2"

metabolite(114) = "Arachidonic acid"

metabolite(115) = "Citric acid +1"

metabolite(116) = "Citric acid +2"

metabolite(117) = "Citric acid +3"

metabolite(118) = "Citric acid +4"

metabolite(119) = "Citric acid +5"

metabolite(120) = "Citric acid +6"

metabolite(121) = "Citric acid"

metabolite(122) = "dihydroxyacetone phosphate (DHAP)"

metabolite(123) = "dihydroxyacetone phosphate (DHAP) +1"

metabolite(124) = "dihydroxyacetone phosphate (DHAP) +2"

metabolite(125) = "Dihydroxyacetone Phosphate (DHAP)+3"

metabolite(126) = "Docosahexaenoic acid"

metabolite(127) = "Eicosapentaenoic acid"

metabolite(128) = "fructose 1,6-diphosphate"

metabolite(129) = "Fructose 1,6-bisphosphate+1"

metabolite(130) = "Fructose 1,6-bisphosphate+6"

metabolite(131) = "G6P"

metabolite(132) = "G6P +1"

metabolite(133) = "G6P +2"

metabolite(134) = "G6P +3"

metabolite(135) = "G6P +4"

metabolite(136) = "G6P +5"

metabolite(137) = "G6P +6"

metabolite(138) = "GLN"

metabolite(139) = "GLN +1"

metabolite(140) = "GLN +2"

metabolite(141) = "GLN +3"

metabolite(142) = "GLN +4"

metabolite(143) = "GLN +5"

metabolite(144) = "GLU"

metabolite(145) = "GLU+1"

metabolite(146) = "GLU+2"

metabolite(147) = "GLU+3"

metabolite(148) = "GLU+4"

metabolite(149) = "GLU+5"

metabolite(150) = "Glyceraldehyde 3-phosphate"

metabolite(151) = "Glyceraldehyde 3-phosphate +1"

metabolite(152) = "Glyceraldehyde 3-phosphate +2"

metabolite(153) = "Glyceraldehyde 3-phosphate +3"

metabolite(154) = "Glycerol 3-phosphate"

metabolite(155) = "Glycerol 3-phosphate+1"

metabolite(156) = "Glycerol 3-phosphate+2"

metabolite(157) = "Glycerol 3-phosphate+3"

metabolite(158) = "Lactate"

metabolite(159) = "Lactate +1"

metabolite(160) = "Lactate +2"

metabolite(161) = "Linoleic acid"

metabolite(162) = "Linolenic acid"

metabolite(163) = "Malate"

metabolite(164) = "Malate +1"

metabolite(165) = "Malate +2"

metabolite(166) = "Malate +3"

metabolite(167) = "Malate +4"

metabolite(168) = "Mevalonic Acid"

metabolite(169) = "Mevalonic Acid +1"

metabolite(170) = "Mevalonic Acid +2"

metabolite(171) = "Mevalonic Acid +3"

metabolite(172) = "Mevalonic Acid +4"

metabolite(173) = "Mevalonic Acid +5"

metabolite(174) = "Mevalonic Acid +6"

metabolite(175) = "Oleic acid"

metabolite(176) = "PEP"

metabolite(177) = "PEP+1"

metabolite(178) = "PEP+2"

metabolite(179) = "PEP+3"

metabolite(180) = "Palmitic acid"

```
metabolite(181) = "Palmitoleic acid"
metabolite(182) = "Pyruvic acid"
metabolite(183) = "Pyruvic acid +1"
metabolite(184) = "Pyruvic acid +2"
metabolite(185) = "Pyruvic acid +3"
metabolite(186) = "Stearic acid"
metabolite(187) = "Succinic acid"
metabolite(188) = "Succinic acid +1"
metabolite(189) = "Succinic acid +2"
metabolite(190) = "Succinic acid +3"
metabolite(191) = "Succinic acid +4"

sample(1) = "5_L_CNS"
sample(2) = "5_L_SPL"
sample(3) = "5_LR_CNS"
sample(4) = "5_LR_SPL"
sample(5) = "5_R_CNS"
sample(6) = "5_R_SPL"
sample(7) = "4-2L_CNS"
sample(8) = "4-2L_SPL"
sample(9) = "4-L_CNS"
sample(10) = "4-L_SPL"
sample(11) = "4-LR_CNS"
sample(12) = "4-LR_SPL"
sample(13) = "5-2L_CNS"
sample(14) = "5-2L_SPL"
sample(15) = "5-L_CNS"
sample(16) = "5-L_SPL"
sample(17) = "R_CNS"
sample(18) = "R_SPL"
sample(19) = "Blank_Water"



'set sheet1 as "All samples"
If ActiveSheet.Name <> "All Samples" Then ActiveSheet.Name = "All Samples"

Dim i As Integer                    ' Integers used in 'For' loops
Dim j As Integer
Dim k As Integer
Dim m As Integer
Dim n As Integer

Dim p As String

'Sheets.Add After:=Sheets("All Samples")
ActiveSheet.Name = "Array" 'creates first worksheet
```

```
'name rows/columns
For i = 1 To 191

Cells(i + 1, 1).Value = metabolite(i)
Next i

For j = 1 To 19
Cells(1, j + 2).Value = sample(j)
Next j

'populate array
Worksheets("All Samples").Select

For k = 2 To 4325

For m = 56 To 59

For n = 1 To 19

    Worksheets("All Samples").Select
    If Cells(k, 1).Value = metabolite(m) And Cells(k, 4).Value = sample(n) Then
    p = Cells(k, 15).Value
    Worksheets("Array").Select
    Cells(m + 1, n + 2).Value = p
    End If
    Next n
    Next m
    Next k


End Sub
```
****

```
Sub Fluidigm_Array_Jan17()

'declare array of variable 'metabolite' as string

Dim Gene(1 To 48) As String
Dim sample(1 To 22) As String

Gene(1) = "ATP5b"
Gene(2) = "CTP2"
Gene(3) = "ACLY"
Gene(4) = "CTP1b"
Gene(5) = "SCD"
Gene(6) = "ACACA"
Gene(7) = "CPT1a"
Gene(8) = "B2M"
Gene(9) = "FASN"
Gene(10) = "HMGCR(1)"
Gene(11) = "HMGCS1"
Gene(12) = "HMGCR(2)"
Gene(13) = "CYP51"
Gene(14) = "DHCR7"
Gene(15) = "DHCR24"
Gene(16) = "RNF20"
Gene(17) = "FDFT1"
Gene(18) = "LSS"
Gene(19) = "MSMO"
Gene(20) = "MVK"
Gene(21) = "SQLE"
Gene(22) = "GAPDH"
Gene(23) = "TM7SF"
Gene(24) = "ENOX2"
Gene(25) = "PMVK"
Gene(26) = "MVD"
Gene(27) = "IDI1"
Gene(28) = "FDPS"
Gene(29) = "NSDHL"
Gene(30) = "SC5D"
Gene(31) = "HSD17b"
Gene(32) = "UBE2D2"
Gene(33) = "EBP"
Gene(34) = "IDI2"
Gene(35) = "SREBF1c"
Gene(36) = "SREBF1a"
Gene(37) = "SREBF2"
Gene(38) = "SPP1"
Gene(39) = "IRF4"
Gene(40) = "CYC1"
```

```
Gene(41) = "VEGFa"
Gene(42) = "ACSS2"
Gene(43) = "IL15"
Gene(44) = "ACSS1"
Gene(45) = "ICAM1"
Gene(46) = "MERTK"
Gene(47) = "PBX1"
Gene(48) = "TYW1"

sample(1) = "49 C1"
sample(2) = "49 C2"
sample(3) = "49 C3"
sample(4) = "11 C1"
sample(5) = "12 C1"
sample(6) = "12 C2"
sample(7) = "12 C3"
sample(8) = "66 C1"
sample(9) = "66 C2"
sample(10) = "36 C1"
sample(11) = "Mouse"
sample(12) = ".-RT"
sample(13) = "49 S1"
sample(14) = "49 S2"
sample(15) = "49 S3"
sample(16) = "11 S1"
sample(17) = "12 S1"
sample(18) = "12 S2"
sample(19) = "12 S3"
sample(20) = "66 S1"
sample(21) = "66 S2"
sample(22) = "36 S1"

'set sheet1 as "All samples"
If ActiveSheet.Name <> "Array1" Then ActiveSheet.Name = "Array1"

Dim i As Integer                    ' Integers used in 'For' loops
Dim j As Integer
Dim k As Integer
Dim m As Integer
Dim n As Integer

Dim p As String

Sheets.Add After:=Sheets("Array1")
ActiveSheet.Name = "Fluidigm Ct Array" 'creates first worksheet

'name rows/columns
```

```
For i = 1 To 48

Cells(i + 1, 1).Value = Gene(i)

Next i

For j = 1 To 22
Cells(1, j + 2).Value = sample(j)
Next j

'populate array 2 (second samples)
Worksheets("Array1").Select

For k = 13 To 916

For m = 1 To 48

For n = 1 To 22

    Worksheets("Array1").Select
    If Cells(k, 5).Value = Gene(m) And Cells(k, 2).Value = sample(n) Then
    p = Cells(k, 7).Value
    Worksheets("Fluidigm Ct Array").Select
    Cells(m + 1, n + 2).Value = p
    End If
    Next n
    Next m
    Next k

'populate array
Worksheets("Array1").Select

For k = 917 To 1777
For m = 1 To 48
For n = 1 To 22

    Worksheets("Array1").Select
    If Cells(k, 5).Value = Gene(m) And Cells(k, 2).Value = sample(n) Then
    p = Cells(k, 7).Value
    Worksheets("Fluidigm Ct Array").Select
    Cells(m + 1, n + 26).Value = p
    End If
    Next n
    Next m
    Next k

For k = 1778 To 2316
```

```
For m = 1 To 48
For n = 1 To 22

    Worksheets("Array1").Select
    If Cells(k, 5).Value = Gene(m) And Cells(k, 2).Value = sample(n) Then
    p = Cells(k, 7).Value
    Worksheets("Fluidigm Ct Array").Select
    Cells(m + 1, n + 50).Value = p
    End If
    Next n
    Next m
    Next k
End Sub
```

****

```
Sub Select_cholesterol_pathway_perturbation_from_array_Apr17()

'Dim b As Long
Dim a As Long
'Dim number As Integer
Dim lastrow As Integer

Sheets.Add After:=Sheets("SEMresACMarch17")
ActiveSheet.Name = "Cholesterol genes" 'creates second worksheet"

Worksheets("SEMresACMarch17").Select
'For b = 5 To 52

    For a = 1 To 34542

    Worksheets("SEMresACMarch17").Select
        If ActiveSheet.Cells(a, 2).Value2 = "CYP51A1" Or ActiveSheet.Cells(a, 2).Value2 = "DHCR7" Or ActiveSheet.Cells(a, 2).Value2 = "DHCR24" Or ActiveSheet.Cells(a, 2).Value2 = "FDFT1" Or ActiveSheet.Cells(a, 2).Value2 = "HMGCR" Or ActiveSheet.Cells(a, 2).Value2 = "HMGCS1" Or ActiveSheet.Cells(a, 2).Value2 = "LSS" Or ActiveSheet.Cells(a, 2).Value2 = "MSMO1" Or ActiveSheet.Cells(a, 2).Value2 = "MVK" Or ActiveSheet.Cells(a, 2).Value2 = "LSS" Or ActiveSheet.Cells(a, 2).Value2 = "SQLE" Or ActiveSheet.Cells(a, 2).Value2 = "TM7SF2" Or ActiveSheet.Cells(a, 2).Value2 = "PMVK" Or ActiveSheet.Cells(a, 2).Value2 = "MVD" Or ActiveSheet.Cells(a, 2).Value2 = "IDI1" Or ActiveSheet.Cells(a, 2).Value2 = "FDPS" Or ActiveSheet.Cells(a, 2).Value2 = "NSDHL" Or ActiveSheet.Cells(a, 2).Value2 = "SC5D" Or ActiveSheet.Cells(a, 2).Value2 = "SC5DL" Or ActiveSheet.Cells(a, 2).Value2 = "SC4MOL" Or ActiveSheet.Cells(a, 2).Value2 = "EBP" Or ActiveSheet.Cells(a, 2).Value2 = "IDI2" Or ActiveSheet.Cells(a, 2).Value2 = "HSD17B7" Then

        Range(Cells(a, 1), Cells(a, 8)).Copy
        Sheets("Cholesterol genes").Select
        With ActiveSheet
            lastrow = .Cells(.Rows.Count, 3).End(xlUp).Row
            Sheets("Cholesterol genes").Cells(lastrow + 1, 1).Select
            Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
            Application.CutCopyMode = False
            End With
            End If
            Next a
            End Sub
```

****

```
Sub Select_cholesterol_pathway_perturbation_from_array_REH_Apr17()

'Dim b As Long
Dim a As Long
'Dim number As Integer
Dim lastrow As Integer

Sheets.Add After:=Sheets("REHresACMarch17")
ActiveSheet.Name = "Cholesterol genes" 'creates second worksheet"

Worksheets("REHresACMarch17").Select
'For b = 5 To 52

  For a = 1 To 38612

  Worksheets("REHresACMarch17").Select
    If ActiveSheet.Cells(a, 2).Value2 = "CYP51A1" Or ActiveSheet.Cells(a, 2).Value2 = "DHCR7" Or ActiveSheet.Cells(a, 2).Value2 = "DHCR24" Or ActiveSheet.Cells(a, 2).Value2 = "FDFT1" Or ActiveSheet.Cells(a, 2).Value2 = "HMGCR" Or ActiveSheet.Cells(a, 2).Value2 = "HMGCS1" Or ActiveSheet.Cells(a, 2).Value2 = "LSS" Or ActiveSheet.Cells(a, 2).Value2 = "MSMO1" Or ActiveSheet.Cells(a, 2).Value2 = "MVK" Or ActiveSheet.Cells(a, 2).Value2 = "LSS" Or ActiveSheet.Cells(a, 2).Value2 = "SQLE" Or ActiveSheet.Cells(a, 2).Value2 = "TM7SF2" Or ActiveSheet.Cells(a, 2).Value2 = "PMVK" Or ActiveSheet.Cells(a, 2).Value2 = "MVD" Or ActiveSheet.Cells(a, 2).Value2 = "IDI1" Or ActiveSheet.Cells(a, 2).Value2 = "FDPS" Or ActiveSheet.Cells(a, 2).Value2 = "NSDHL" Or ActiveSheet.Cells(a, 2).Value2 = "SC5D" Or ActiveSheet.Cells(a, 2).Value2 = "SC5DL" Or ActiveSheet.Cells(a, 2).Value2 = "SC4MOL" Or ActiveSheet.Cells(a, 2).Value2 = "EBP" Or ActiveSheet.Cells(a, 2).Value2 = "IDI2" Or ActiveSheet.Cells(a, 2).Value2 = "HSD17B7" Then

    Range(Cells(a, 1), Cells(a, 8)).Copy
    Sheets("Cholesterol genes").Select
    With ActiveSheet
          lastrow = .Cells(.Rows.Count, 3).End(xlUp).Row
          Sheets("Cholesterol genes").Cells(lastrow + 1, 1).Select
          Selection.PasteSpecial Paste:=xlPasteAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=False
          Application.CutCopyMode = False
          End With
          End If
          Next a
          End Sub
```

****

R scripts

```
#SEM/REH cells DESEQ2 Analysis

#http://bioconductor.org/packages/devel/bioc/vignettes/DESeq2/inst/doc/DESeq2.html
.libPaths("C:/R/win-library/3.3")
library(BiocInstaller)
biocLite(lib.loc = "C:/R/win-library/3.3", lib="C:/R/win-library/3.3")
biocValid() library(readr)
library(DESeq2)
library("AnnotationDbi")
library("org.Hs.eg.db")
####
SEM1 <- read_delim("C:/list_human.csv.gene.count-with-in-vitro.csv",
            ",", escape_double = FALSE, trim_ws = TRUE)

SEM1 <- as.data.frame(SEM1)
row.names(SEM1) <- SEM1$X1
SEM1$X1 <- NULL
SEM1 <- as.matrix(SEM1)



countData1 <- as.matrix(SEM1)
storage.mode(countData1) = "integer"

colData1 <- read.delim2("C:/SEM.colData1.txt",sep="\t",row.names=1)

SEM1dds <- DESeqDataSetFromMatrix(countData = countData1,
                          colData = colData1,
                          design = ~ Site)
SEM1dds
SEM1dds <- DESeq(SEM1dds)

resSEM1 <- results(SEM1dds, contrast=c("Site","CNS","SPLEEN"))
summary(resSEM1)
SEM1res <- as.data.frame(resSEM1)

SEM1res$symbol <- mapIds(org.Hs.eg.db,
                  keys=row.names(SEM1res),
                  column="SYMBOL",
                  keytype="ENSEMBL",
                  multiVals="first")

rld1<- rlogTransformation(SEM1dds, blind=TRUE)
vsd1<-varianceStabilizingTransformation(SEM1dds, blind=TRUE)

library('RColorBrewer')
library('gplots')
```

```
select <- order(rowMeans(counts(SEM1dds,normalized=TRUE)),decreasing=TRUE)[1:30]
hmcol<- colorRampPalette(brewer.pal(9, 'GnBu'))(100)
heatmap.2(counts(SEM1dds,normalized=TRUE)[select,], col = hmcol,
        Rowv = FALSE, Colv = FALSE, scale='none',
        dendrogram='none', trace='none', margin=c(10,6))
dev.copy(svg,'SEM1DESeq2_heatmap1.svg')
dev.off()
heatmap.2(assay(rld1)[select,], col = hmcol,
        Rowv = FALSE, Colv = FALSE, scale='none',
        dendrogram='none', trace='none', margin=c(10, 6))
dev.copy(svg,'SEM1DESeq2_heatmap2.svg')
dev.off()
heatmap.2(assay(vsd1)[select,], col = hmcol,
        Rowv = FALSE, Colv = FALSE, scale='none',
        dendrogram='none', trace='none', margin=c(10, 6))
dev.copy(svg,'SEM1DESeq2_heatmap3.svg')
dev.off()
###########

distsRL <- dist(t(assay(rld1)))
mat<- as.matrix(distsRL)
hc <- hclust(distsRL)
heatmap.2(mat, Rowv=as.dendrogram(hc),
        symm=TRUE, trace='none',
        col = rev(hmcol), margin=c(13, 13))
dev.copy(svg,'SEM1deseq2_heatmaps_samplebysample.svg')
dev.off()
############
print(plotPCA(rld1, intgroup=c('Site')))
dev.copy(svg,'SEM1deseq2_pca.svg')
dev.off()
##########
plotMA(SEM1dds,ylim=c(-2,2),main='DESeq2')
dev.copy(svg,'SEM1deseq2_MAplot.svg')
dev.off()
```

```
############################################################
#SEM/REH analysis without TC control

.libPaths("C:/R/win-library/3.3")
library(readr)
library(DESeq2)
####
SEM <- read_delim("C:/list_human.csv.gene.count.csv",
            ",", escape_double = FALSE, trim_ws = TRUE)

SEM <- as.data.frame(SEM)
row.names(SEM) <- SEM$X1
SEM$X1 <- NULL
SEM <- as.matrix(SEM)

countData <- as.matrix(SEM)
storage.mode(countData) = "integer"

colData <- read.delim2("C:/SEM.colData.txt",sep="\t",row.names=1)

SEMdds <- DESeqDataSetFromMatrix(countData = countData,
                    colData = colData,
                    design = ~ Site)
SEMdds$Site <- relevel(SEMdds$Site, ref="SPLEEN")
SEMdds

SEMdds <- DESeq(SEMdds)
SEMdds$Site <- relevel(SEMdds$Site, ref="SPLEEN")

resSEM <- results(SEMdds)
summary(resSEM)
SEMres <- as.data.frame(resSEM)

rld<- rlogTransformation(SEMdds, blind=TRUE)
vsd<-varianceStabilizingTransformation(SEMdds, blind=TRUE)

#####
dev.off()
library('RColorBrewer')
library('gplots')
select <- order(rowMeans(counts(SEMdds,normalized=TRUE)),decreasing=TRUE)[1:30]
hmcol<- colorRampPalette(brewer.pal(9, 'GnBu'))(100)
heatmap.2(counts(SEMdds,normalized=TRUE)[select,], col = hmcol,
     Rowv = FALSE, Colv = FALSE, scale='none',
     dendrogram='none', trace='none', margin=c(10,6))
dev.copy(svg,'SEM-DESeq2_heatmap1.svg')
dev.off()
```

```
heatmap.2(assay(rld)[select,], col = hmcol,
        Rowv = FALSE, Colv = FALSE, scale='none',
        dendrogram='none', trace='none', margin=c(10, 6))
dev.copy(svg,'SEM-DESeq2_heatmap2.svg')
dev.off()
heatmap.2(assay(vsd)[select,], col = hmcol,
        Rowv = FALSE, Colv = FALSE, scale='none',
        dendrogram='none', trace='none', margin=c(10, 6))
dev.copy(svg,'SEM-DESeq2_heatmap3.svg')
dev.off()
###########

distsRL <- dist(t(assay(rld)))
mat<- as.matrix(distsRL)
hc <- hclust(distsRL)
heatmap.2(mat, Rowv=as.dendrogram(hc),
        symm=TRUE, trace='none',
        col = rev(hmcol), margin=c(13, 13))
dev.copy(svg,'SEM-deseq2_heatmaps_samplebysample.svg')
dev.off()
############
print(plotPCA(rld, intgroup=c('Site')))
dev.copy(svg,'SEM-deseq2_pca.svg')
dev.off()
################
dev.off()
plotMA(SEMdds,ylim=c(-2,2),main='DESeq2')
dev.copy(svg,'SEM-deseq2_MAplot.svg')
dev.off()
```

```
###############################################################
#Cholesterol gene waterfall plots

.libPaths("C:/R/win-library/3.3")
library(readr)
library("ggplot2", lib.loc="C:/R/win-library/3.3")
library('RColorBrewer')
dev.off()
Chol <- read_csv("C:/Cholesterol genes.csv")

Chol <- as.data.frame(Chol)
Chol$`-Log10 Adjusted p-value` <- -log10(Chol$`Adjusted p-value`)

Chol$X1 <- factor(Chol$X1, levels = Chol$X1[order(Chol$log2FoldChange)])

bccol<- colorRampPalette(brewer.pal(9, 'GnBu'))(100)

ggplot(Chol, aes(x=X1, y=log2FoldChange, fill=`-Log10 Adjusted p-value`)) + geom_bar(colour="black",stat="identity") +
  xlab("Gene") +   ylab("Log2 Fold-change") +
  ggtitle("Cholesterol synthesis genes") + coord_flip() +
  scale_fill_gradientn(colours=topo.colors(7),
                       breaks=c(1.3, 8, 14),
                       labels=c("1.3 (Adj. p=0.05)","8","14"), limits=c(1.3,14), na.value = "white")

ggplot(Chol, aes(x=X1, y=log2FoldChange, fill=`-Log10 Adjusted p-value`)) + geom_bar(colour="black",stat="identity") +
  xlab("Gene") +   ylab("Log2 Fold-change, CNS:Spleen") +
  ggtitle("Cholesterol synthesis genes") + coord_flip() +
  scale_fill_gradientn(colours=bccol,
                       breaks=c(1.3, 8, 14),
                       labels=c("1.3 (Adj. p=0.05)","8","14"), limits=c(1.3,14),na.value = "transparent")

svg(file="C:/Cholesterol genes Barchart.svg")
ggplot(Chol, aes(x=X1, y=log2FoldChange, fill=`-Log10 Adjusted p-value`)) + geom_bar(colour="black",stat="identity") +
  xlab("Gene") +   ylab("Log2 Fold-change, CNS:Spleen") +
  ggtitle("Cholesterol synthesis genes") + coord_flip() +
  scale_fill_gradientn(colours=bccol,
                       breaks=c(1.3, 8, 14),
                       labels=c("1.3 (Adj. p=0.05)","8","14"), limits=c(1.3,14),na.value = "transparent")
dev.off()
```

```
############################################################
#Venn diagram

library(VennDiagram)
library(gridExtra)
library(readxl)

RNASeqData <- read_excel("C:/CombinedRNASeqData.xlsx")

REH <- as.data.frame(RNASeqData$REH)
REH <- subset(REH,!NA)
SEM <- as.data.frame(RNASeqData$SEM)
REHandSEM <- as.data.frame(RNASeqData$Both)
REH <- subset(REH!NA)

dev.off()
grid.newpage()
h = draw.pairwise.venn(area1 = nrow(na.omit(REH)), area2 = nrow(na.omit(SEM)),
             , cross.area =  nrow(na.omit(REHandSEM)),fontfamily = "Arial", fontface = "bold", cat.fontfamily = "Arial",
cat.fontface = "bold",
             category = c("REH", "SEM"), lty = rep("solid", 2), lwd=1 ,
             fill = c("#0177e1",
                  "#ff949e"), ext.text=FALSE)#scaled = FALSE)
grid.arrange(gTree(children=h), top="GSEA of RNA Seq\n data using\nKEGG pathways")

dev.off()
svg(file="C:/SEM and REH KEGG pathways venn.svg")
grid.newpage()
grid.arrange(gTree(children=h), top="GSEA of RNA Seq\n data using\nKEGG pathways")
dev.off()
```

```
############################################################
#Cholesterol data from multiplex PCR
library(readr)
library(readxl)
library("ggplot2", lib.loc="C:/R/win-library/3.3")
library('RColorBrewer')
dev.off()
FluiChol <- read_excel("C:/TC Jan 17 Fluidigm chip analysis Chol-lipid export Foldchanges.xlsx")

FluiChol <- as.data.frame(FluiChol)
FluiChol$`-Log10 p-value` <- -log10(FluiChol$`p-value`)

FluiChol$Gene <- factor(FluiChol$Gene, levels = FluiChol$Gene[order(FluiChol$`Mean FC`)])

bccol<- colorRampPalette(brewer.pal(9, 'GnBu'))(100)

ggplot(FluiChol, aes(x=Gene, y=`Mean FC`, fill=`-Log10 p-value`)) + geom_bar(colour="black",stat="identity") +
  xlab("Gene") +   ylab("Mean FC") +
  ggtitle("FluiCholesterol synthesis genes") + coord_flip() +
  scale_fill_gradientn(colours=topo.colors(7),
                   breaks=c(1.3, 8, 14),
                   labels=c("1.3 (p=0.05)","8","14"), limits=c(1.3,14), na.value = "white")

ggplot(FluiChol, aes(x=Gene, y=`Mean FC`, fill=`-Log10 p-value`)) + geom_bar(colour="black",stat="identity") +
  xlab("Gene") +   ylab("Log2 Fold-change, CNS:Spleen") +
  ggtitle("FluiCholesterol synthesis genes") + coord_flip() +
  scale_fill_gradientn(colours=bccol,
                   breaks=c(1.3, 8, 14),
                   labels=c("1.3 (p=0.05)","8","14"), limits=c(1.3,14),na.value = "transparent")
#########
ggplot() + geom_bar(data=FluiChol, aes(x=Gene, y=`Mean FC`, fill=`-Log10 p-value`), colour="black",stat="identity") +
  geom_point(data=FluiChol, aes(x=Gene, y=c(`Patient 1`))) +
  geom_point(data=FluiChol, aes(x=Gene, y=c(`Patient 2`))) +
  geom_point(data=FluiChol, aes(x=Gene, y=c(`Patient 3`))) +
  geom_point(data=FluiChol, aes(x=Gene, y=c(`Patient 4`))) +
  geom_point(data=FluiChol, aes(x=Gene, y=c(`Patient 5`))) +
  xlab("Gene") +   ylab("Log2 Fold-change, CNS:Spleen") +
  ggtitle("FluiCholesterol synthesis genes") + coord_flip(ylim=c(-2.5,2.5)) +
  scale_fill_gradientn(colours=bccol,
                   breaks=c(1.3, 8, 14),
                   labels=c("1.3 (p=0.05)","8","14"), limits=c(1.3,14),na.value = "transparent")
#################

svg(file="C:/FluiCholesterol genes Barchart.svg")
ggplot(FluiChol, aes(x=Gene, y=`Mean FC`, fill=`-Log10 p-value`)) + geom_bar(colour="black",stat="identity") +
  xlab("Gene") +   ylab("Log2 Fold-change, CNS:Spleen") +
  ggtitle("FluiCholesterol synthesis genes") + coord_flip() +
```

```
  scale_fill_gradientn(colours=bccol,
                       breaks=c(1.3, 8, 14),
                       labels=c("1.3 (p=0.05)","8","14"), limits=c(1.3,14),na.value = "transparent")
dev.off()
```

```
############################################################
#Cholesterol data from multiplex PCR - idividual graphs (similar script used for Fatty acid and SREBP/ACCS gene analysis)

library(readr)
library(readxl)
library("ggplot2", lib.loc="C:/R/win-library/3.3")
library('RColorBrewer')
library(gridExtra)
library(grid)
dev.off()
FluiChols <- read_excel("C:/TC Jan 17 Fluidigm chip analysis Chol-lipid export(analysis4) Foldchanges.xlsx")
FluiChols <- as.data.frame(FluiChols)

rownames(FluiChols) <- FluiChols$Gene
FluiChols <- subset(FluiChols[,2:10])
FluiChols <- log(FluiChols,2)

p1Chol <- subset(FluiChols[,1:3])
p1Chol$Mean <- rowMeans(p1Chol)
p2Chol <- FluiChols[c(-1,-2,-3,-5,-6,-7,-8,-9)]
p3Chol <- subset(FluiChols[,5:7])
p3Chol$Mean <- rowMeans(p3Chol)
p4Chol <- FluiChols[c(-1,-2,-3,-4,-5,-6,-7,-9)]
p5Chol <- FluiChols[c(-1,-2,-3,-4,-5,-6,-7,-8)]

p1Chol$names <- factor(row.names(p1Chol), levels = row.names(p1Chol))
p1Chol$names = with(p1Chol, factor(names, levels = rev(levels(names))))

p1 <- ggplot(p1Chol, aes(x=p1Chol$names, y=`Mean`)) + geom_bar(colour="black",fill="blue",stat="identity") +
  xlab("Gene") +   ylab("Mean Log2 FC")  +
  ggtitle("Patient 1") + coord_flip(ylim=c(-2.5,2.5))

p2 <- ggplot(p2Chol, aes(x=p1Chol$names, y=`Patient 2`)) + geom_bar(colour="black",fill="blue",stat="identity")  +
  xlab("Gene") +   ylab("Mean Log2 FC") +
  ggtitle("Patient 2") + coord_flip(ylim=c(-2.5,2.5))

p3 <- ggplot(p3Chol, aes(x=p1Chol$names, y=`Mean`)) + geom_bar(colour="black",fill="blue",stat="identity") +
  xlab("Gene") +   ylab("Mean Log2 FC") +
  ggtitle("Patient 3") + coord_flip(ylim=c(-2.5,2.5))

p4 <- ggplot(p4Chol, aes(x=p1Chol$names, y=`Patient 4`)) + geom_bar(colour="black",fill="blue",stat="identity") +
  xlab("Gene") +   ylab("Mean Log2 FC") +
  ggtitle("Patient 4") + coord_flip(ylim=c(-2.5,2.5))

p5 <- ggplot(p5Chol, aes(x=p1Chol$names, y=`Patient 5`)) + geom_bar(colour="black",fill="blue",stat="identity") +
  xlab("Gene") +   ylab("Mean Log2 FC") +
  ggtitle("Patient 5") + coord_flip(ylim=c(-2.5,2.5))
```

```
grid.arrange(p1, p2, p3, p4, p5, ncol = 3)
################################

FluiChol$Gene <- factor(FluiChol$Gene, levels = FluiChol$Gene[order(FluiChol$`Mean FC`)])

p1Chol$names <- factor(row.names(p1Chol), levels = row.names(p1Chol))
p1Chol$names = with(p1Chol, factor(names, levels = p1Chol$names[order(p1Chol$Mean)]))
p2Chol$names <- factor(row.names(p1Chol), levels = row.names(p1Chol))
p2Chol$names = with(p2Chol, factor(names, levels = p2Chol$names[order(p2Chol$`Patient 2`)]))

p3Chol$names <- factor(row.names(p1Chol), levels = row.names(p1Chol))
p3Chol$names = with(p3Chol, factor(names, levels = p3Chol$names[order(p3Chol$Mean)]))
p4Chol$names <- factor(row.names(p1Chol), levels = row.names(p1Chol))
p4Chol$names = with(p4Chol, factor(names, levels = p4Chol$names[order(p4Chol$`Patient 4`)]))
p5Chol$names <- factor(row.names(p1Chol), levels = row.names(p1Chol))
p5Chol$names = with(p5Chol, factor(names, levels = p5Chol$names[order(p5Chol$`Patient 5`)]))

p1 <- ggplot(p1Chol, aes(x=p1Chol$names, y=`Mean`)) +
  xlab("Gene") +   ylab("Mean Log2 FC")  +
  geom_bar(stat = 'identity', colour="black", aes(fill = `Mean`>0), position = 'dodge') +
  scale_fill_manual(values=c("red","blue"),guide = 'none') +
  ggtitle("Patient 1") + coord_flip(ylim=c(-2.5,2.5))

p2 <- ggplot(p2Chol, aes(x=p2Chol$names, y=`Patient 2`)) + geom_bar(colour="black",fill="blue",stat="identity")  +
  xlab("Gene") +   ylab("Mean Log2 FC") +
  geom_bar(stat = 'identity', colour="black", aes(fill = `Patient 2`>0), position = 'dodge') +
  scale_fill_manual(values=c("red","blue"),guide = 'none') +
  ggtitle("Patient 2") + coord_flip(ylim=c(-2.5,2.5))

p3 <- ggplot(p3Chol, aes(x=p3Chol$names, y=`Mean`)) + geom_bar(colour="black",fill="blue",stat="identity") +
  xlab("Gene") +   ylab("Mean Log2 FC") +
  geom_bar(stat = 'identity', colour="black", aes(fill = `Mean`>0), position = 'dodge') +
  scale_fill_manual(values=c("red","blue"),guide = 'none') +
  ggtitle("Patient 3") + coord_flip(ylim=c(-2.5,2.5))

p4 <- ggplot(p4Chol, aes(x=p4Chol$names, y=`Patient 4`)) + geom_bar(colour="black",fill="blue",stat="identity") +
  xlab("Gene") +   ylab("Mean Log2 FC") +
  geom_bar(stat = 'identity', colour="black", aes(fill = `Patient 4`>0), position = 'dodge') +
  scale_fill_manual(values=c("red","blue"),guide = 'none') +
  ggtitle("Patient 4") + coord_flip(ylim=c(-2.5,2.5))

p5 <- ggplot(p5Chol, aes(x=p5Chol$names, y=`Patient 5`)) + geom_bar(colour="black",fill="blue",stat="identity") +
  xlab("Gene") +   ylab("Mean Log2 FC") +
  geom_bar(stat = 'identity', colour="black", aes(fill = `Patient 5`>0), position = 'dodge') +
  scale_fill_manual(values=c("red","blue"),guide = 'none') +
  ggtitle("Patient 5") + coord_flip(ylim=c(-2.5,2.5))
```

```
grid.arrange(p1, p2, p3, p4, p5, ncol = 3, top=textGrob("Cholesterol synthesis genes Log2 fold-change \n CNS:Spleen by individual patient sample", gp=gpar(fontsize=15, font =2)))


svg(file="C:/Fluidigm individual patient cholesterol genes Barchart.svg")

grid.arrange(p1, p2, p3, p4, p5, ncol = 3, top=textGrob("Cholesterol synthesis genes Log2 fold-change \n CNS:Spleen by individual patient sample", gp=gpar(fontsize=15, font =2)))

dev.off()

################

p1 <- ggplot(p1Chol, aes(x=p1Chol$names, y=`Mean`)) +

  xlab("Gene") +   ylab("Mean FC")  +

  geom_bar(stat = 'identity', colour="black", aes(fill = `Mean`>0), position = 'dodge') +

  scale_fill_manual(values=c("red","blue"),guide = 'none') +

  ggtitle("Patient 1") + coord_flip(ylim=c(-2.5,2.5)) +

  geom_point(data=p1Chol, aes(x=names, y=c(`Patient 1`))) +

  geom_point(data=p1Chol, aes(x=names, y=c(`Patient 1__1`))) +

  geom_point(data=p1Chol, aes(x=names, y=c(`Patient 1__2`)))


p2 <- ggplot(p2Chol, aes(x=p2Chol$names, y=`Patient 2`)) + geom_bar(colour="black",fill="blue",stat="identity")  +

  xlab("Gene") +   ylab("Mean Log2 FC") +

  geom_bar(stat = 'identity', colour="black", aes(fill = `Patient 2`>0), position = 'dodge') +

  scale_fill_manual(values=c("red","blue"),guide = 'none') +

  ggtitle("Patient 2") + coord_flip(ylim=c(-2.5,2.5))


p3 <- ggplot(p3Chol, aes(x=p3Chol$names, y=`Mean`)) + geom_bar(colour="black",fill="blue",stat="identity") +

  xlab("Gene") +   ylab("Mean Log2 FC") +

  geom_bar(stat = 'identity', colour="black", aes(fill = `Mean`>0), position = 'dodge') +

  scale_fill_manual(values=c("red","blue"),guide = 'none') +

  ggtitle("Patient 3") + coord_flip(ylim=c(-2.5,2.5)) +

  geom_point(data=p3Chol, aes(x=p3Chol$names, y=c(`Patient 3`))) +

  geom_point(data=p3Chol, aes(x=p3Chol$names, y=c(`Patient 3__1`))) +

  geom_point(data=p3Chol, aes(x=p3Chol$names, y=c(`Patient 3__2`)))


p4 <- ggplot(p4Chol, aes(x=p4Chol$names, y=`Patient 4`)) + geom_bar(colour="black",fill="blue",stat="identity") +

  xlab("Gene") +   ylab("Mean Log2 FC") +

  geom_bar(stat = 'identity', colour="black", aes(fill = `Patient 4`>0), position = 'dodge') +

  scale_fill_manual(values=c("red","blue"),guide = 'none') +

  ggtitle("Patient 4") + coord_flip(ylim=c(-2.5,2.5))


p5 <- ggplot(p5Chol, aes(x=p5Chol$names, y=`Patient 5`)) + geom_bar(colour="black",fill="blue",stat="identity") +

  xlab("Gene") +   ylab("Mean Log2 FC") +

  geom_bar(stat = 'identity', colour="black", aes(fill = `Patient 5`>0), position = 'dodge') +

  scale_fill_manual(values=c("red","blue"),guide = 'none') +

  ggtitle("Patient 5") + coord_flip(ylim=c(-2.5,2.5))


grid.arrange(p1, p2, p3, p4, p5, ncol = 3, top=textGrob("Cholesterol synthesis genes Log2 fold-change \n CNS:Spleeen by individual patient sample", gp=gpar(fontsize=15, font =2)))


svg(file="C:/Fluidigm individual patient cholesterol genes Barchart1.svg")
```

```
grid.arrange(p1, p2, p3, p4, p5, ncol = 3, top=textGrob("Cholesterol synthesis genes Log2 fold-change \n CNS:Spleeen by individual patient sample", gp=gpar(fontsize=15, font =2)))

dev.off()
```

```
############################################################
#Other gene data from multiplex PCR - idividual gene graphs

library(readr)
library(readxl)
library("ggplot2", lib.loc="C:/R/win-library/3.3")
library('RColorBrewer')
library(gridExtra)
library(grid)
library(stringr)
dev.off()

FluiChols <- read_excel("C:/TC Jan 17 Fluidigm chip analysis Chol-lipid export(analysis6 Other) Foldchanges.xlsx")
FluiChols <- as.data.frame(FluiChols)

rownames(FluiChols) <- FluiChols$Gene
FluiChols <- subset(FluiChols[,2:10])
FluiChols<- log(FluiChols,2)

FluiChols$`Patient 1 (t(12;21))` <- rowMeans(FluiChols[,1:3])
FluiChols$`Patient 2 (t(12;21))` <- FluiChols$`Patient 2`
FluiChols$`Patient 2`<- NULL
FluiChols$`Patient 3 (t(12;21))` <- rowMeans(FluiChols[,4:6])
FluiChols$`Patient 4 (t(4;11))` <- FluiChols$`Patient 4`
FluiChols$`Patient 4`<- NULL
FluiChols$`Patient 5 (t(4;11))` <- FluiChols$`Patient 5`
FluiChols$`Patient 5`<- NULL

p1Chol <- subset(FluiChols[,c(1:3)])
p1Chol <- as.data.frame(t(p1Chol))
p3Chol <- subset(FluiChols[,c(4:6)])
p3Chol <- as.data.frame(t(p3Chol))

FluiChols$`Patient 1` <- NULL
FluiChols$`Patient 1__1` <- NULL
FluiChols$`Patient 1__2` <- NULL
FluiChols$`Patient 3` <- NULL
FluiChols$`Patient 3__1` <- NULL
FluiChols$`Patient 3__2` <- NULL

tFluichols <- as.data.frame(t(FluiChols))

g1Chol <-tFluichols[c(-2,-3,-4,-5,-6,-7,-8)]
g1Chol$'1' <- c(p1Chol[1,1],NA,NA,NA,NA)
g1Chol$'1__1' <- c(p1Chol[2,1],NA,NA,NA,NA)
g1Chol$'1__2' <- c(p1Chol[3,1],NA,NA,NA,NA)
```

```
g1Chol$'3' <- c(NA,NA,p3Chol[1,1],NA,NA)
g1Chol$'3__1' <- c(NA,NA,p3Chol[2,1],NA,NA)
g1Chol$'3__2' <- c(NA,NA,p3Chol[3,1],NA,NA)

g1Chol$names <- factor(row.names(g1Chol), levels = row.names(g1Chol))

g2Chol <-tFluichols[c(-1,-3,-4,-5,-6,-7,-8)]
g2Chol$'1' <- c(p1Chol[1,2],NA,NA,NA,NA)
g2Chol$'1__1' <- c(p1Chol[2,2],NA,NA,NA,NA)
g2Chol$'1__2' <- c(p1Chol[3,2],NA,NA,NA,NA)

g2Chol$'3' <- c(NA,NA,p3Chol[1,2],NA,NA)
g2Chol$'3__1' <- c(NA,NA,p3Chol[2,2],NA,NA)
g2Chol$'3__2' <- c(NA,NA,p3Chol[3,2],NA,NA)

g2Chol$names <- factor(row.names(g2Chol), levels = row.names(g2Chol))

g3Chol <-tFluichols[c(-1,-2,-4,-5,-6,-7,-8)]
g3Chol$'1' <- c(p1Chol[1,3],NA,NA,NA,NA)
g3Chol$'1__1' <- c(p1Chol[2,3],NA,NA,NA,NA)
g3Chol$'1__2' <- c(p1Chol[3,3],NA,NA,NA,NA)

g3Chol$'3' <- c(NA,NA,p3Chol[1,3],NA,NA)
g3Chol$'3__1' <- c(NA,NA,p3Chol[2,3],NA,NA)
g3Chol$'3__2' <- c(NA,NA,p3Chol[3,3],NA,NA)

g3Chol$names <- factor(row.names(g3Chol), levels = row.names(g3Chol))

g4Chol <-tFluichols[c(-1,-2,-3,-5,-6,-7,-8)]
g4Chol$'1' <- c(p1Chol[1,4],NA,NA,NA,NA)
g4Chol$'1__1' <- c(p1Chol[2,4],NA,NA,NA,NA)
g4Chol$'1__2' <- c(p1Chol[3,4],NA,NA,NA,NA)

g4Chol$'3' <- c(NA,NA,p3Chol[1,4],NA,NA)
g4Chol$'3__1' <- c(NA,NA,p3Chol[2,4],NA,NA)
g4Chol$'3__2' <- c(NA,NA,p3Chol[3,4],NA,NA)

g4Chol$names <- factor(row.names(g4Chol), levels = row.names(g4Chol))

g5Chol <-tFluichols[c(-1,-2,-3,-4,-5,-6,-8)]
g5Chol$'1' <- c(p1Chol[1,7],NA,NA,NA,NA)
g5Chol$'1__1' <- c(p1Chol[2,7],NA,NA,NA,NA)
g5Chol$'1__2' <- c(p1Chol[3,7],NA,NA,NA,NA)

g5Chol$'3' <- c(NA,NA,p3Chol[1,7],NA,NA)
g5Chol$'3__1' <- c(NA,NA,p3Chol[2,7],NA,NA)
g5Chol$'3__2' <- c(NA,NA,p3Chol[3,7],NA,NA)
```

```
g5Chol$names <- factor(row.names(g5Chol), levels = row.names(g5Chol))

g6Chol <-tFluichols[c(-1,-2,-3,-4,-5,-6,-7)]
g6Chol$'1' <- c(p1Chol[1,8],NA,NA,NA,NA)
g6Chol$'1__1' <- c(p1Chol[2,8],NA,NA,NA,NA)
g6Chol$'1__2' <- c(p1Chol[3,8],NA,NA,NA,NA)

g6Chol$'3' <- c(NA,NA,p3Chol[1,8],NA,NA)
g6Chol$'3__1' <- c(NA,NA,p3Chol[2,8],NA,NA)
g6Chol$'3__2' <- c(NA,NA,p3Chol[3,8],NA,NA)

g6Chol$names <- factor(row.names(g6Chol), levels = row.names(g6Chol))

########################

g1 <- ggplot(g1Chol, aes(x=names, y=VEGFa)) +
  xlab("") +   ylab("Mean Log2 FC")  +
  geom_bar(stat = 'identity', colour="black", aes(fill = `VEGFa`>0), position = 'dodge') +
  scale_fill_manual(values=c("blue","blue"),guide = 'none') +
  ggtitle("VEGFa")  +
  scale_x_discrete(labels= function(names) str_wrap(names, width = 10)) +  #theme(plot.margin = 
unit(c(1,1,0.5,4),"lines")) +
  geom_point(data=na.omit(g1Chol), aes(x=g1Chol$names, y=g1Chol$`1`)) +
  geom_point(data=na.omit(g1Chol), aes(x=g1Chol$names, y=g1Chol$`1__1`)) +
  geom_point(data=na.omit(g1Chol), aes(x=g1Chol$names, y=g1Chol$`1__2`)) +
  geom_point(data=na.omit(g1Chol), aes(x=g1Chol$names, y=g1Chol$`3`)) +
  geom_point(data=na.omit(g1Chol), aes(x=g1Chol$names, y=g1Chol$`3__1`)) +
  geom_point(data=na.omit(g1Chol), aes(x=g1Chol$names, y=g1Chol$`3__2`)) +
  theme(axis.title.x = element_text(face="bold",size=16), axis.text.x = element_text(face="bold",size=9.5),
        axis.title.y = element_text(face="bold",size=16), axis.text.y = element_text(face="bold",size=14),
        plot.title=element_text(size=18,hjust=0.5,face="bold"))
g1

g2 <- ggplot(g2Chol, aes(x=names, y=IRF4)) +
  xlab("") +   ylab("Mean Log2 FC")  +
  geom_bar(stat = 'identity', colour="black", aes(fill = `IRF4`>0), position = 'dodge') +
  scale_fill_manual(values=c("blue","blue"),guide = 'none') +
  ggtitle("IRF4") +
  scale_x_discrete(labels= function(names) str_wrap(names, width = 10)) +  #theme(plot.margin = 
unit(c(1,1,0.5,4),"lines")) +
  geom_point(data=na.omit(g2Chol), aes(x=g2Chol$names, y=g2Chol$`1`)) +
  geom_point(data=na.omit(g2Chol), aes(x=g2Chol$names, y=g2Chol$`1__1`)) +
  geom_point(data=na.omit(g2Chol), aes(x=g2Chol$names, y=g2Chol$`1__2`)) +
  geom_point(data=na.omit(g2Chol), aes(x=g2Chol$names, y=g2Chol$`3`)) +
  geom_point(data=na.omit(g2Chol), aes(x=g2Chol$names, y=g2Chol$`3__1`)) +
  geom_point(data=na.omit(g2Chol), aes(x=g2Chol$names, y=g2Chol$`3__2`)) +
  theme(axis.title.x = element_text(face="bold",size=16), axis.text.x = element_text(face="bold",size=9.5),
```

```
      axis.title.y = element_text(face="bold",size=16), axis.text.y = element_text(face="bold",size=14),
      plot.title=element_text(size=18,hjust=0.5,face="bold"))

g2

g3 <- ggplot(g3Chol, aes(x=names, y=MERTK)) +
  xlab("") +   ylab("Mean Log2 FC")  +
  geom_bar(stat = 'identity', colour="black", aes(fill = `MERTK`>0), position = 'dodge') +
  scale_fill_manual(values=c("red","blue"),guide = 'none') +
  ggtitle("MERTK")
  scale_x_discrete(labels= function(names) str_wrap(names, width = 10)) +  #theme(plot.margin =
unit(c(1,1,0.5,4),"lines")) +
  geom_point(data=na.omit(g3Chol), aes(x=g3Chol$names, y=g3Chol$`1`)) +
  geom_point(data=na.omit(g3Chol), aes(x=g3Chol$names, y=g3Chol$`1__1`)) +
  geom_point(data=na.omit(g3Chol), aes(x=g3Chol$names, y=g3Chol$`1__2`)) +
  geom_point(data=na.omit(g3Chol), aes(x=g3Chol$names, y=g3Chol$`3`)) +
  geom_point(data=na.omit(g3Chol), aes(x=g3Chol$names, y=g3Chol$`3__1`)) +
  geom_point(data=na.omit(g3Chol), aes(x=g3Chol$names, y=g3Chol$`3__2`)) +
  theme(axis.title.x = element_text(face="bold",size=16), axis.text.x = element_text(face="bold",size=9.5),
      axis.title.y = element_text(face="bold",size=16), axis.text.y = element_text(face="bold",size=14),
      plot.title=element_text(size=18,hjust=0.5,face="bold"))
g3
g4 <- ggplot(g4Chol, aes(x=names, y=IL15)) +
  xlab("") +   ylab("Mean Log2 FC")  +
  geom_bar(stat = 'identity', colour="black", aes(fill = `IL15`>0), position = 'dodge') +
  scale_fill_manual(values=c("red","blue"),guide = 'none') +
  ggtitle("IL15")
  scale_x_discrete(labels= function(names) str_wrap(names, width = 10)) +  #theme(plot.margin =
unit(c(1,1,0.5,4),"lines")) +
  geom_point(data=na.omit(g4Chol), aes(x=g4Chol$names, y=g4Chol$`1`)) +
  geom_point(data=na.omit(g4Chol), aes(x=g4Chol$names, y=g4Chol$`1__1`)) +
  geom_point(data=na.omit(g4Chol), aes(x=g4Chol$names, y=g4Chol$`1__2`)) +
  geom_point(data=na.omit(g4Chol), aes(x=g4Chol$names, y=g4Chol$`3`)) +
  geom_point(data=na.omit(g4Chol), aes(x=g4Chol$names, y=g4Chol$`3__1`)) +
  geom_point(data=na.omit(g4Chol), aes(x=g4Chol$names, y=g4Chol$`3__2`)) +
  theme(axis.title.x = element_text(face="bold",size=16), axis.text.x = element_text(face="bold",size=9.5),
      axis.title.y = element_text(face="bold",size=16), axis.text.y = element_text(face="bold",size=14),
      plot.title=element_text(size=18,hjust=0.5,face="bold"))
g4
g5 <- ggplot(g5Chol, aes(x=names, y=ICAM1)) +
  xlab("") +   ylab("Mean Log2 FC")  +
  geom_bar(stat = 'identity', colour="black", aes(fill = `ICAM1`>0), position = 'dodge') +
  scale_fill_manual(values=c("red","blue"),guide = 'none') +
  ggtitle("ICAM1") +# coord_flip(ylim=c(-2.5,2.5)) +
  scale_x_discrete(labels= function(names) str_wrap(names, width = 10)) +  #theme(plot.margin =
unit(c(1,1,0.5,4),"lines")) +
  geom_point(data=na.omit(g5Chol), aes(x=g5Chol$names, y=g5Chol$`1`)) +
  geom_point(data=na.omit(g5Chol), aes(x=g5Chol$names, y=g5Chol$`1__1`)) +
  geom_point(data=na.omit(g5Chol), aes(x=g5Chol$names, y=g5Chol$`1__2`)) +
```

```
  geom_point(data=na.omit(g5Chol), aes(x=g5Chol$names, y=g5Chol$`3`)) +
  geom_point(data=na.omit(g5Chol), aes(x=g5Chol$names, y=g5Chol$`3__1`)) +
  geom_point(data=na.omit(g5Chol), aes(x=g5Chol$names, y=g5Chol$`3__2`)) +
  theme(axis.title.x = element_text(face="bold",size=16), axis.text.x = element_text(face="bold",size=9.5),
        axis.title.y = element_text(face="bold",size=16), axis.text.y = element_text(face="bold",size=14),
        plot.title=element_text(size=18,hjust=0.5,face="bold"))

g5
g6 <- ggplot(g6Chol, aes(x=names, y=SPP1)) +
  xlab("") +   ylab("Mean Log2 FC")  +
  geom_bar(stat = 'identity', colour="black", aes(fill = `SPP1`>0), position = 'dodge') +
  ##theme_bw()
  #scale_fill_manual(values="blue",guide = 'none') +
  scale_fill_manual(values=c("red","blue"),guide = 'none') +
  ggtitle("SPP1") +# coord_flip(ylim=c(-2.5,2.5)) +
  scale_x_discrete(labels= function(names) str_wrap(names, width = 10)) +  #theme(plot.margin = 
unit(c(1,1,0.5,4),"lines")) +
  geom_point(data=na.omit(g6Chol), aes(x=g6Chol$names, y=g6Chol$`1`)) +
  geom_point(data=na.omit(g6Chol), aes(x=g6Chol$names, y=g6Chol$`1__1`)) +
  geom_point(data=na.omit(g6Chol), aes(x=g6Chol$names, y=g6Chol$`1__2`)) +
  geom_point(data=na.omit(g6Chol), aes(x=g6Chol$names, y=g6Chol$`3`)) +
  geom_point(data=na.omit(g6Chol), aes(x=g6Chol$names, y=g6Chol$`3__1`)) +
  geom_point(data=na.omit(g6Chol), aes(x=g6Chol$names, y=g6Chol$`3__2`)) +
  theme(axis.title.x = element_text(face="bold",size=16), axis.text.x = element_text(face="bold",size=9.5),
        axis.title.y = element_text(face="bold",size=16), axis.text.y = element_text(face="bold",size=14),
        plot.title=element_text(size=18,hjust=0.5,face="bold"))

g6

grid.arrange(g1, g2, g3, g4, g5, g6, ncol = 3, top=textGrob("Log2 Fold-change in expression CNS:Spleen\nof genes 
implicated in CNS ALL", gp=gpar(fontsize=20, font =2)))

dev.off()
svg(file="C:/Fluidigm individual patient Other genes Barchart1.svg")
grid.arrange(g1, g2, g3, g4, g5, g6, ncol = 2, top=textGrob("Log2 Fold-change in expression CNS:Spleen\nof genes 
implicated in CNS ALL", gp=gpar(fontsize=20, font =2)))
dev.off()
```

```
##################################################################
#Primary CNS ALL (relapse) analysis
library(readr)
library("ggplot2", lib.loc="C:/R/win-library/3.3")
library('RColorBrewer')
library(readxl)
library(gplots)
library(car)

dev.off()
Chol <- read_csv("C:/VanDongenCNSrelapsedata.csv")

Chol <- as.data.frame(Chol)
Chol$`-Log10 p-value` <- -log10(Chol$`p-value`)

Chol$Gene <- factor(Chol$Gene, levels = Chol$Gene[order(Chol$log2FoldChange)])

bccol<- colorRampPalette(brewer.pal(9, 'GnBu'))(100)

ggplot(Chol, aes(x=Gene, y=log2FoldChange, fill=`-Log10 p-value`)) + geom_bar(colour="black",stat="identity") +
  xlab("Gene") +   ylab("Log2 Fold-change") +
  ggtitle("Cholesterol synthesis genes") + coord_flip() +
  scale_fill_gradientn(colours=topo.colors(7),
                    breaks=c(1.3, 30, 60),
                    labels=c("1.3 (Adj. p=0.05)","5","10 (Adj. p=1x10e-10)"), limits=c(1.3,10), na.value = "white")

ggplot(Chol, aes(x=Gene, y=log2FoldChange, fill=`-Log10 p-value`)) + geom_bar(colour="black",stat="identity") +
  xlab("Gene") +   ylab("Log2 Fold-change, CNS:Spleen") +
  ggtitle("Cholesterol synthesis genes") + coord_flip() +
  scale_fill_gradientn(colours=bccol,
                    breaks=c(1.3, 30, 60),
                    labels=c("1.3 (Adj. p=0.05)","5","10"), limits=c(1.3,10),na.value = "transparent")
dev.off()
svg(file="C:/VD Cholesterol genes barchart.svg")
ggplot(Chol, aes(x=Gene, y=log2FoldChange, fill=`-Log10 p-value`)) + geom_bar(colour="black",stat="identity") +
  xlab("Gene") +   ylab("Log2 Fold-change, CNS:Spleen") +
  ggtitle("Cholesterol synthesis genes") + coord_flip() +
  scale_fill_gradientn(colours=bccol,
                    breaks=c(1.3, 30, 60),
                    labels=c("1.3 (Adj. p=0.05)","5","10"), limits=c(1.3,10),na.value = "transparent")
dev.off()

Chol2 <- read_excel("C:/Cluster.xlsx")
Chol2 <- as.data.frame(Chol2)

row.names(Chol2) <- Chol2$X__1
Chol2$X__1 <- NULL
```

```
set.seed(20)
VDCluster <- kmeans(Chol2, 2, nstart = 20)
VDCluster
table(VDCluster$cluster, rownames(Chol2))

clusters <- hclust(dist(Chol2))
plot(clusters)
scaledChol <- scale(Chol2)

heatmap.2(scaledChol, scale="none", trace = "none",
        rowsep = 8, sepcolor = "black",
        dendrogram = "row",cexRow=1,
        col=bccol,
        key=FALSE,lhei=c(1,50))

heatmap.2(t(scaledChol),scale="none", trace = "none",
        colsep = 42, sepcolor = "black",
        dendrogram = "col",cexCol=1,
        col=bccol,
        key=FALSE,lwid=c(1,50),revC=TRUE,lhei=c(1,1))
dev.off()
svg(file="C:/VD Cholesterol genes heatmap.svg")
heatmap.2(t(scaledChol),scale="none", trace = "none",
        colsep = 42, sepcolor = "black",
        dendrogram = "col",cexCol=1,
        col=bccol,
        #col=cm.colors(256),
        #sepwidth = 1,
        key=FALSE,lwid=c(1,50),revC=TRUE,lhei=c(1,1))
dev.off()
```

```
#############################################################
#TARGET histogram analysis

.libPaths("C:/R/win-library/3.3")
library(readxl)
CoGchol <- read_excel("C:/CoG9906 chol genes prior to ln txxlsx.xlsx")

CoGchol$Gene <- NULL
CoGchol <- as.matrix(CoGchol)
dev.off()
hist(CoGchol, freq=FALSE, col="lightgreen", main="Distribution of\ncholesterol genes in CoG P9906\ntrial samples prior to log transformation", xlab ="")
curve(dnorm(x, mean=mean(CoGchol), sd=sd(CoGchol)), add=TRUE, lwd=2)
lCoGChol <- log(CoGchol)
hist(lCoGChol, freq=FALSE, col="lightgreen",main="Distribution of\ncholesterol genes in CoG P9906\ntrial samples after to log transformation", xlab ="")
curve(dnorm(x, mean=mean(lCoGChol), sd=sd(lCoGChol)), add=TRUE, lwd=2)
dev.off()

svg(file="C:/Z-scores after ln transformation/histogram prior to ln tx.svg")
hist(CoGchol, freq=FALSE, col="lightgreen", main="Distribution of\ncholesterol genes in CoG P9906\ntrial samples prior to log transformation", xlab ="")
curve(dnorm(x, mean=mean(CoGchol), sd=sd(CoGchol)), add=TRUE, lwd=2)
dev.off()

svg(file="C:/Z-scores after ln transformation/histogram post ln tx.svg")
hist(lCoGChol, freq=FALSE, col="lightgreen",main="Distribution of\ncholesterol genes in CoG P9906\ntrial samples after to log transformation", xlab ="")
curve(dnorm(x, mean=mean(lCoGChol), sd=sd(lCoGChol)), add=TRUE, lwd=2)
dev.off()
```

```
###############################################################
#TARGET survival analysis
.libPaths("C:/R/win-library/3.3")
library(readxl)
library(rms)
## Get data
CoGdata <- read_excel("C:/Cog9906analysis3 with ln-transform for multivariate analysis.xlsx")
##Remove bottom row (empty)
CoGdata<-CoGdata[1:207,]
##renames "2+ genes zs>1.5" column
colnames(CoGdata)[which(names(CoGdata)=="2+ genes upregulated z-s >=1.5")] <- "Cholesterol_synthesis_upregulated"
## Fit Cox prortional hazard model
CoxModelCoG <- coxph(Surv(CoGdata$`Event Free Survival Time in Day`, CoGdata$`Isolated CNS relapse?`)
            ~ CoGdata$`Cholesterol_synthesis_upregulated` + CoGdata$`Day 29 MRD >0.01` + CoGdata$`High WCC?`,
            data=CoGdata)
summary(CoxModelCoG)

##### export Cox hazard as table
# Prepare the columns
HR <- round(exp(coef(CoxModelCoG)), 2)
CI <- round(exp(confint(CoxModelCoG)), 2)
P <- round(coef(summary(CoxModelCoG))[,5], 5)
# Names the columns of CI
colnames(CI) <- c("Lower", "Higher")
# Bind columns together as dataset
table2 <- as.data.frame(cbind(HR, CI, P))
table2
write.table(table2, file="C:/coxtable.csv", sep=",")
#######
## Prepare Kaplan-Meier curve
CoGkm <- npsurv(Surv(`Event Free Survival Time in Day`,`Isolated CNS relapse?`) ~ Cholesterol_synthesis_upregulated,
data=CoGdata)
summary(CoGkm)
survplot(CoGkm,xlab="Days", ylab="CNS Relapse-free Survival Probablility",
      conf = "none", xlim = c(0,1460), label.curves = list(keys=c("chol. synthesis not upregulated","chol. synthesis
upregulated"),keyloc=("none")),
      col=c("black","blue"),lty=c(2,1), lwd=c(1,2), time.inc=365, n.risk=TRUE, adj.n.risk=0,
cex.n.risk=0.5,fun=function(y)100*y, ylim=c(0,100))
dev.off()
embed_fonts("C:/Documents/Experiments/In silico analyses/ALL/CoGP9906/Z-scores after ln transformation/R-
plots/KMcurve;zs1.5;2gene.pdf", outfile="C:/Documents/Experiments/In silico analyses/ALL/CoGP9906/Z-scores after ln
transformation/R-plots/KMcurve;zs1.5;2gene_e.pdf")
par(font.axis="1")
par(mar=c(5, 4, 4, 2)+0.1)
svg(file="C:/Documents/Experiments/In silico analyses/ALL/CoGP9906/Z-scores after ln transformation/R-
plots/KMcurve;zs1.5;2gene.svg", family="Arial Black")
par(font.axis="9")
par(mar=c(5, 4, 4, 10)+0.1)
survplot(CoGkm,xlab="Days", ylab="CNS Relapse-free Probability (%)",
```

```
conf = "none", xlim = c(0,1460), label.curves = list(keys=c("chol. synthesis not upregulated","chol. synthesis upregulated"),keyloc=("none")),

col=c("black","blue"),lty=c(2,1), lwd=c(1,2), time.inc=365, n.risk=TRUE, adj.n.risk=0, cex.n.risk=0.6,fun=function(y)100*y, ylim=c(0,100))

dev.off()

par(font.axis="1")

par(mar=c(5, 4, 4, 2)+0.1)

########
```

```
##############################################################
#RNASeq IL7R analysis
library(readr)
library("ggplot2", lib.loc="C:/R/win-library/3.3")
library('RColorBrewer')
dev.off()
IL7R <- read_excel("C:/RNASeq IL7R data.xlsx")
IL7R <- as.data.frame(IL7R)
IL7R$`-Log10 Adjusted p-value` <- -log10(IL7R$`padj`)

IL7R$Gene <- factor(IL7R$Gene, levels = IL7R$Gene[order(IL7R$log2FoldChange)])

bccol<- colorRampPalette(brewer.pal(9, 'GnBu'))(100)
ggplot(IL7R, aes(x=Gene, y=log2FoldChange, fill=`-Log10 Adjusted p-value`)) + geom_bar(colour="black",stat="identity") +
  xlab("Gene") +   ylab("Log2 Fold-change") +
  #  ylim(-0.5,1) +
  ggtitle("IL7Ra expression") +# coord_flip() +
  scale_fill_gradientn(colours=topo.colors(7),
                       breaks=c(1.3, 2, 3),
                       labels=c("1.3 (Adj. p=0.05)","2","3"), limits=c(1.3,3), na.value = "transparent")+
  theme_bw()
ggplot(IL7R, aes(x=Gene, y=log2FoldChange, fill=`-Log10 Adjusted p-value`)) + geom_bar(colour="black",stat="identity") +
  xlab("Gene") +   ylab("Log2 Fold-change, CNS:Spleen") +
  ggtitle("IL7Ra expression") +# coord_flip() +
  scale_fill_gradientn(colours=bccol,
                       breaks=c(1.3, 2, 3),
                       labels=c("1.3 (Adj. p=0.05)","2","3"), limits=c(1.3,3),na.value = "transparent")
dev.off()
dev.off()
svg(file="C:/IL7R gene Barchart1.svg")
ggplot(IL7R, aes(x=Gene, y=log2FoldChange, fill=`-Log10 Adjusted p-value`)) + geom_bar(colour="black",stat="identity") +
  xlab("Gene") +   ylab("Log2 Fold-change, CNS:Spleen") +
  ggtitle("IL7Ra expression") +# coord_flip() +
  scale_fill_gradientn(colours=bccol,
                       breaks=c(1.3, 2, 3),
                       labels=c("1.3 (Adj. p=0.05)","2","3"), limits=c(1.3,3),na.value = "transparent") +
  theme(axis.title=element_text(size=16),axis.text=element_text(size=12),
        plot.title=element_text(size=18,hjust=0.5,face="bold"))
dev.off()
```

```
###########################################################
#VanDongen IL7R analysis
library(readxl)

## Get data
Van_Dongen_data_IL7Ra_for_r_analysis <- read_excel("C:/Van Dongen data IL7Ra for r analysis.xlsx")

boxplot(Van_Dongen_data_IL7Ra_for_r_analysis$`Mean across probes` ~ Van_Dongen_data_IL7Ra_for_r_analysis$Site)

boxplot(Van_Dongen_data_IL7Ra_for_r_analysis$`226218_at` ~ Van_Dongen_data_IL7Ra_for_r_analysis$Site)

boxplot(Van_Dongen_data_IL7Ra_for_r_analysis$`205798_at` ~ Van_Dongen_data_IL7Ra_for_r_analysis$Site)

library(ggplot2)

p <- ggplot(Van_Dongen_data_IL7Ra_for_r_analysis, aes(x=Site, y=`Mean across probes`)) +
  labs(list(title = "IL7R expression", x = "Source of ALL cells", y = "gene expression")) +
  theme(axis.title.x = element_text(face="bold",size=16), axis.text.x = element_text(face="bold",size=14),
        axis.title.y = element_text(face="bold",size=16), axis.text.y = element_text(face="bold",size=14),
        plot.title=element_text(size=18,hjust=0.5,face="bold"))

## define custom median function
plot.median <- function(x) {
  m <- median(x)
  c(y = m, ymin = m, ymax = m)
}

p2 <- p + geom_boxplot(aes(ymin=..lower.., ymax=..upper..),fill=c("red","red","blue"),colour="black") +
  geom_dotplot(binaxis='y', stackdir='center', method="histodot", binwidth=0.15)
p2 + theme_bw() + theme(axis.title.x = element_text(face="bold",size=16), axis.text.x =
element_text(face="bold",size=14),
                axis.title.y = element_text(face="bold",size=16), axis.text.y = element_text(face="bold",size=14),
                plot.title=element_text(size=18,hjust=0.5,face="bold"))
########
library(grid)
# Create a text
grob <- grobTree(textGrob("        p-value (Mann-Whitney)\nBone Marrow (diagnosis or relapse)\n        vs CNS relapse
<0.00001",
                 x=0.54,  y=0.14, hjust=0,
                 gp=gpar(col="black", fontsize=14, fontface="bold")))
# Plot
p2  + theme_bw() + theme(axis.title.x = element_text(face="bold",size=16), axis.text.x =
element_text(face="bold",size=14),
                 axis.title.y = element_text(face="bold",size=16), axis.text.y = element_text(face="bold",size=14),
                 plot.title=element_text(size=18,hjust=0.5,face="bold")) +
  annotation_custom(grob)

VDDMWtest <- read_excel("C:/Documents/Experiments/In silico analyses/ALL/IL17Ra analyses/Van Dongen data IL7Ra for
r analysis (M-W test).xlsx")
```

```
wilcox.test(IL7Ra ~ Site, data=VDDMWtest)

VDDBMD <- read_excel("C:/Documents/Experiments/In silico analyses/ALL/IL17Ra analyses/Van Dongen data IL7Ra for r analysis (M-W test, BMD).xlsx")

wilcox.test(IL7Ra ~ Site, data=VDDBMD)

VDDBMR <- read_excel("C:/Documents/Experiments/In silico analyses/ALL/IL17Ra analyses/Van Dongen data IL7Ra for r analysis (M-W test, BMR).xlsx")

wilcox.test(IL7Ra ~ Site, data=VDDBMR)


###################Save to file

dev.off()

svg(file="C:/Van Dongen data boxplot (IL7Ra)1.svg")

p2  + theme_bw() + theme(axis.title.x = element_text(face="bold",size=16), axis.text.x = element_text(face="bold",size=14),

                axis.title.y = element_text(face="bold",size=16), axis.text.y = element_text(face="bold",size=14),

                plot.title=element_text(size=18,hjust=0.5,face="bold")) +

  annotation_custom(grob)


dev.off()
```

```
###########################################################
#TARGET IL7R analysis
library(readxl)
library(rms)
library(tidyr)
library(extrafont)
library(extrafontdb)

## Get data
CoGdata <- read_excel("C:/Documents/Experiments/In silico analyses/ALL/CoGP9906/Minitab-SPSS/Cog9906analysis3(2) with ln-transform for spss multivariate1.xlsx")

##Remove bottom row (empty)
CoGdata<-CoGdata[1:207,]

##renames "2+ genes zs>1.5" column
colnames(CoGdata)[which(names(CoGdata)=="2+ genes upregulated z-s >=1.5")] <- "Cholesterol_synthesis_upregulated"

## Fit Cox prortional hazard model
CoxModelCoG <- coxph(Surv(CoGdata$`Event Free Survival Time in Day`, CoGdata$`Isolated CNS relapse?`)
            ~ CoGdata$`IL7R z-score >=1.2` + CoGdata$`Cholesterol_synthesis_upregulated` + CoGdata$`Day 29 MRD >0.01` + CoGdata$`High WCC?` + CoGdata$`CNS status 3` + CoGdata$`High Age at diagnosis?` + CoGdata$`MLL Status_1`,
            data=CoGdata)
summary(CoxModelCoG)

##### export Cox hazard as table
########## https://datascienceplus.com/how-to-export-regression-results-from-r-to-ms-word/ #####
# Prepare the columns

HR <- round(exp(coef(CoxModelCoG)), 2)
CI <- round(exp(confint(CoxModelCoG)), 2)
p <- round(coef(summary(CoxModelCoG))[,5], 5)
# Names the columns of CI
colnames(CI) <- c("Lower", "Higher")
# Bind columns together as dataset
table2 <- as.data.frame(cbind(HR, CI, p))
table2$a <- "("
table2$b <- "-"
table2$c <- ")"

table2 <- table2[,c("HR","a","Lower","b","Higher","c", "p")]
# Merge all columns in one
table2 = unite(table2, "HR (95% CI)", c(HR, a, Lower, b, Higher, c), sep = ",", remove=T)
# add space between the estimates of HR and CI
table2[,1] <- gsub("\\(", " (", table2[,1])
table2
write.table(table2, file="C:/coxtable-IL7R;zs1-2 with cholesterol.csv", sep=",")
```

```
####################################################
## Prepare Kaplan-Meier curve
CoGdata$`IL7R.Upregulated` <-  CoGdata$`IL7R z-score >=1.2`
CoGkm <- npsurv(Surv(`Event Free Survival Time in Day`,`Isolated CNS relapse?`) ~ `IL7R.Upregulated`, data=CoGdata)
summary(CoGkm)
survdiff(Surv(`Event Free Survival Time in Day`,`Isolated CNS relapse?`) ~ `IL7R.Upregulated`, data=CoGdata)
par(mar=c(5.1,7.1,4.1,8.1))
survplot(CoGkm,xlab="Days", ylab="iCNS Relapse-free\nProbablility", cex.xlab = 2, cex.ylab=2,
        conf = "none", xlim = c(0,1460), label.curves = list(keys=c("IL7R not upregulated","IL7R upregulated"),keyloc=("none")),
        col=c("black","blue"),lty=c(2,1), lwd=c(1,2), time.inc=365, n.risk=TRUE, adj.n.risk=0, cex.n.risk=1,fun=function(y)100*y, ylim=c(0,100))
par(mar=c(5.1,4.1,4.1,2.1))

###############save curve as .svg
dev.off()
svg(file="C:/Documents/Experiments/In silico analyses/ALL/IL17Ra analyses/KMcurve;zs1.2;IL7R(1).svg")
par(font.axis="9")
par(mar=c(5.1,7.1,4.1,8.1))
survplot(CoGkm,xlab="Days", ylab="iCNS Relapse-free\nProbablility", cex.xlab = 2, cex.ylab=2,
        conf = "none", xlim = c(0,1460), label.curves = list(keys=c("IL7R not upregulated","IL7R upregulated"),keyloc=("none")),
        col=c("black","blue"),lty=c(2,1), lwd=c(1,2), time.inc=365, n.risk=TRUE, adj.n.risk=0, cex.n.risk=1,fun=function(y)100*y, ylim=c(0,100))
dev.off()
par(font.axis="1")
par(mar=c(5, 4, 4, 2)+0.1)
dev.off()
```

```
##############################################################
#Untargeted metabolomics volcano plots (similar for negative mode)


library(readxl)
library(calibrate)


TwoDVolcPos <- read_excel("C:/2D Volcano plot Pos.xlsx",col_types=c("text","numeric","numeric","numeric",
"numeric","numeric","text","numeric"))


# EvL
# https://www.r-bloggers.com/using-volcano-plots-in-r-to-visualize-microarray-and-rna-seq-results/


with(TwoDVolcPos, plot(EvLLog2FC, Log10pEvL, pch=20, cex=(0.00015*MeanAbundanceCapAt10000),
        main="Volcano Plot\n(Positive Mode) Early vs Late", xlim=c(-6,6),ylim=c(0,8),
        xlab="Log2 Fold-Change Early vs Late", ylab="-Log10 p-value Early vs Late",
        cex.lab=1.75, cex.axis=1.75, cex.main=1.75, cex.sub=1.75))


# Add colored points: red if padj<0.05, orange of log2FC>1, green if both, blue if named)
with(subset(TwoDVolcPos, Log10pEvL>1.3 ), points(EvLLog2FC, Log10pEvL, pch=20,
col="red",cex=(0.00015*MeanAbundanceCapAt10000)))
with(subset(TwoDVolcPos, abs(EvLLog2FC)>1), points(EvLLog2FC, Log10pEvL, pch=20,
col="orange",cex=(0.00015*MeanAbundanceCapAt10000)))
with(subset(TwoDVolcPos, Log10pEvL>1.3 & abs(EvLLog2FC)>1), points(EvLLog2FC, Log10pEvL, pch=20,
col="green",cex=(0.00015*MeanAbundanceCapAt10000)))
with(subset(TwoDVolcPos, Name!=""), points(EvLLog2FC, Log10pEvL, pch=20,
col="blue",cex=(0.00015*MeanAbundanceCapAt10000)))
with(subset(TwoDVolcPos, Name!=""), textxy(EvLLog2FC, Log10pEvL, labs=Name, cex=1.75, offset=0.6))


dev.off()
dev.off()
svg(file="C:/2D Volcano plot Pos EvL.svg")


with(TwoDVolcPos, plot(EvLLog2FC, Log10pEvL, pch=20, cex=(0.00015*MeanAbundanceCapAt10000),
        main="Volcano Plot\n(Positive Mode) Early vs Late", xlim=c(-6,6),ylim=c(0,8),
        xlab="Log2 Fold-Change Early vs Late", ylab="-Log10 p-value Early vs Late",
        cex.lab=1.75, cex.axis=1.75, cex.main=1.75, cex.sub=1.75))


# Add colored points: red if padj<0.05, orange of log2FC>1, green if both, blue if named)
with(subset(TwoDVolcPos, Log10pEvL>1.3 ), points(EvLLog2FC, Log10pEvL, pch=20,
col="red",cex=(0.00015*MeanAbundanceCapAt10000)))
with(subset(TwoDVolcPos, abs(EvLLog2FC)>1), points(EvLLog2FC, Log10pEvL, pch=20,
col="orange",cex=(0.00015*MeanAbundanceCapAt10000)))
with(subset(TwoDVolcPos, Log10pEvL>1.3 & abs(EvLLog2FC)>1), points(EvLLog2FC, Log10pEvL, pch=20,
col="green",cex=(0.00015*MeanAbundanceCapAt10000)))
with(subset(TwoDVolcPos, Name!=""), points(EvLLog2FC, Log10pEvL, pch=20,
col="blue",cex=(0.00015*MeanAbundanceCapAt10000)))


with(subset(TwoDVolcPos, Name!=""), textxy(EvLLog2FC, Log10pEvL, labs=Name, cex=1.75, offset=0.6))


dev.off()
dev.off()
```

```
####
#   EvU

with(TwoDVolcPos, plot(EvULog2FC, Log10pEvU, pch=20, cex=(0.00015*MeanAbundanceCapAt10000),
        main="Volcano Plot (Positive Mode)\nEarly vs Control", xlim=c(-6,6),ylim=c(0,8),
        xlab="Log2 Fold-Change Early vs Control", ylab="-Log10 p-value Early vs Control",
        cex.lab=1.75, cex.axis=1.75, cex.main=1.75, cex.sub=1.75))

# Add colored points: red if padj<0.05, orange of log2FC>1, green if both, blue if named)
with(subset(TwoDVolcPos, Log10pEvU>1.3 ), points(EvULog2FC, Log10pEvU, pch=20,
col="red",cex=(0.00015*MeanAbundanceCapAt10000)))
with(subset(TwoDVolcPos, abs(EvULog2FC)>1), points(EvULog2FC, Log10pEvU, pch=20,
col="orange",cex=(0.00015*MeanAbundanceCapAt10000)))
with(subset(TwoDVolcPos, Log10pEvU>1.3 & abs(EvULog2FC)>1), points(EvULog2FC, Log10pEvU, pch=20,
col="green",cex=(0.00015*MeanAbundanceCapAt10000)))
with(subset(TwoDVolcPos, Name!=""), points(EvULog2FC, Log10pEvU, pch=20,
col="blue",cex=(0.00015*MeanAbundanceCapAt10000)))

with(subset(TwoDVolcPos, Name!=""), textxy(EvULog2FC, Log10pEvU, labs=Name, cex=1.75, offset=0.6))



dev.off()
svg(file="C:/2D Volcano plot Pos EvU.svg")
with(TwoDVolcPos, plot(EvULog2FC, Log10pEvU, pch=20, cex=(0.00015*MeanAbundanceCapAt10000),
        main="Volcano Plot (Positive Mode)\nEarly vs Control", xlim=c(-6,6),ylim=c(0,8),
        xlab="Log2 Fold-Change Early vs Control", ylab="-Log10 p-value Early vs Control",
        cex.lab=1.75, cex.axis=1.75, cex.main=1.75, cex.sub=1.75))

# Add colored points: red if padj<0.05, orange of log2FC>1, green if both, blue if named)
with(subset(TwoDVolcPos, Log10pEvU>1.3 ), points(EvULog2FC, Log10pEvU, pch=20,
col="red",cex=(0.00015*MeanAbundanceCapAt10000)))
with(subset(TwoDVolcPos, abs(EvULog2FC)>1), points(EvULog2FC, Log10pEvU, pch=20,
col="orange",cex=(0.00015*MeanAbundanceCapAt10000)))
with(subset(TwoDVolcPos, Log10pEvU>1.3 & abs(EvULog2FC)>1), points(EvULog2FC, Log10pEvU, pch=20,
col="green",cex=(0.00015*MeanAbundanceCapAt10000)))
with(subset(TwoDVolcPos, Name!=""), points(EvULog2FC, Log10pEvU, pch=20,
col="blue",cex=(0.00015*MeanAbundanceCapAt10000)))
with(subset(TwoDVolcPos, Name!=""), textxy(EvULog2FC, Log10pEvU, labs=Name, cex=1.75, offset=0.6))
dev.off()
dev.off()
```

```
###############################################################
#Untargeted metabolomics allopurinol plots

#################http://www.sthda.com/english/wiki/correlation-test-between-two-variables-in-r##################
library("ggpubr")
library(readxl)

xva <- read_excel("C:/Xanthine vs allopurinol.xlsx")
xva1 <- xva
xva <- subset(xva,xva$Allopurinol >10000)
ggscatter(xva, x = "Xanthine", y = "Allopurinol",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Xanthine abundance", ylab = "Allopurinol abundance", cor.coef.size = 8) +
  theme(axis.title = element_text(face="bold",size=25), axis.text  = element_text(size=20))

pearsontest <- cor.test(xva$Xanthine, xva$Allopurinol,
                        method = "pearson")
pearsontest

dev.off()
dev.off()
svg(file="C:/Xanthine vs allopurinol.svg")
ggscatter(xva, x = "Xanthine", y = "Allopurinol",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Xanthine abundance", ylab = "Allopurinol abundance", cor.coef.size = 8) +
  theme(axis.title = element_text(face="bold",size=25), axis.text  = element_text(size=20))
dev.off()
dev.off()
svg(file="C:/Xanthine vs allopurinol(1).svg")

ggscatter(xva1, x = "Xanthine", y = "Allopurinol",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Xanthine abundance", ylab = "Allopurinol abundance", cor.coef.size = 8) +
  theme(axis.title = element_text(face="bold",size=25), axis.text  = element_text(size=20))
dev.off()
```

```
##############################################################
#Cholesterol abundance boxplots
library(readxl)
library(ggplot2)

CSFChol1 <- read_excel("C:/CSF free cholesterol for r chart.xlsx")
CSFChol1 <-na.omit(CSFChol1)

CSFChol2 <- subset(CSFChol1[(56:58),])
CSFChol1 <- subset(CSFChol1[(1:55),])

r <- ggplot(CSFChol1, aes(Group,`TotalChol(ug)`,fill=Colour)) + geom_boxplot(show.legend = FALSE) +
scale_fill_manual(values=c("blue","red")) +
  geom_point(data=CSFChol1, aes(x=Group,y=`TotalChol(ug)`),show.legend = FALSE) +
  xlab("") + ylab("Total Cholesterol in CSF(ug)") + ggtitle("Total Cholesterol Abundance in CSF From\nChildren at Diagnosis
With Leukaemia Compared\nWith Matched Leukaemic Controls on Maintenance\nChemotherapy and Unmatched
Controls") +
  #  ylim(0.5,0.95) +
  theme(legend.position="none") +
  theme_bw() +
  theme(axis.title.x = element_text(face="bold",size=16), axis.text.x = element_text(face="bold",size=14),
      axis.title.y = element_text(face="bold",size=16), axis.text.y = element_text(face="bold",size=14),
      plot.title=element_text(size=18,hjust=0.5,face="bold"))
r
############
dev.off()
dev.off()

svg(file="C:/CSFCholAbundanceBoxplot.svg")
r
dev.off()
dev.off()
```

```
#############################################################
#Cellular holesterol abundance boxplots
library(readxl)
library(ggplot2)
library(grid)
library(gridExtra)

## Get data

CellChol1 <- read_excel("C:/CholesterolGCMSforRe.xlsx")
CellChol2 <- subset(CellChol1[(19:20),])
CellChol1 <- subset(CellChol1[(1:18),])

r <- ggplot(CellChol2, aes(Group,`Cholesterol(ug/10(6)cells)`)) +geom_bar(stat="identity", fill=c("red","blue")) +
  geom_point(data=CellChol1, aes(x=Group,y=`Cholesterol(ug/10(6)cells)`)) +
  xlab("") + ylab("Total Cholesterol (ug/10(6) cells)") +# ggtitle("Total Cholesterol Abundance in Cells\nFrom CNS and Spleen of MiceWith ALL") +
  #  theme(axis.title.y = element_text(face="bold"), axis.text.y = element_text(face="bold"))+
  theme_bw() +
  theme(axis.title.x = element_text(face="bold",size=16), axis.text.x = element_text(face="bold",size=14),
        axis.title.y = element_text(face="bold",size=16), axis.text.y = element_text(face="bold",size=14),
        plot.title=element_text(size=18,hjust=0.5,face="bold"))
r
title1=textGrob("Total Cholesterol Abundance in Cells\nfrom the CNS and Spleen of\nmice with ALL", gp=gpar(fontface="bold",fontsize=20))
grid.arrange(r,ncol=1, top=title1)
dev.off()
dev.off()
svg(file="C:/CholesterolGCMSbar.svg")

grid.arrange(r,ncol=1, top=title1)
dev.off()
dev.off()
```

```
##############################################################
#Cellular fatty acid abundance boxplots (similar for other multi-boxplot graphs)
library(readxl)
library(ggplot2)
library(grid)
library(gridExtra)

## Get data

CNSSPLCellsFA <- read_excel("C:MevalonateFAsRBoxplot.xlsx")
#Mevalonate

p <- ggplot(CNSSPLCellsFA, aes(x=Site, y=`Mevalonate`)) +
  labs(list(title = "", x = "", y = "Mevalonate Abundance")) +
  theme(axis.title.x = element_text(face="bold"), axis.text.x = element_text(face="bold")) +
  theme(axis.title.y = element_text(face="bold"), axis.text.y = element_text(face="bold"))

## dotplot with box
p2 <- p + geom_boxplot(aes(ymin=..lower.., ymax=..upper..),fill=c("red","blue"),colour="black") +
  geom_dotplot(binaxis='y', stackdir='center', method="histodot") +#, binwidth=20)
  theme_bw()+
  theme(axis.title.x = element_text(face="bold",size=16), axis.text.x = element_text(face="bold",size=14),
        axis.title.y = element_text(face="bold",size=16), axis.text.y = element_text(face="bold",size=14),
        plot.title=element_text(size=18,hjust=0.5,face="bold"))



title1=textGrob("Mevalonate abundance in cells from\nthe CNS and Spleens of mice\nwith ALL",
gp=gpar(fontface="bold",fontsize=20))
grid.arrange(p2,ncol=1, top=title1)
#save to file
dev.off()
dev.off()
svg(file="C:/MevalonatBoxplot.svg")
grid.arrange(p2,ncol=1, top=title1)
dev.off()
dev.off()
################################
#Palmitate / Palmitoleate

d <- ggplot(CNSSPLCellsFA, aes(x=Site, y=`Palmitate`)) +
  labs(list(title = "Palmitate", x = "", y = "Palmitate Abundance")) +
  theme(axis.title.x = element_text(face="bold"), axis.text.x = element_text(face="bold")) +
  theme(axis.title.y = element_text(face="bold"), axis.text.y = element_text(face="bold"))

## dotplot with box
d2 <- d + geom_boxplot(aes(ymin=..lower.., ymax=..upper..),fill=c("red","blue"),colour="black") +
  geom_dotplot(binaxis='y', stackdir='center', method="histodot") +#, binwidth=20)
  theme_bw() +
```

```
  theme(axis.title.x = element_text(face="bold",size=16), axis.text.x = element_text(face="bold",size=14),
        axis.title.y = element_text(face="bold",size=16), axis.text.y = element_text(face="bold",size=14),
        plot.title=element_text(size=18,hjust=0.5,face="bold"))

e <- ggplot(CNSSPLCellsFA, aes(x=Site, y=`Palmitoleate`)) +
  labs(list(title = "Palmitoleate", x = "", y = "Palmitoleate Abundance")) +
  theme(axis.title.x = element_text(face="bold"), axis.text.x = element_text(face="bold")) +
  theme(axis.title.y = element_text(face="bold"), axis.text.y = element_text(face="bold"))

## dotplot with box
e2 <- e + geom_boxplot(aes(ymin=..lower.., ymax=..upper..),fill=c("red","blue"),colour="black") +
  geom_dotplot(binaxis='y', stackdir='center', method="histodot") +
  ylim(500000,10000000) +
  theme_bw()+
  theme(axis.title.x = element_text(face="bold",size=16), axis.text.x = element_text(face="bold",size=14),
        axis.title.y = element_text(face="bold",size=16), axis.text.y = element_text(face="bold",size=14),
        plot.title=element_text(size=18,hjust=0.5,face="bold"))

title1=textGrob("Palmitate and Palmitoleate abundance in cells\n from the CNS and Spleens of\nmice with ALL",
gp=gpar(fontface="bold",fontsize=20))
grid.arrange(d2,e2,ncol=2, top=title1)

#save to file
dev.off()
dev.off()

svg(file="C:/PalmitatepalmitoleateBoxplot.svg")
grid.arrange(d2,e2,ncol=2, top=title1)

dev.off()
dev.off()
##############

###############################
#Stearate / Oleate

k <- ggplot(CNSSPLCellsFA, aes(x=Site, y=`Stearate`)) +
  labs(list(title = "Stearate", x = "", y = "Stearate Abundance")) +
  theme(axis.title.x = element_text(face="bold"), axis.text.x = element_text(face="bold")) +
  theme(axis.title.y = element_text(face="bold"), axis.text.y = element_text(face="bold"))

## dotplot with box
k2 <- k + geom_boxplot(aes(ymin=..lower.., ymax=..upper..),fill=c("red","blue"),colour="black") +
  geom_dotplot(binaxis='y', stackdir='center', method="histodot") +#, binwidth=20)
  theme_bw()+
  theme(axis.title.x = element_text(face="bold",size=16), axis.text.x = element_text(face="bold",size=14),
        axis.title.y = element_text(face="bold",size=16), axis.text.y = element_text(face="bold",size=14),
        plot.title=element_text(size=18,hjust=0.5,face="bold"))
```

```
l <- ggplot(CNSSPLCellsFA, aes(x=Site, y=`Oleate`)) +
  labs(list(title = "Oleate", x = "", y = "Oleate Abundance")) +
  theme(axis.title.x = element_text(face="bold"), axis.text.x = element_text(face="bold")) +
  theme(axis.title.y = element_text(face="bold"), axis.text.y = element_text(face="bold"))

## dotplot with box
l2 <- l + geom_boxplot(aes(ymin=..lower.., ymax=..upper..),fill=c("red","blue"),colour="black") +
  geom_dotplot(binaxis='y', stackdir='center', method="histodot") +#, binwidth=20)
  theme_bw()+
  theme(axis.title.x = element_text(face="bold",size=16), axis.text.x = element_text(face="bold",size=14),
        axis.title.y = element_text(face="bold",size=16), axis.text.y = element_text(face="bold",size=14),
        plot.title=element_text(size=18,hjust=0.5,face="bold"))

title1=textGrob("Stearate and Oleate abundance in cells\nfrom the CNS and Spleens of\nmice with ALL",
gp=gpar(fontface="bold",fontsize=20))

grid.arrange(k2,l2,ncol=2, top=title1)

#save to file
dev.off()
dev.off()

svg(file="C:/StearateoleateBoxplot.svg")
grid.arrange(k2,l2,ncol=2, top=title1)

dev.off()
dev.off()
############################################
#Palmitate:Palmitoleate

r <- ggplot(CNSSPLCellsFA, aes(x=Site, y=`Palmitate:Palmitoleate`)) +
  labs(list(title = "Palmitate/Palmitoleate", x = "", y = "Palmitate:Palmitoleate Ratio")) +
  theme(axis.title.x = element_text(face="bold"), axis.text.x = element_text(face="bold")) +
  theme(axis.title.y = element_text(face="bold"), axis.text.y = element_text(face="bold"))

## dotplot with box
r2 <- r + geom_boxplot(aes(ymin=..lower.., ymax=..upper..),fill=c("red","blue"),colour="black") +
  geom_dotplot(binaxis='y', stackdir='center', method="histodot") +#, binwidth=20)
  theme_bw()+
  theme(axis.title.x = element_text(face="bold",size=16), axis.text.x = element_text(face="bold",size=14),
        axis.title.y = element_text(face="bold",size=16), axis.text.y = element_text(face="bold",size=14),
        plot.title=element_text(size=18,hjust=0.5,face="bold"))

s <- ggplot(CNSSPLCellsFA, aes(x=Site, y=`Stearate:Oleate`)) +
  labs(list(title = "Stearate/Oleate", x = "", y = "Stearate:Oleate Ratio")) +
  theme(axis.title.x = element_text(face="bold"), axis.text.x = element_text(face="bold")) +
  theme(axis.title.y = element_text(face="bold"), axis.text.y = element_text(face="bold"))
```

```
## dotplot with box
s2 <- s + geom_boxplot(aes(ymin=..lower.., ymax=..upper..),fill=c("red","blue"),colour="black") +
  geom_dotplot(binaxis='y', stackdir='center', method="histodot") +#, binwidth=20)
  theme_bw()+
  theme(axis.title.x = element_text(face="bold",size=16), axis.text.x = element_text(face="bold",size=14),
        axis.title.y = element_text(face="bold",size=16), axis.text.y = element_text(face="bold",size=14),
        plot.title=element_text(size=18,hjust=0.5,face="bold"))

title1=textGrob("Palmitate:Palmitoleate and Stearate:Oleate ratios\n in Cells from the CNS and Spleens\nof mice with ALL", gp=gpar(fontface="bold",fontsize=20))
grid.arrange(r2,s2,ncol=2, top=title1)

#save to file
dev.off()
dev.off()

svg(file="C:/RatiopalmitatetopalmitoleateBoxplot.svg")
grid.arrange(r2,s2,ncol=2, top=title1)


dev.off()
dev.off()

####################################################################
### CSF Cholesterol/CE analysis

##Chol ester subsets

CE1 <- read_excel("C:/CSF_nmoles_norm for r stacked bar chart2.xlsx")

#CE2 <- subset(CE1[,1:7])

m <- ggplot(CE1, aes(Sample,Value)) + geom_bar(stat="identity",aes(fill = ChE)) +
  xlab("Sample") + ylab("Proportion of total Cholesterol Esters") +
  ggtitle("Bar Chart Showing Cholesterol Ester\nSubtypes in CSF From Mice\nWith and Without Leukaemia") +
  scale_fill_brewer(palette = "Set1") +
  theme_bw() +
  theme(axis.title.x = element_text(face="bold",size=16), axis.text.x = element_text(face="bold",size=14),
        axis.title.y = element_text(face="bold",size=16), axis.text.y = element_text(face="bold",size=14),
        plot.title=element_text(size=18,hjust=0.5,face="bold"))

m

dev.off()
dev.off()
svg(file="C:/DCSFChESubsets.svg")
m
```

```
dev.off()
dev.off()
```

```
##################################################################
### CSF metabolite timeline linecharts

.libPaths("C:/R/win-library/3.3")
library(readxl)
library(ggplot2)
library(gridExtra)
library(grid)


############## Creatine

cr <- read_excel("C:/creatine data for chart csf at relapse.xlsx")
cr1 <- subset(cr[,c(1,2)])
cr1 <- na.omit(cr1)

cra <- ggplot(cr1, aes(x=Days, y=P1437)) +  xlim(2,414) +ylim(1,3.2) +
  geom_line(size=1,colour="#762a83") + geom_point(size=1.5,colour="#762a83") + ylab("") + xlab("") +
  labs(title="P1437") +
  geom_vline(xintercept = 330, size=1) + annotate("text", x=335,y=55, hjust=0,,label="time of relapse") +
  theme(plot.title = element_text(size=18,hjust = 0.5,margin = margin(t = 10, b = -20))) +
  theme_bw()

cra
cr2 <- subset(cr[,c(1,3)])
cr2 <- na.omit(cr2)

crb <- ggplot(cr2, aes(x=Days, y=P6539), main="P6539")+ xlim(31,718) + ylim(1,3.2) +
  geom_line(size=1, colour="#5ab4ac") + geom_point(size=2, colour="#5ab4ac") + ylab("") + xlab("")+
  labs(title="P6539") +
  geom_vline(xintercept = 520, size=1) + annotate("text", x=530,y=55, hjust=0,,label="time of relapse") +
  theme(plot.title = element_text(size=18,hjust = 0.5,margin = margin(t = 10, b = -20)))+
  theme_bw()
crb
cr3 <- subset(cr[,c(1,4)])
cr3 <- na.omit(cr3)

crc <- ggplot(cr3, aes(x=Days, y=P4234)) + xlim(2050,3550) +ylim(1,3.2) +
  geom_line(size=1, colour="#01665e") + geom_point(size=2, colour="#01665e") + ylab("") + xlab("") +
  labs(title="P4234") +
  geom_vline(xintercept = 2567, size=1) + annotate("text", x=2600,y=55, hjust=0,,label="time of relapse") +
  geom_vline(xintercept = 3240, size=1) +
  theme(plot.title = element_text(size=18,hjust = 0.5,margin = margin(t = 10, b = -20)))+
  theme_bw()
crc

cr4 <- subset(cr[,c(1,5)])
```

```
cr4 <- na.omit(cr4)



crd <- ggplot(cr4, aes(x=Days, y=P6941)) + xlim(350,1150) +ylim(1,3.2) +
  geom_line(size=1, colour="#8c510a") + geom_point(size=2, colour="#8c510a") + ylab("") + xlab("") +
  labs(title="P6941") +
  geom_vline(xintercept = 1030, size=1) + annotate("text", x=1020,y=55, hjust=1,,label="time of relapse") +
  theme(plot.title = element_text(size=18,hjust = 0.5,margin = margin(t = 10, b = -20)))+
  theme_bw()
crd


grid.arrange(cra,crb,crc,crd,ncol=2, left=textGrob("uM creatine", rot=90, gp = gpar(fontface = "bold", fontsize=18)),
bottom=textGrob("Days"), top=textGrob("Changes in csf creatine abundance\nin serial CSF samples\nwith times of relapse
highlighted",gp = gpar(fontface = "bold", fontsize=22)))


####Save to file
dev.off()
svg(file="C:/Creatinelinechart.svg",width=10,
    height=7)
grid.arrange(cra,crb,crc,crd,ncol=2, left=textGrob("uM creatine", rot=90, gp = gpar(fontface = "bold", fontsize=18)),
bottom=textGrob("Days"), top=textGrob("Changes in csf creatine abundance\nin serial CSF samples\nwith times of relapse
highlighted",gp = gpar(fontface = "bold", fontsize=22)))
dev.off()
dev.off()


######### Xanthine
xa <- read_excel("C:/xanthine data for chart csf at relapse(normalised)1.xlsx")


xa1 <- subset(xa[,c(1,2)])
xa1 <- na.omit(xa1)


xaa <- ggplot(xa1, aes(x=Days, y=P1437)) +  xlim(2,414) +ylim(4,24) +
  geom_line(size=1,colour="#762a83") + geom_point(size=1.5,colour="#762a83") + ylab("") + xlab("") +
  labs(title="P1437") +
  geom_vline(xintercept = 330, size=1) + annotate("text", x=335,y=20, hjust=0,,label="time of relapse") +
  theme(plot.title = element_text(size=18,hjust = 0.5,margin = margin(t = 10, b = -20))) +
  theme_bw()


xaa


xa2 <- subset(xa[,c(1,3)])
xa2 <- na.omit(xa2)


xab <- ggplot(xa2, aes(x=Days, y=P6539), main="P6539")+ xlim(31,718) + ylim(4,24) +
  geom_line(size=1, colour="#5ab4ac") + geom_point(size=2, colour="#5ab4ac") + ylab("") + xlab("")+
  labs(title="P6539") +
  geom_vline(xintercept = 520, size=1) + annotate("text", x=530,y=20, hjust=0,,label="time of relapse") +
  theme(plot.title = element_text(size=18,hjust = 0.5,margin = margin(t = 10, b = -20)))+
  theme_bw()
```

```
xab

xa3 <- subset(xa[,c(1,4)])
xa3 <- na.omit(xa3)

xac <- ggplot(xa3, aes(x=Days, y=P4234)) + xlim(2050,3550) +ylim(4,24) +
  geom_line(size=1, colour="#01665e") + geom_point(size=2, colour="#01665e") + ylab("") + xlab("") +
  labs(title="P4234") +
  geom_vline(xintercept = 2567, size=1) + annotate("text", x=2600,y=20, hjust=0,,label="time of relapse") +
  geom_vline(xintercept = 3240, size=1) +
  theme(plot.title = element_text(size=18,hjust = 0.5,margin = margin(t = 10, b = -20)))+
  theme_bw()
xac

xa4 <- subset(xa[,c(1,5)])
xa4 <- na.omit(xa4)

xad <- ggplot(xa4, aes(x=Days, y=P6941)) + xlim(350,1150) +ylim(4,24) +
  geom_line(size=1, colour="#8c510a") + geom_point(size=2, colour="#8c510a") + ylab("") + xlab("") +
  labs(title="P6941") +
  geom_vline(xintercept = 1030, size=1) + annotate("text", x=1020,y=20, hjust=1,,label="time of relapse") +
  theme(plot.title = element_text(size=18,hjust = 0.5,margin = margin(t = 10, b = -20)))+
  theme_bw()
xad

grid.arrange(xaa,xab,xac,xad,ncol=2, left=textGrob("Abundance of Xanthine", rot=90, gp = gpar(fontface = "bold", fontsize=18)), bottom=textGrob("Days"), top=textGrob("Changes in csf xanthine abundance\nin serial CSF samples\nwith times of relapse highlighted",gp = gpar(fontface = "bold", fontsize=22)))

####Save to file
dev.off()
svg(file="C:/Xanthinelinechart.svg",width=10,
    height=7)
grid.arrange(xaa,xab,xac,xad,ncol=2, left=textGrob("Abundance of Xanthine", rot=90, gp = gpar(fontface = "bold", fontsize=18)), bottom=textGrob("Days"), top=textGrob("Changes in csf xanthine abundance\nin serial CSF samples\nwith times of relapse highlighted",gp = gpar(fontface = "bold", fontsize=22)))
dev.off()

dev.off()
###############

############### Dimethylarginine

di <- read_excel("C:/Dimethylarginine data for chart csf at relapse3.xlsx")
#di <- as.data.frame(di)

di1 <- subset(di[,c(1,2)])
di1 <- na.omit(di1)
```

```
dia <- ggplot(di1, aes(x=Days, y=P1437)) +  xlim(2,414) +ylim(8,32) +
  geom_line(size=1,colour="#762a83") + geom_point(size=1.5,colour="#762a83") + ylab("") + xlab("") +
  labs(title="P1437") +
  geom_vline(xintercept = 330, size=1) + annotate("text", x=335,y=28, hjust=0,,label="time of relapse") +
  theme(plot.title = element_text(size=18,hjust = 0.5,margin = margin(t = 10, b = -20))) +
  theme_bw()

dia

di2 <- subset(di[,c(1,3)])
di2 <- na.omit(di2)

dib <- ggplot(di2, aes(x=Days, y=P6539), main="P6539")+ xlim(31,718) + ylim(8,32) +
  geom_line(size=1, colour="#5ab4ac") + geom_point(size=2, colour="#5ab4ac") + ylab("") + xlab("")+
  labs(title="P6539") +
  geom_vline(xintercept = 520, size=1) + annotate("text", x=530,y=28, hjust=0,,label="time of relapse") +
  theme(plot.title = element_text(size=18,hjust = 0.5,margin = margin(t = 10, b = -20)))+
  theme_bw()
dib

di3 <- subset(di[,c(1,4)])
di3 <- na.omit(di3)


dic <- ggplot(di3, aes(x=Days, y=P4234)) + xlim(2050,3550) +ylim(8,32) +
  geom_line(size=1, colour="#01665e") + geom_point(size=2, colour="#01665e") + ylab("") + xlab("") +
  labs(title="P4234") +
  geom_vline(xintercept = 2567, size=1) + annotate("text", x=2600,y=28, hjust=0,,label="time of relapse") +
  geom_vline(xintercept = 3240, size=1) +
  theme(plot.title = element_text(size=18,hjust = 0.5,margin = margin(t = 10, b = -20)))+
  theme_bw()
dic

di4 <- subset(di[,c(1,5)])
di4 <- na.omit(di4)


did <- ggplot(di4, aes(x=Days, y=P6941)) + xlim(350,1150) +ylim(8,32) +
  geom_line(size=1, colour="#8c510a") + geom_point(size=2, colour="#8c510a") + ylab("") + xlab("") +
  labs(title="P6941") +
  geom_vline(xintercept = 1030, size=1) + annotate("text", x=1020,y=28, hjust=1,,label="time of relapse") +
  theme(plot.title = element_text(size=18,hjust = 0.5,margin = margin(t = 10, b = -20)))+
  theme_bw()
did
```

```
grid.arrange(dia,dib,dic,did,ncol=2, left=textGrob("uM dimethylarginine", rot=90, gp = gpar(fontface = "bold", fontsize=18)), bottom=textGrob("Days"), top=textGrob("Changes in csf dimethylarginine abundance\nin serial CSF samples\nwith times of relapse highlighted",gp = gpar(fontface = "bold", fontsize=22)))


####Save to file
dev.off()
svg(file="C:/dimethylargininelinechart.svg",width=10,
   height=7)
grid.arrange(dia,dib,dic,did,ncol=2, left=textGrob("Abundance of Dimethylarginine", rot=90, gp = gpar(fontface = "bold", fontsize=18)), bottom=textGrob("Days"), top=textGrob("Changes in csf dimethylarginine abundance\nin serial CSF samples\nwith times of relapse highlighted",gp = gpar(fontface = "bold", fontsize=22)))


##################
dev.off()
dev.off()

############### Phenylalanine

Ph <- read_excel("C:/Phenylalanine data for chart csf at relapse.xlsx")
#Ph <- as.data.frame(Ph)

Ph1 <- subset(Ph[,c(1,2)])
Ph1 <- na.omit(Ph1)



Pha <- ggplot(Ph1, aes(x=Days, y=P1437)) +  xlim(2,414) +ylim(16,62) +
  geom_line(size=1,colour="#762a83") + geom_point(size=1.5,colour="#762a83") + ylab("") + xlab("") +
  labs(title="P1437") +
  geom_vline(xintercept = 330, size=1) + annotate("text", x=335,y=52, hjust=0,,label="time of relapse") +
  theme(plot.title = element_text(size=18,hjust = 0.5,margin = margin(t = 10, b = -20))) +
  theme_bw()

Pha

Ph2 <- subset(Ph[,c(1,3)])
Ph2 <- na.omit(Ph2)

Phb <- ggplot(Ph2, aes(x=Days, y=P6539), main="P6539")+ xlim(31,718) + ylim(16,62) +
  geom_line(size=1, colour="#5ab4ac") + geom_point(size=2, colour="#5ab4ac") + ylab("") + xlab("")+
  labs(title="P6539") +
  geom_vline(xintercept = 520, size=1) + annotate("text", x=530,y=52, hjust=0,,label="time of relapse") +
  theme(plot.title = element_text(size=18,hjust = 0.5,margin = margin(t = 10, b = -20)))+
  theme_bw()
Phb

Ph3 <- subset(Ph[,c(1,4)])
Ph3 <- na.omit(Ph3)
```

```
Phc <- ggplot(Ph3, aes(x=Days, y=P4234)) + xlim(2050,3550) +ylim(16,62) +
  geom_line(size=1, colour="#01665e") + geom_point(size=2, colour="#01665e") + ylab("") + xlab("") +
  labs(title="P4234") +
  geom_vline(xintercept = 2567, size=1) + annotate("text", x=2600,y=52, hjust=0,,label="time of relapse") +
  geom_vline(xintercept = 3240, size=1) +
  theme(plot.title = element_text(size=18,hjust = 0.5,margin = margin(t = 10, b = -20)))+
  theme_bw()
Phc

Ph4 <- subset(Ph[,c(1,5)])
Ph4 <- na.omit(Ph4)


Phd <- ggplot(Ph4, aes(x=Days, y=P6941)) + xlim(350,1150) +ylim(16,62) +
  geom_line(size=1, colour="#8c510a") + geom_point(size=2, colour="#8c510a") + ylab("") + xlab("") +
  labs(title="P6941") +
  geom_vline(xintercept = 1030, size=1) + annotate("text", x=1020,y=52, hjust=1,,label="time of relapse") +
  theme(plot.title = element_text(size=18,hjust = 0.5,margin = margin(t = 10, b = -20)))+
  theme_bw()
Phd

grid.arrange(Pha,Phb,Phc,Phd,ncol=2, left=textGrob("uM Phenylalanine", rot=90, gp = gpar(fontface = "bold",
fontsize=18)), bottom=textGrob("Days"), top=textGrob("Changes in csf Phenylalanine abundance\nin serial CSF
samples\nwith times of relapse highlighted",gp = gpar(fontface = "bold", fontsize=22)))

####Save to file
dev.off()
svg(file="C:/Phenylalaninelinechart.svg",width=10,
    height=7)
grid.arrange(Pha,Phb,Phc,Phd,ncol=2, left=textGrob("Abundance of Phenylalanine", rot=90, gp = gpar(fontface = "bold",
fontsize=18)), bottom=textGrob("Days"), top=textGrob("Changes in csf phenylalanine abundance\nin serial CSF
samples\nwith times of relapse highlighted",gp = gpar(fontface = "bold", fontsize=22)))
dev.off()
dev.off()

############### Pseudouridine

Ps <- read_excel("C:/Pseudouridine data for chart csf at relapse.xlsx")
#Ps <- as.data.frame(Ps)

Ps1 <- subset(Ps[,c(1,2)])
Ps1 <- na.omit(Ps1)


Psa <- ggplot(Ps1, aes(x=Days, y=P1437)) +  xlim(2,414) +ylim(9,42) +
  geom_line(size=1,colour="#762a83") + geom_point(size=1.5,colour="#762a83") + ylab("") + xlab("") +
  labs(title="P1437") +
  geom_vline(xintercept = 330, size=1) + annotate("text", x=335,y=35, hjust=0,,label="time of relapse") +
  theme(plot.title = element_text(size=18,hjust = 0.5,margin = margin(t = 10, b = -20))) +
```

```
  theme_bw()

Psa

Ps2 <- subset(Ps[,c(1,3)])
Ps2 <- na.omit(Ps2)

Psb <- ggplot(Ps2, aes(x=Days, y=P6539), main="P6539")+ xlim(31,718) + ylim(9,42) +
  geom_line(size=1, colour="#5ab4ac") + geom_point(size=2, colour="#5ab4ac") + ylab("") + xlab("")+
  labs(title="P6539") +
  geom_vline(xintercept = 520, size=1) + annotate("text", x=530,y=35, hjust=0,,label="time of relapse") +
  theme(plot.title = element_text(size=18,hjust = 0.5,margin = margin(t = 10, b = -20)))+
  theme_bw()
Psb

Ps3 <- subset(Ps[,c(1,4)])
Ps3 <- na.omit(Ps3)


Psc <- ggplot(Ps3, aes(x=Days, y=P4234)) + xlim(2050,3550) +ylim(9,42) +
  geom_line(size=1, colour="#01665e") + geom_point(size=2, colour="#01665e") + ylab("") + xlab("") +
  labs(title="P4234") +
  geom_vline(xintercept = 2567, size=1) + annotate("text", x=2600,y=35, hjust=0,,label="time of relapse") +
  geom_vline(xintercept = 3240, size=1) +
  theme(plot.title = element_text(size=18,hjust = 0.5,margin = margin(t = 10, b = -20)))+
  theme_bw()
Psc

Ps4 <- subset(Ps[,c(1,5)])
Ps4 <- na.omit(Ps4)


Psd <- ggplot(Ps4, aes(x=Days, y=P6941)) + xlim(350,1150) +ylim(9,42) +
  geom_line(size=1, colour="#8c510a") + geom_point(size=2, colour="#8c510a") + ylab("") + xlab("") +
  labs(title="P6941") +
  geom_vline(xintercept = 1030, size=1) + annotate("text", x=1020,y=35, hjust=1,,label="time of relapse") +
  theme(plot.title = element_text(size=18,hjust = 0.5,margin = margin(t = 10, b = -20)))+
  theme_bw()
Psd

grid.arrange(Psa,Psb,Psc,Psd,ncol=2, left=textGrob("Abundance of Pseudouridine", rot=90, gp = gpar(fontface = "bold",
fontsize=18)), bottom=textGrob("Days"), top=textGrob("Changes in csf Pseudouridine abundance\nin serial CSF
samples\nwith times of relapse highlighted",gp = gpar(fontface = "bold", fontsize=22)))

####Save to file
dev.off()
svg(file="C:/Pseudouridinelinechart.svg",width=10,
    height=7)
```

```
grid.arrange(Psa,Psb,Psc,Psd,ncol=2, left=textGrob("uM creatine", rot=90, gp = gpar(fontface = "bold", fontsize=18)),
bottom=textGrob("Days"), top=textGrob("Changes in csf pseudouridine abundance\nin serial CSF samples\nwith times of
relapse highlighted",gp = gpar(fontface = "bold", fontsize=22)))

dev.off()
```

```
####################################################################
### CSF metabolite timeline linecharts with boxplot (palmitate-palioleate ratio); similir for other timelines + boxplots


.libPaths("C:/R/win-library/3.3")
library(readxl)
library(ggplot2)
library(gridExtra)
library(grid)
library(reshape2)


Palmitatecsfdata <- read_excel("C:/Documents/Experiments/Metabolomics data/Vaildation of untargeted data/Relapse CSF May 17 run/TC CSF relapse confirmation run/ReportOutput/Palmitate ratio data for chart csf at relapse1.xlsx")
Palmitatecsfdata2 <- read_excel("C:/Documents/Experiments/Metabolomics data/Vaildation of untargeted data/Relapse CSF May 17 run/TC CSF relapse confirmation run/ReportOutput/Palmitate ratio data for chart csf at relapse2.xlsx")
Palmitate1<- melt(Palmitatecsfdata, id.var='Days')
Palmitate1<- na.omit(Palmitate1)
Palmitate2<- melt(Palmitatecsfdata2, id.var='Days')
Palmitate2<- na.omit(Palmitate2)


ggplot() + geom_boxplot(data=Palmitate2, aes(x=Palmitate2$Days, y=Palmitate2$value, group=Palmitate2$Days))


boxplot(Palmitate2$value~Palmitate2$Days)


l<- ggplot(Palmitate1, aes(x=Days, y=value, color=variable)) +
  geom_boxplot(data=Palmitate2, aes(x=Palmitate2$Days, y=Palmitate2$value, group=Palmitate2$Days, fill="#d8b365"),width=0.5, show.legend=FALSE) +
  geom_line(size=1.4) +
  geom_point(size=3)+
  theme_bw() +
  theme(legend.title=element_blank(),axis.title = element_text(size=20),axis.text=element_text(size=16), legend.text=element_text(size=14)) +
  scale_colour_manual(values=c("#762a83",#"#762a83","#762a83","#762a83","#762a83",
                        "#5ab4ac",#"#5ab4ac","#5ab4ac","#5ab4ac","#5ab4ac","#5ab4ac",
                        "#01665e",#"#01665e","#01665e","#01665e","#01665e","#01665e",
                        "#8c510a")) +#,"#8c510a","#8c510a","#8c510a","#8c510a","#8c510a")) +
  scale_x_log10(breaks=c(0,8,28,183,365,730,1460,3400),labels=c("0","8","28","183","365","730","1460","3400")) +
  xlab(" ") + ylab("Palmitate:Palmitoloeate ratio")


title1=textGrob("Palmitate:Palmitoleate ratio in\n CSF of children with ALL", gp=gpar(fontface="bold", fontsize=20))
grid.arrange(l,ncol=1, top=title1,bottom=textGrob("Days",gp=gpar(fontface="bold",fontsize=16)))


dev.off()
svg(file="C:/Palmitate-palmitoloeate csfchart (logaxis)2.svg")
grid.arrange(l,ncol=1, top=title1,bottom=textGrob("Days",gp=gpar(fontface="bold",fontsize=16)))
dev.off()
```

```
##################################################################
### 13C glucose percentages

library(readxl)
library(ggplot2)
library(grid)
library(gridExtra)

## Get data

CSFgluc1 <- read_excel("C:/13C-glucose percentages for R.xlsx")

CSFgluc20 <- subset(CSFgluc1,Group=="20 mins")
CSFgluc202 <- subset(CSFgluc20[(9:10),])
CSFgluc201 <- subset(CSFgluc20[(1:8),])

r <- ggplot(CSFgluc202, aes(`CSF or Plasma`,`% of labelled glucose`)) +geom_bar(stat="identity", fill=c("red","blue")) +
  geom_point(data=CSFgluc201, aes(x=`CSF or Plasma`,y=`% of labelled glucose`)) +
  xlab("20 minutes") + ylab("Percentage of 13C-labelled glucose") +
  theme_bw() +
  theme(axis.title.x = element_text(face="bold",size=20), axis.text.x = element_text(face="bold",size=15),
        axis.title.y = element_text(face="bold",size=20), axis.text.y = element_text(face="bold",size=15))

r
################################
CSFgluc40 <- subset(CSFgluc1,Group=="40mins")
CSFgluc402 <- subset(CSFgluc40[(5:6),])
CSFgluc401 <- subset(CSFgluc40[(1:4),])

s <- ggplot(CSFgluc402, aes(`CSF or Plasma`,`% of labelled glucose`)) +geom_bar(stat="identity", fill=c("red","blue")) +
  geom_point(data=CSFgluc401, aes(x=`CSF or Plasma`,y=`% of labelled glucose`)) +
  xlab("40 minutes") + ylab("Percentage of 13C-labelled glucose") +
  theme_bw() +
  theme(axis.title.x = element_text(face="bold",size=20), axis.text.x = element_text(face="bold",size=15)) +
  theme(axis.title.y = element_text(face="bold",size=20), axis.text.y = element_text(face="bold",siz=15))
s

title1=textGrob("Percentage of 13C-labelled glucose in\nthe CSF and Plasma of mic with ALL\n20 and 40 minutes after injection", gp=gpar(fontface="bold",fontsize=20))
grid.arrange(r,s,ncol=2, top=title1)

###################################
dev.off()
dev.off()
svg(file="C:/13Cglucose percentages for R.svg")
grid.arrange(r,s,ncol=2, top=title1)
dev.off()
dev.off()
```