This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.
TET Mediated 5’Hydroxymethylation in the Pathogenesis of Non Alcoholic Fatty Liver Disease

Marcus J Lyall

Doctor of Philosophy
The University of Edinburgh
2017
Declaration
I declare that this thesis has been composed by me, and that all of the work is my own unless otherwise stated.

Dr Marcus J Lyall
August 2017
Acknowledgements

The work within this thesis has been made possible primarily though the support and encouragement of others prior to and during my academic studies and the patience and vision of collaborative groups.

Following my move to Edinburgh from Dundee in 2007 as a junior doctor I was fortunate enough to find myself in the metabolic unit at the western general hospital. Professor Mark Strachan encouraged me into the field of Diabetes and Endocrinology, a decision I have never regretted, and has been a clinical and personal mentor throughout my medical career. Unsure where my research interests would be, I was guided towards the Edinburgh Clinical Academic Track (ECAT) scheme by Professor Brian Walker. ECAT provided a remarkable opportunity to explore the phenomenal research opportunities within Edinburgh University without which I would not have been able to enjoy three such rewarding years in this intriguing field.

I am hugely indebted to the ECAT team, Professor Brian Walker, Professor John Iredale and Professor Neil Henderson and Joanne Ness for their solid guidance, pragmatic advice and also to Professor Andrew Jackson for the many impromptu supportive meetings on our mutual train commute home. Collectively they ensured my chosen research plan was relevant, achievable and had prospects for future exploration.

This project has spanned laboratories in three separate institutes proving a great logistical and technical challenge. This has been made possible by the trust and vision of Professor Richard Meehan and Dr David Hay in affording me time in their respective laboratories. To an equal extent I am hugely indebted to the students and post-doctoral staff within those labs who are too numerous to mention but you know who you are!! Particular thanks must go to Dr John Thomson. Not only a good friend and colleague but a remarkably talented scientist who has afforded me limitless time in teaching me the complex world of epigenome bioinformatics. Great thanks must also go to Dr Kate Cameron for teaching me the intricacies of embryonic stem cell differentiation and characterisation and the many times she fed cells in my absence! Finally, considerable thanks must go to the Drake lab members of past and present, in particular Dr Jessy Cartier, Dr Khulan Batbayar and Dr Catherine Rose who were given the unenviable task of turning me into a moderately competent scientist. Finally, I am of course greatly indebted to my primary supervisor Dr Mandy Drake, not only for her remarkable patience and encouragement but also her scientific expertise in guiding my research and challenging my experimental ideas.
Abstract

Non-alcoholic fatty liver disease (NAFLD) now affects around one in four adults in the human population and parallels the global increase in obesity. Within the spectrum of NAFLD, simple steatosis is associated with insulin resistance and type 2 diabetes while progression to steatohepatitis (NASH) is associated with an increased risk of liver cirrhosis and all-cause mortality. The molecular pathology of NAFLD is incompletely understood, however observational studies in human cohorts suggest the regulation of DNA methylation may play a role. 5-hydroxymethylcytosine (5hmC) is a cytosine modification generated from 5-methylcytosine (5mC) by the Ten eleven translocase isoenzymes (Tets) as part of a demethylation process. The aim of this project was to examine the role of Tet enzyme activity on the pathogenesis and progression of NAFLD.

Detailed characterisation of two established murine dietary interventions allowed the selection of a NAFLD mouse model which broadly recapitulated the metabolic, histological and transcriptional features of human disease. Using DNA immunoprecipitation coupled with whole genome next generation sequencing and RNA micro expression arrays I examined the effect of high fat diet feeding (HFD) on hepatic DNA 5hmC levels within annotated gene regions. Whilst the global 5hmC profile was not altered by HFD, there was profound genic enrichment of 5hmC in upregulated mediators of cholesterol synthesis and transport (Lss, Sc4mol, Fdps, Hsd17b7, Cyp17a1, Mvd, Cyp1a2, Dhcr7 and Apoa4) with no enrichment in genes with other pathological functions (drug detoxification, inflammation, cell cycle regulation). Induced peaks of 5hmC enrichment were subsequently abolished following rescue of the NAFLD phenotype by conversion to control diet.

Cross species validation was performed in vitro utilising embryonic stem cell derived hepatocytes challenged with a cocktail of high energy substrates. My in vivo findings were broadly replicated with specific 5hmC enrichment in genes synthesising lipotoxic molecules (PLIN2, CIDEC, APOA4, ACADVL, HMGCS2, APOA5, CYP2J2, IGFBP1, PPAP2C, ACSL1, APOC3, ANGPTL4, NRG1) with no enrichment in upregulated genes of alternative function.

To determine whether or not the 5hmC enrichment seen is of functional relevance, I studied Tet1<sup>−/−</sup> C57BL/6J mice. Tet1<sup>−/−</sup> mice are grossly normal in appearance, however loss of Tet1 conferred a striking resistance to diet induced obesity with reduced body fat mass, improved insulin-sensitivity and near complete absence of NAFLD compared to wild type littermates.
Furthermore, the HFD fed Tet1\(^{-}\) liver transcriptome showed a ‘protective’ profile, with suppression of genes for lipid synthesis, inflammation and fibrosis.

Thus, in multiple cross-species models of NAFLD, over nutrition induces genic hydroxymethylation specifically within activated genes driving the synthesis and transport of lipid molecules. Such changes are reversible with resolution of the NAFLD phenotype strengthening functional association. Tet1 deficiency conveys an obesity and NAFLD resistant phenotype. I therefore introduce Tet1 mediated hydroxymethylation as a novel mechanism for NAFLD pathogenesis.
Lay Summary

Around 25% of the world’s population have too much fat in their liver, a disease called ‘non-alcoholic fatty liver disease’ or NAFLD. NAFLD is predominantly a result of chronic overnutrition and the number of people with of NAFLD has increased rapidly in parallel with the rise in obesity and sedentary living. NAFLD is associated with type 2 diabetes, a greatly increased risk of heart disease and stroke and higher rates of common cancers. In addition to this, a proportion of patients with NAFLD go on to develop liver failure for which the only treatment is liver transplantation. Indeed, it is likely that NAFLD will be the dominant indication for liver transplant in the US and UK in the coming 20 years. Despite such high global rates, the exact mechanism by which fat accumulates in the liver and contributes to negative health outcomes is not yet defined. It is imperative therefore that we understand the NAFLD disease process such that new treatments can be found to limit the burden of disease.

The liver is a central organ in regulating energy balance. To do this, genes in the liver are activated or deactivated in response to signals from the gut and other organs. Understanding how genes are switched on and off is essential for our understanding of liver function during health and disease. One way in which this may be regulated is through chemical modifications to one of the DNA building blocks (known as bases) named cytosine. In mammalian DNA, a large proportion of cytosine bases are modified with a chemical tag known as a methyl group (referred to as 5mC) and regions of DNA where high levels of 5mC exist are more likely to be inactive. In 2009 it was discovered that 5mC can be converted into a chemically similar but functionally very different base called 5-hydroxymethylcytosine (5hmC). This process is performed by a group of proteins present in all tissues called Ten eleven translocase (TET) enzymes, which convert 5mC to 5hmC. In this thesis, I investigated this process of TET mediated 5mC to 5hmC conversion in NAFLD.

I demonstrate for the first time that in mice and human liver cells, the induction of NAFLD is associated with the generation of high levels of the 5hmC mark within genes that are activated, to drive the synthesis and transport of fats. Furthermore, I show that in mice this process appears to be reversible when NAFLD is reversed by calorie restriction. Finally, this study demonstrates that in mice lacking one of the TET enzymes (TET1) are protected from obesity when fed a high fat diet and do not develop NAFLD, perhaps secondary to a failure to activate a wide variety of genes involved in lipid synthesis and transport as well as a number of
processes associated with NAFLD complications. It is hoped that future research in this field will lead to more effective, personalised care for patients with this globally prevalent disease.
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^1$H-MRS</td>
<td>proton magnetic resonance spectroscopy</td>
</tr>
<tr>
<td>5caC</td>
<td>5-carboxyl-cytosine</td>
</tr>
<tr>
<td>5fC</td>
<td>5-formylcytosine</td>
</tr>
<tr>
<td>5hmC</td>
<td>5-hydroxymethylation</td>
</tr>
<tr>
<td>5mC</td>
<td>5-methylcytosine</td>
</tr>
<tr>
<td>Aacs</td>
<td>Acetoacetyl-CoA Synthetase</td>
</tr>
<tr>
<td>ACACA</td>
<td>Acetyl-CoA Carboxylase Alpha</td>
</tr>
<tr>
<td>Acacb</td>
<td>Acetyl-CoA Carboxylase Beta</td>
</tr>
<tr>
<td>ACADVL</td>
<td>Acyl-CoA Dehydrogenase, Very Long Chain</td>
</tr>
<tr>
<td>ACC1</td>
<td>Acetyl CoA carboxylase</td>
</tr>
<tr>
<td>Acly</td>
<td>ATP Citrate Lyase</td>
</tr>
<tr>
<td>Acox</td>
<td>Peroxisomal acyl-coenzyme A oxidase 1</td>
</tr>
<tr>
<td>Acsl1</td>
<td>Acyl-CoA Synthetase Long-Chain Family Member 1</td>
</tr>
<tr>
<td>Acss2</td>
<td>Acyl-CoA Synthetase Short-Chain Family Member 2</td>
</tr>
<tr>
<td>Adam32</td>
<td>ADAM Metallopeptidase Domain 32</td>
</tr>
<tr>
<td>AdipoR2</td>
<td>adiponectin receptor 2</td>
</tr>
<tr>
<td>Agpat2</td>
<td>1-Acylglycerol-3-Phosphate O-Acyltransferase 2</td>
</tr>
<tr>
<td>Ahcy</td>
<td>Adenosylhomocysteinase</td>
</tr>
<tr>
<td>AKR1B10</td>
<td>Aldo-Keto Reductase Family 1, Member B10 (Aldose Reductase)</td>
</tr>
<tr>
<td>AKR1B15</td>
<td>Aldo-Keto Reductase Family 1, Member B15</td>
</tr>
<tr>
<td>Akr1b7</td>
<td>aldo-keto reductase family 1, member B7</td>
</tr>
<tr>
<td>AKR1C2</td>
<td>Aldo-Keto Reductase Family 1, Member C2</td>
</tr>
<tr>
<td>AKR1C4</td>
<td>Aldo-Keto Reductase Family 1, Member C4</td>
</tr>
<tr>
<td>ALB</td>
<td>Albumin</td>
</tr>
<tr>
<td>AMPK</td>
<td>5'-AMP-activated protein kinase</td>
</tr>
<tr>
<td>ANGPTL4</td>
<td>Angiopoietin Like 4</td>
</tr>
<tr>
<td>ApoA4</td>
<td>Apolipoprotein A4</td>
</tr>
<tr>
<td>APOA5</td>
<td>Apolipoprotein A5</td>
</tr>
<tr>
<td>APOB100</td>
<td>apolipoproteinB-100 polypeptide chain</td>
</tr>
<tr>
<td>APOC3</td>
<td>Apolipoprotein C3</td>
</tr>
<tr>
<td>Arhgap5</td>
<td>Rho GTPase Activating Protein 5</td>
</tr>
<tr>
<td>Ccl4</td>
<td>C-C motif chemokine ligand 4</td>
</tr>
<tr>
<td>CD36</td>
<td>cluster of differentiation 36</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Cdc20</td>
<td>cell division cycle 20</td>
</tr>
<tr>
<td>CDD</td>
<td>choline deficient diet</td>
</tr>
<tr>
<td>Ces</td>
<td>Carboxylesterase</td>
</tr>
<tr>
<td>CGIs</td>
<td>CpG islands</td>
</tr>
<tr>
<td>ChIP</td>
<td>Chromatin Immunoprecipitation</td>
</tr>
<tr>
<td>Chkb</td>
<td>choline kinase beta</td>
</tr>
<tr>
<td>Chrebp</td>
<td>Carbohydrate-responsive element-binding protein</td>
</tr>
<tr>
<td>CIDEC</td>
<td>Cell Death Inducing DFFA Like Effector C</td>
</tr>
<tr>
<td>Coll1a1</td>
<td>collagen type I alpha 1 chain</td>
</tr>
<tr>
<td>CpG</td>
<td>cytosine-phosphate-guanine</td>
</tr>
<tr>
<td>Cxcl1</td>
<td>chemokine (C-X-C motif) ligand</td>
</tr>
<tr>
<td>Cxcl10</td>
<td>C-X-C Motif Chemokine Ligand 10</td>
</tr>
<tr>
<td>Cyp17a1</td>
<td>Cytochrome P450 Family 17 Subfamily A Member 1</td>
</tr>
<tr>
<td>CYP1A1</td>
<td>Cytochrome P450 Family 1 Subfamily A Member 1</td>
</tr>
<tr>
<td>CYP2J2</td>
<td>Cytochrome P450 Family 2 Subfamily J Member 2</td>
</tr>
<tr>
<td>DGAT1</td>
<td>Diacylglycerol acyltransferase1</td>
</tr>
<tr>
<td>DGAT2</td>
<td>Diacylglycerol acyltransferase2</td>
</tr>
<tr>
<td>Dhcr24</td>
<td>24-Dehydrocholesterol Reductase</td>
</tr>
<tr>
<td>Dhcr7</td>
<td>7-Dehydrocholesterol Reductase</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNL</td>
<td>de novo lipogenesis</td>
</tr>
<tr>
<td>DNMT1</td>
<td>DNA methyltransferases1</td>
</tr>
<tr>
<td>DNMT2</td>
<td>DNA methyltransferases 2</td>
</tr>
<tr>
<td>DNMT3a</td>
<td>DNA methyltransferases 3a</td>
</tr>
<tr>
<td>DNMT3b</td>
<td>DNA methyltransferases 3b</td>
</tr>
<tr>
<td>Elovl3</td>
<td>ELOVL Fatty Acid Elongase 3</td>
</tr>
<tr>
<td>Elovl6</td>
<td>ELOVL Fatty Acid Elongase 6</td>
</tr>
<tr>
<td>ER</td>
<td>endoplasmic reticulum</td>
</tr>
<tr>
<td>ESCheps</td>
<td>ESC derived hepatocytes</td>
</tr>
<tr>
<td>FA</td>
<td>fatty acids</td>
</tr>
<tr>
<td>Fasn</td>
<td>Fatty Acid Synthase</td>
</tr>
<tr>
<td>FATP</td>
<td>Fatty Acid Transport Protein</td>
</tr>
<tr>
<td>Fdps</td>
<td>Farnesyl Diphosphate Synthase</td>
</tr>
<tr>
<td>Fgf21</td>
<td>Fibroblast Growth Factor 21</td>
</tr>
<tr>
<td>Fgfr1</td>
<td>Fibroblast Growth Factor Receptor 1</td>
</tr>
</tbody>
</table>
FH  Fumarate Hydratase
Foxp1  Forkhead Box P1
FXR  farnesoid X receptor
G6PD  Glucose-6-Phosphate Dehydrogenase
GlcNAc  N-acetylglucosamine
GLP-1  glucagon-like peptide-1
Gnmt  Glycine N-Methyltransferase
Gpnmb  Glycoprotein Nmb
GSK-3β  Glycogen synthase kinase 3
Gsta1  Glutathione S-Transferase Alpha 1
H&E  Haematoxylin and eosin
H+E  haematoxylin and eosin
HCC  hepatocellular carcinoma
hESC  Human embryonic stem cells
HIF  hypoxia inducible factors
Hmg1  High Mobility Group Box 1
HMGCS2  3-Hydroxy-3-Methylglutaryl-CoA Synthase 2
HMGCS2  3-Hydroxy-3-Methylglutaryl-CoA Synthase 2
HMW  high molecular weight
HNF4α  Hepatocyte Nuclear Factor 4 Alpha
Hprt  Hypoxanthine-guanine phosphoribosyltransferase
HSCs  hepatic stellate cells
Hsd17b7  Hydroxysteroid (17-Beta) Dehydrogenase 7
ICD-11  11th International Classification of Diseases
IGFBP1  Insulin Like Growth Factor Binding Protein 1
Il-6  Interleukin 6
iPSCs  induced pluripotent stem cells
Irs2  Insulin Receptor Substrate 2
KC  Kuppfer cells
Lpl  Lipoprotein Lipase
Lss  Lanosterol Synthase
LXR  Nuclear Receptor Subfamily 1 Group H Member 3
Ly6d  Lymphocyte Antigen 6 Complex, Locus
Mat2a  Methionine Adenosyltransferase 2A
MCDD  Methionine Choline Deficient Diet
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>MeCP2</td>
<td>methyl CpG binding protein 2</td>
</tr>
<tr>
<td>MLxlpl</td>
<td>MLX Interacting Protein-Like</td>
</tr>
<tr>
<td>Mmp12</td>
<td>Matrix Metallopeptidase 12</td>
</tr>
<tr>
<td>Mmp13</td>
<td>Matrix Metallopeptidase 13</td>
</tr>
<tr>
<td>Mthfd1</td>
<td>Methylene tetrahydrofolate Dehydrogenase (NADP+ Dependent) 1</td>
</tr>
<tr>
<td>Mtp</td>
<td>microsomal triglyceride transfer protein</td>
</tr>
<tr>
<td>Mvd</td>
<td>Mevalonate Diphasphate Decarboxylase</td>
</tr>
<tr>
<td>Myc</td>
<td>V-Myc Avian Myelocytomatosis Viral Oncogene Homolog</td>
</tr>
<tr>
<td>NAFLD</td>
<td>non-alcoholic fatty liver disease</td>
</tr>
<tr>
<td>Nanog</td>
<td>Nanog homeobox</td>
</tr>
<tr>
<td>NAS</td>
<td>NAFLD activity score</td>
</tr>
<tr>
<td>NASH</td>
<td>non-alcoholic steatohepatitis</td>
</tr>
<tr>
<td>NASH-CRN</td>
<td>National Institute for Health NASH Clinical Research Network</td>
</tr>
<tr>
<td>NEFAs</td>
<td>non esterified fatty acids</td>
</tr>
<tr>
<td>NRG1</td>
<td>Neuregulin 1</td>
</tr>
<tr>
<td>NRL</td>
<td>Nod-like receptors</td>
</tr>
<tr>
<td>Nupr1</td>
<td>Nuclear Protein 1, Transcriptional Regulator</td>
</tr>
<tr>
<td>Oct3</td>
<td>octamer-binding transcription factor 4</td>
</tr>
<tr>
<td>OGT</td>
<td>O-linked N-acetylglucosaminyltransferase</td>
</tr>
<tr>
<td>PAPP2C</td>
<td>Phospholipid Phosphatase 1</td>
</tr>
<tr>
<td>PCK</td>
<td>Phosphoenolpyruvate Carboxykinase 1</td>
</tr>
<tr>
<td>PCK1</td>
<td>Phosphoenolpyruvate Carboxykinase 1</td>
</tr>
<tr>
<td>Pcyt1a</td>
<td>Phosphate Cytidylyltransferase 1, Choline, Alpha</td>
</tr>
<tr>
<td>PLIN1</td>
<td>Perilipin 1</td>
</tr>
<tr>
<td>PLIN2</td>
<td>Perilipin 2</td>
</tr>
<tr>
<td>PNPLA3</td>
<td>Patatin-like phospholipase domain-containing protein 3</td>
</tr>
<tr>
<td>Pnpla3</td>
<td>Patatin Like Phospholipase Domain Containing 3</td>
</tr>
<tr>
<td>Pnpla5</td>
<td>Patatin Like Phospholipase Domain Containing 5</td>
</tr>
<tr>
<td>PPARα</td>
<td>Peroxisome proliferator-activated receptor alpha</td>
</tr>
<tr>
<td>PPARγ</td>
<td>Peroxisome proliferator-activated receptor gamma</td>
</tr>
<tr>
<td>Prrx1</td>
<td>Paired Related Homeobox 1</td>
</tr>
<tr>
<td>PSR</td>
<td>Picosirius red</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RNAPII</td>
<td>RNA polymerase II</td>
</tr>
<tr>
<td>RXR</td>
<td>retinoid X receptor</td>
</tr>
</tbody>
</table>
SAMe  S-adenosylmethionine
Sc4mol  Methylsterol Monooxygenase 1
Scd1  Stearoyl-CoA Desaturase 1
SDH  succinate dehydrogenase
Slpi  Secretory Leukocyte Peptidase Inhibitor
SNPS  single nucleotide polymorphisms
SOCS3  suppressor of cytokine signalling 3
Sqle  Squalene Epoxidase
Sreb1c  Sterol regulatory element-binding protein 1c
Sreb2  Sterol Regulatory Element Binding Transcription Factor 2
T2DM  Type 2 Diabetes Mellitus
Tet1  Ten eleven translocase 1
Tet2  Ten eleven translocase 2
Tet3  Ten eleven translocase 3
TG  Triglycerides
TGH  Triglyceride hydroxylase
Timp1  TIMP Metallopeptidase Inhibitor 1
TLR  Toll-like receptors
Tlr2  Toll-like receptor 2
Tlr12  Toll-like receptor 12
Tnfr  tumour necrosis factor receptor
TNFα  tumour necrosis factor alpha
Txnip  Osgin 1
UGT1A1  UDP Glucuronosyltransferase Family 1 Member A1
UPR  unfolded protein response
VEGF  vascular endothelial growth factor
vLDL  Very-low-density lipoprotein
WHO  World Health Organisation
Wnt3a  Wnt Family Member 3A
# Table of Contents

CHAPTER 1: INTRODUCTION ..................................................................................................................1

Background..............................................................................................................................................1

1.1 What is NAFLD? .................................................................................................................................2

1.2 Pathological Findings and Scoring .....................................................................................................3

1.2.1 Hepatic (simple) Steatosis ...........................................................................................................3

1.2.2 NASH ...........................................................................................................................................3

1.2.3 Cirrhosis .......................................................................................................................................4

1.3 NAFLD – A global epidemic ............................................................................................................6

1.3.1 NAFLD Prevalence .......................................................................................................................6

1.3.2 How common are NASH and NAFLD Cirrhosis? .......................................................................7

1.3.3 Who Gets NAFLD? ......................................................................................................................8

1.3.4 Genetics of NAFLD .....................................................................................................................10

1.4 What is the Global disease burden of NAFLD and what are the Clinical Needs? .......................10

1.5 Molecular Mechanisms of NAFLD Pathogenesis and Progression ..............................................11

1.5.1 Mechanisms of Lipid Accumulation ............................................................................................11

  Excess FA delivery due to adipose tissue dysfunction ....................................................................12

  De novo lipogenesis ..........................................................................................................................14

  Impaired fat export ...........................................................................................................................14

1.5.2 The Pathogenesis of NAFLD progressing to NASH and Hepatic Fibrosis .............................15

  Hepatocellular injury and apoptosis .................................................................................................15

  Inflammation .....................................................................................................................................16

  Fibrosis ...............................................................................................................................................19

1.6 Current and Future NAFLD Therapies .........................................................................................19

  Weight loss Intervention ...................................................................................................................19

  Pharmacotherapy trials .....................................................................................................................20

1.7 Modelling NAFLD in vitro and in vivo ..........................................................................................21

1.8 Epigenetics and NAFLD .................................................................................................................22

1.8.1 The Epigenetic Concept ..............................................................................................................22

1.8.2 DNA Methylation - Distribution, Regulation and Functional relevance ..............................23

1.8.3 CpG Islands as landmarks in mammalian genomes .................................................................24

1.8.4 Methylation Within and Out with promoter Sequences ..........................................................25

1.8.5 5hmC and Tet proteins and DNA demethylation pathways ....................................................27
2.5.4 Fluorescence microscopy.................................................................46
BODIPY 493/503 lipid marker.................................................................46
Mitosox Staining.....................................................................................46
2.6 Molecular Biology..............................................................................47
2.6.1 Expression Analysis........................................................................47
RNA extraction.......................................................................................47
RNA Quantification and Integrity ..........................................................47
Synthesis of complementary DNA (cDNA) by reverse transcription........48
Quantitative PCR (qPCR) of Gene Expression.......................................48
RNA biotinylated labelling and amplification:........................................50
Microarray Expression Analysis..............................................................50
RNA sequencing....................................................................................50
2.6.2 Epigenome Profiling.......................................................................51
DNA extraction......................................................................................51
DNA Quantification and Integrity ..........................................................51
DNA Sonication.....................................................................................51
Hydroxymethylcytosine DNA Immunoprecipitation.............................52
Quantitative PCR of Immunoprecipitated Products..............................53
Whole Genome Amplification (WGA)......................................................53
Ion Torrent Proton Sequencing..............................................................53
2.6.3 Liquid Chromatography Mass Spectrometry.................................54
2.6.4 Genotyping of TET1 Knock Out Offspring......................................54
2.7 Bioinformatic and Statistical Analysis .............................................55
2.7.1 Microarray Gene expression.........................................................55
2.7.2 RNA sequencing analysis .............................................................55
2.7.3 Epigenome profiling.....................................................................56
2.7.4 Null Hypothesis Testing Statistical Methods.................................57
2.8 Primer Sequences..............................................................................58
2.8.1 Gene Expression Quantitative PCR Analysis Mouse .....................58
2.8.2 Gene Expression Quantitative PCR Analysis Human.....................60
2.8.3 hmeDIP DNA Quantification Mouse..............................................61
2.8.4 hmeDIP DNA Quantification Human............................................62
2.8.5 Tet1 Knockout mouse genotyping primers......................................62
CHAPTER 3: EVALUATION OF CHOLINE AND METHIONINE DEFICIENT MICE: NEW INSIGHTS INTO THE MECHANISM OF STEATOSIS AND INSULIN RESISTANCE.........63
3.1 Introduction.......................................................................................63
5.3.3 High Fat Diet Induced Reversible 5hmC Enrichment in Genic Regions of Cholesterol Synthesis and Transport genes without changing global 5hmC content....
5.3.4 Genic 5hmC Enrichment as a Marker of Transcriptional Activity...
5.3.5 Altered TCA Metabolites...
4.4 Discussion...
4.4.1 DIO model Performance...
4.4.2 The Murine NAFLD transcriptome...
4.4.3 Murine NAFLD Drives Genic 5hmC Enrichment within Key Cholesterol Synthesis Mediators...
4.4.4 5hmC as a marker of tissue state...
4.4.5 Experiment Limitations...
4.5 Conclusions...

CHAPTER 5: EXAMINATION OF THE NAFLD HYDROXYMETHYLOME IN ESC DERIVED HEPATOCYTES...
5.1 Introduction...
5.1.1 AIMS:...
5.1.2 HYPOTHESIS:...
5.2 Materials and methods...
5.2.1 ESC cell maintenance and differentiation...
5.2.2 LPO treatment and NAFLD phenotyping...
5.2.3 Expression analysis...
5.2.4 5hmC Analysis...
5.3 Results...
5.3.1 Differentiation of hESC derived cells with hepatocyte like morphology and function...
5.3.2 ESCheps retain 5hmC and display a mature hepatocyte TET expression profile...
5.3.3 Modelling NAFLD in ESCheps...
5.3.4 LPO induces upregulation of NAFLD Pathways in ESCheps...
5.4.1 ESCheps are a novel model for examining hepatic 5hmC regulation in vitro...
5.4.2 LPO model performance...
5.4.3 5hmC profiles in ESCheps is typical of ESCs not Differentiated hepatocytes...
5.4.4 LPO intervention induces genic 5hmC enrichment in lipid synthesis and transport mediators...
5.4.5 Experimental limitations...
Conclusion...
CHAPTER 6: TET1−/− RESPONSE TO OBESOGENIC DIETARY INTERVENTION .......... 133
6.1 Introduction ............................................................................................................. 133
6.1.1 AIMS .................................................................................................................. 134
6.1.2 HYPOTHESIS .................................................................................................. 134
6.2 Material and Methods ........................................................................................... 135
6.2.1 Animals ............................................................................................................. 135
6.2.2 Genotyping ...................................................................................................... 135
6.2.3 Liver and serum measurements ...................................................................... 135
6.2.5 Expression analysis ......................................................................................... 135
6.2.6 5hmC Analysis ............................................................................................... 136
6.2.7 Energy homeostasis analysis .......................................................................... 136
6.3 Results ................................................................................................................ 137
6.3.1 Frequency of Tet1 progeny and effect on the hydroxymethylome .......... 137
6.3.2 Metabolic effect of Tet1 deficiency ................................................................. 140
6.3.3 Tet1 deficiency confers a protective hepatic transcriptome ....................... 143
6.3.4 Further investigation of the Tet1 ................................................................. 146
Discussion ................................................................................................................. 147
Conclusion ............................................................................................................... 150
CHAPTER 7: THESIS DISCUSSION ........................................................................... 151
General Discussion: ................................................................................................. 151
7.1 What this thesis adds ......................................................................................... 151
7.2 Limitations and opportunities ........................................................................ 154
7.3 Future work ....................................................................................................... 155
7.4 Exciting themes in current NAFLD Research – Pathogenesis, biomarkers and interventions ........................................................................................................... 156
   The gut microbiome ............................................................................................. 156
   Non-invasive Biomarkers of progression to NASH ........................................... 156
   Restoration of energy balance ............................................................................ 157
7.5 Conclusion ......................................................................................................... 157
BIBLIOGRAPHY ............................................................................................................. 159
Presentation and Publication of data from this thesis ......................................... 189
APPENDICES .............................................................................................................. 190
Appendix 1: Gene list and fold changes for 17 week high fat diet fed mouse liver tissue .................................................................................................................. 190
Appendix 2: Gene list and fold changes for human ESC derived hepatocytes treated with LPO energy substrate cocktail versus untreated cells ....................... 193
CHAPTER 1: INTRODUCTION

Background

In 2016 a report was published in the Lancet surveying 19.2 million individuals worldwide and concluding that for the first time the number of overweight people now outnumber those malnourished. Since 1975 the prevalence of obesity has tripled and now stands at 10.8% and 14.9% for men and women globally. We are in a time of unprecedented inequality with a surge in disorders related to calorie excess punctuated by pockets of severe deprivation and malnutrition. At the current rate of change, the WHO target to halt the global obesity increase by 2020 is likely to be unachievable. Considerable effort is now being made in public health to promote active living and dietary reform while at the same time developing measures to limit the disease burden associated with long term calorie excess. Among the diseases which have shown an increase in association with obesity is non-alcoholic fatty liver disease (NAFLD). This is of particular concern as studies undertaken to estimate the prevalence of NAFLD revealed a much greater prevalence than expected, in excess of one third in the adult population of some developed nations. As such there has been a renewed focus on the characterisation and classification of this disease and also on the biological processes surrounding its development and associated negative health outcomes.

The early release ‘Beta’ version of the 11th International Classification of Diseases (ICD-11) defines NAFLD as ‘fatty liver related to insulin resistance in the absence of significant alcohol consumption’. This refers to a spectrum of hepatic pathological findings found in people with no past or present history of alcohol excess who typically display one or more clinical features of the metabolic syndrome (central obesity, hypertension, Type 2 diabetes (T2DM) dyslipidaemia). The unifying feature is the presence of fat in large lipid droplets within hepatocytes known as hepatic steatosis. A sub population of patients with hepatic steatosis display a more aggressive, inflammatory phenotype termed non-alcoholic steatohepatitis (NASH) which may in turn progress to irreversible liver cirrhosis and ultimately liver failure or primary liver cancer (hepatocellular carcinoma).
1.1 What is NAFLD?

Preceding our current understanding of NAFLD, patients presenting with liver disease of this type were presumed to either underestimate their lifetime alcohol consumption or were deemed to have an idiopathic form of liver injury. Early reports from Japan in the 1960s described an association between insulin resistance, obesity and hepatic steatosis. Some 20 years later, American authors further described the association between obesity, diabetes and hepatic steatosis and proposed a functional association. Furthermore, in 1979 Adler et al described distinct subtypes of liver histology in obese patients ranging from simple fatty change to an inflammatory hepatitis, fibrosis and advanced cirrhosis, a finding replicated a larger cohort ten years later. These initial studies suggested for the first time that this obesity related liver pathology had the potential to progress to a severe life threatening disease. It was Ludwig et al at the Mayo clinic in 1980 who first proposed the term ‘non-alcoholic steatohepatitis’ or NASH.

In the intervening years since the initial NAFLD descriptions there has been a global acceleration in rates of obesity, T2DM and sedentary living and parallel to this NAFLD is now the most common form of chronic liver disease globally.

The most commonly observed form of NAFLD is the uncomplicated accumulation of fats within hepatocytes typically termed hepatic (or simple) steatosis. Given the near identical histological appearance of alcoholic liver disease and NAFLD, it is important to define the term ‘non-alcoholic’ from a lifestyle perspective. The National Institute for Health NASH Clinical Research Network (NASH-CRN) has adopted a practically useful threshold value for alcohol intake of 10g ethanol per day for women and 20g ethanol per day for men (equating to approximately one or two drinks respectively). This cut off is both helpful from a public engagement perspective and is based on a conservative estimate of the level required for alcohol related liver cirrhosis. As such this, or a similar threshold has been adopted widely in epidemiological studies of NAFLD prevalence. The term NASH is used to describe livers in which steatosis is complicated by substantial inflammation and hepatocellular injury and may or may not be associated with a degree of hepatic fibrosis (the laying down of new collagen between cells). In a minority of cases NASH can progress to irreversible hepatic cirrhosis where much of the liver parenchyma is replaced by fibrotic tissue and loss of all the original pathological features is common making diagnosis of the original aetiology difficult. In these cases the liver is unable to detoxify the blood or excrete bilirubin resulting in jaundice and encephalopathy, with ensuing hepatic failure. Patients with NAFLD, with or without NAFLD cirrhosis, are at increased risk of hepatocellular carcinoma (HCC).
Ultimately at the point of hepatic failure, orthotopic liver transplantation is the only therapeutic option and even in these cases disease recurrence is common in the transplanted organ 20.

1.2 Pathological Findings and Scoring

A uniform method of classification for each disease stage is essential for accurate estimation of disease prevalence, monitoring health outcomes and interpreting therapeutic trials. A number of scoring systems have evolved with the aim of patient risk stratification for liver and all cause morbidity/mortality with minimal intra-observer variation 21–24. Of these, the most recent and widely adopted was designed and validated by the National Institute for Health NASH-CRN pathology committee 24. The NASH-CRN scoring system encompasses the NAFLD activity score (NAS), fibrosis stage and also the presence of characteristic NASH histopathological features 24.

1.2.1 Hepatic (simple) Steatosis

Hepatic steatosis is defined as a liver composed at greater than 5.5% fat by weight percentage. This is based on the 95th percentile in the non overweight US population with no features of the metabolic syndrome and without a history of alcohol excess 3. Practically, this is defined on histological examination as the presence of large lipid droplets (macrovacuoles) in ≥5% of hepatocytes. Human NAFLD results in the accumulation multiple lipid species including triglycerides, diacylglycerol, ceramides, free fatty acids (FA), sphingolipids and cholesterol esters, however it is predominantly triglycerides which are visible on light microscopy 25. The distribution of lipid droplets is typically scattered around the small branches of the hepatic vein (sinusoidal, zone 3) however a periportal distribution (zone1) is reported in obese adolescents 26. In extensive steatosis, lipid droplets may be scattered thought out the acinus (acinar) or randomly scattered (azonal). The macrosteatotic cells are readily identifiable with a large circular vacuole displacing the nucleus to the cell periphery (Figure 1.1B).

1.2.2 NASH

In a subset of cases steatosis is complicated by the presence of substantial inflammation with evidence of hepatocellular injury and reaches a threshold to be termed NASH. This threshold is very important as the presence of these features confers a significant increase in risk of liver related and all-cause mortality over and above steatosis alone (discussed in 1.4). The presence of multiple inflammatory cell types including macrophages, Th1 cells, NKT cells and plasma cells is common 27. Hepatocellular injury is recognised by either ballooning or the presence of
eosinophilic protein aggregates known as Mallory-Denk bodies. Fibrosis is graded by location and severity deemed mild, moderate, bridging or cirrhosis. While the NASH-CRN scoring system provides the most comprehensive and validated system for differentiating steatosis from NASH, terms such as substantial, mild and moderate confer a degree of both intra and inter observer variation with kappa statistics reported as low as 0.45. Furthermore, the NASH-CRN scoring system was designed by specialist hepatopathologists and the utility of generalising this system to non-specialists has been questioned.

1.2.3 Cirrhosis

Cirrhosis is the endpoint of the majority of chronic liver diseases and describes a pervasive network of complex collagen between vascular structures with disruption of the hepatic architecture. Much of the liver parenchyma is irreversibly lost and loss of hepatic synthetic function (ability to synthesise albumin and clotting factors), detoxification capacity and impaired blood flow leads to clinically decompensated liver disease. Furthermore, patients are at an increased risk of malignant transformation to HCC.
Figure 1.1 Stages of NAFLD.  A) Diagrammatic representation of hepatic steatosis in human NAFLD. Affected hepatocytes are typically found in the peri-sinusoidal distribution (red) but can be periportal (blue), azonal (middle) or panacinar (right). P = portal triad, S = hepatic sinusoid. B) H+E stained liver section of patients with NASH. Macrovesicular steatosis with displacement of the hepatocyte nuclei (Blue arrows) and lobular inflammatory infiltrate (Black arrow) (source Brunet et al 26). C) Markers of hepatocellular injury in NASH. Ballooned hepatocyte (blue arrows) containing rope like structure of Mallory-Denk body (blue arrow). D and E) NASH complicated by bridging fibrosis between vascular structures (D) and global deposition causing cirrhosis (E). Blue = trichrome blue staining for new type 1 collagen deposition. (histology images reproduced with permission 17,30)
1.3 NAFLD – A global epidemic

1.3.1 NAFLD Prevalence

NAFLD prevalence varies considerably by geographical region almost certainly due to ethnic variation and environmental factors. It is clear however that the prevalence of NAFLD regardless of region correlates highly with the increasing prevalence of obesity and T2DM. Furthermore the idea that NAFLD is a ‘Western disease’ is now widely discredited with NAFLD commonly found in the population of all nations although the incidence is lowest among nations where over nutrition is less common.

The reported prevalence of NAFLD is heavily dependent on the diagnostic method. The ‘gold standard’ for identification of NAFLD is liver biopsy with histological analysis. Perhaps the most widely quoted study examining all comers to a US army medical (n=328, age 18-70) centre and performing biopsies on all patients with ultrasound evidence of disease estimated NAFLD prevalence at 46% with the highest prevalence in Hispanic participants. One unselected Canadian autopsy study demonstrated a prevalence of any NAFLD stage of 53%. A similar autopsy study from Northern Greece reported a prevalence 71.1%. Minervini et al retrospectively examined liver biopsies from potential live liver donors and report NAFLD of some stage in the majority (52%) of patients examined. While this diagnostic technique is likely to give the most precise measurement of NAFLD prevalence in a given population, it is an invasive procedure, is expensive and is associated with rare but serious complications including haemorrhage and pneumothorax. As such it cannot be undertaken indiscriminately in large populations and it is likely these studies over estimate NAFLD incidence by way of patient selection bias.

Non-invasive methods for the diagnoses of NAFLD include liver imaging and the measurement of serum ALT and AST enzymes. Liver enzyme analysis is highly insensitive and can be normal even in severe fibrosis. Indeed the majority of patients with hepatic steatosis have normal ALT levels. Therefore, the use of at least one imaging modality is generally required for non-invasive prevalence studies. The ‘gold standard’ imaging modality is proton magnetic resonance spectroscopy ($^{1}$H- MRS) with the greatest correlation to biopsy findings even at low levels of hepatic steatosis. Utilising ($^{1}$H- MRS) to examine 2349 patients from a multi-ethnic background enrolled in the Dallas heart study, Browning et al reported a prevalence of NAFLD (as defined as hepatic triglyceride content of $\geq$ 5.5%) at 31% with again the greatest prevalence among Hispanics (45%) with the lowest prevalence among
African American men (23%) \(^2\). Although precise, \(^{1}\text{H-MRS}\) is expensive and time consuming. Hepatic ultrasound scanning (USS) is a rapid and relatively inexpensive method of liver imaging and compares favourably \(^{1}\text{H- MRS}\) at detecting histologically confirmed NAFLD \(^{38,39}\). A recent meta-analysis of 45 studies performed globally utilising predominantly US assessment provides currently the most comprehensive and indeed most alarming picture of this global epidemic. Taken together the pooled NAFLD prevalence globally was estimated to be 25.34% ranging from 31.79 % in South America to 13.48% in the African population.

1.3.2 How common are NASH and NAFLD Cirrhosis?

As described in 1.1, NASH represents a subset of patients with NAFLD whereby hepatic steatosis is complicated by evidence of hepatocellular injury and substantial lobular or portal inflammation often with a degree of pericellular fibrosis \(^{24}\). The diagnosis of NASH is a pathological finding and as such liver biopsy is required. Therefore the prevalence of NASH is estimated on autopsy or biopsy series which may be confounded by patient selection, sampling error and intra-observer variability in histological grade \(^{40,41}\). The aforementioned unselected NAFLD biopsy studies report NASH prevalence at 12% \(^{42}\), 15% \(^4\) and 2.7% rising to 19% in obese individuals \(^{33}\). Furthermore, a biopsy series form obese patients undergoing bariatric surgery found a NASH prevalence of 37% \(^{43}\). From the available data, it can be speculated that the NASH prevalence in North America is between 5-10 percent of the US total population \(^{44}\). Thus far the prevalence of NASH in the UK population is unknown.

Studying the proportion of patients that progress from NASH to cirrhosis is complicated by the relative absence of NAFLD histological findings once cirrhosis is established. Where there is high diagnostic doubt a patient may be labelled ‘cryptogenic cirrhosis’. Furthermore, patients rarely have repeated biopsies as this is unlikely to influence clinical management and as such studies examining rate and incidence of NAFLD progressions are lacking. Cumulatively small studies suggest 30-40% of patients with NASH show some fibrotic progression over a 5 year period \(^{45,46}\). One study reports a point prevalence of advanced fibrosis of 2.7% suggesting there are greater than 2 million individuals in the US who unknowingly have advanced liver disease and are at high risk of HCC and cirrhosis \(^{47}\).
1.3.3 Who Gets NAFLD?

Widely reported independent risk factors for NAFLD include high BMI (>25 kg/m\(^2\) in European people, >22.5 kg/m\(^2\) in Asia), insulin resistance, glucose intolerance, impaired fasting glucose, T2DM, dyslipidaemia and hypertension) \(^3\)-\(^5\),\(^14\)-\(^42\)\(^-\)^44. Studies focussed on high risk populations suggest a NAFLD prevalence of up to 90% and 57% in morbidly obese and T2DM populations respectively \(^39\),\(^43\) (Figure 1.2). Furthermore, patients with diabetes and NAFLD are more likely to progress to display advanced fibrosis \(^48\). Other non-dietary related risk factors include male gender, increasing age and family history (of T2DM, NAFLD or early cardiovascular death) \(^2\),\(^47\),\(^49\). Interestingly, the protective influence of female gender is present only until the age of fifty and is reversed following the menopause suggesting an important role for oestrogen in preventing hepatic lipid deposition \(^50\). The presence of NAFLD in lean patients has been repeatedly reported in Asia and the US, presumably due to a genetic preference for intra-abdominal rather than peripheral adipose deposition although this requires further study \(^51\),\(^52\).
Figure 1.2: Estimated prevalence of each NAFLD stage in the adult population. Global data is available for steatosis. Data from NASH prevalence is predominantly in US population. Associated risk factors for development and progression of NAFLD are shown. (* as defined by WHO criteria)
1.3.4 Genetics of NAFLD.

Thus far there have been five genome wide association studies (GWAS) of NAFLD risk (Reviewed by Anstee et al and Wood et al). Although multiple single nucleotide polymorphisms (SNPs) were identified in biologically plausible pathways, only one gene demonstrated a repeatedly significant association with hepatic lipid load, PNPLA3 also known as ADIPONUTRIN. Romeo et al were the first to report the sequence variation (rs738409) isoleucine to methionine PNPLA3 amino acid substitution as independently associated with hepatic lipid load in a mixed ethnic population from the Dallas Heart Study. PNPLA3 was also found to have the greatest background allele frequency in Hispanic populations and contributed significantly to the increased risk in this ethnic group. This finding was validated in a second larger study with carriage of a near identical SNP (PNPLA3 rs738408) demonstrating a strong association with severity of steatosis and also lobular inflammation and fibrosis in histological grading. Mechanistically, PNPLA3 is a membrane associated triglyceride hydrolase enzyme with a putative role in the packaging and export of TGs in vLDL particles, and patients homozygous for the PNPLA3 variant display reduced vLDL export. Although no other loci have been repeatedly associated with NAFLD, one study reports that an accumulation of SNPs in genes associated with ‘one carbon metabolism’ correlated with steatosis burden. Further work is warranted to translate these findings into translational value perhaps utilising a ‘personalised medicine’ approach.

1.4 What is the Global disease burden of NAFLD and what are the Clinical Needs?

Patients with NAFLD have a 1.7 fold higher standardised mortality than those without. This disease burden is predominantly but not exclusively borne by the ~20-30% of NAFLD patients who have NASH, and is particularly high in those with significant fibrosis. Interestingly liver related mortality is only the third highest cause of death reinforcing the importance of NAFLD as a multisystem disease and the need for it to be treated as such clinically. Deaths from cardiovascular disease and common cancer types are the predominant causes of death in NAFLD and are significantly higher than the general population with particularly high risk in patients with T2DM. The extent to which these effects are caused by NAFLD per se or simply a reflection of common cardiovascular risk factors remains unclear. Liver related death in NAFLD is predominantly due to cirrhosis and hepatic failure or HCC, and is greater than the general population in patients with NASH but not benign steatosis alone. Bhala et al followed 247 patients with NAFLD, advanced fibrosis or cirrhosis for a mean of 85 months and reported a liver related death or requirement for transplant incidence of 13.4%, similar to that of hepatitis C cirrhosis. Similarly, Ascha et al followed 195 patients with
NAFLD cirrhosis for 4 years and reported a cumulative year incidence of HCC at 2.6% \(^{70}\). Between 2010-2014 there were over 2000 liver transplant procedures in the UK with 7.9% of these resulting from metabolic liver disease (source: NHS Blood and Transplant Service \(^{71}\)). NAFLD is on course to become the leading cause of liver transplantation in the US over the next two decades \(^{20,72}\).

Given the reported prevalence of NAFLD, the continued rise of NAFLD related risk factors and the repeatedly demonstrated higher morbidity and mortality through a variety of intra and extra-hepatic pathologies, NAFLD presents an enormous global health concern. I therefore propose that the current clinical needs in NAFLD management are:

1) Biomarker discovery for the precise and objective discrimination of NASH from simple steatosis in an affordable and patient acceptable process.
2) Biomarker discovery for prediction of patients at risk of liver related death to allow optimum screening and targeting of resources (bariatric surgery).
3) Identification of novel therapeutics for the treatment of NASH, the prevention of fibrosis and the amelioration of NAFLD related extra hepatic complications.

1.5 Molecular Mechanisms of NAFLD Pathogenesis and Progression

Meeting these clinical needs requires a detailed understanding of the molecular pathology of lipid accumulation, NAFLD progression and the role(s) that NAFLD may play in cardiovascular disease and common malignancy.

1.5.1 Mechanisms of Lipid Accumulation

The common feature in NAFLD is an accumulation of multiple lipid species within hepatocytes. This occurs due to an imbalance in lipid synthesis and uptake versus lipid clearance. It can be argued that hepatic lipid sequestration is a normal physiological process, however given its association with negative health outcomes, this is difficult to justify. It is widely accepted that liver tissue is not a physiological reservoir for fat and that steatosis a pathological consequence of chronic positive energy balance \(^{72}\). Our understanding of the physiological processes underlying this imbalance continues to expand \(^{73,74}\).
**Excess FA delivery due to adipose tissue dysfunction**

In the fed state, insulin suppresses triglyceride hydrolysis and promotes triglyceride synthesis in adipocytes. In patients with diabetes, adipose tissue insulin resistance and inflammation induces a chronic state of lipolysis, with the subsequent secretion of FAs into serum which are taken up by the liver. Patients with NAFLD have elevated levels of FAs in serum with a higher rates of adipose tissue lipolysis. Furthermore 59% hepatic of TGs in NAFLD are estimated to be synthesised from plasma FAs with a minority generated by de novo lipogenesis (DNL) and dietary intake as estimated by measuring sequestration of radiolabelled non esterified fatty acids (NEFAs). Furthermore, elevated transcripts of the FA uptake transporter CD36 has been reported in NAFLD suggesting an adaptation to increase FA sequestration from serum.

In addition to insulin resistance and FA release, adipose tissue plays a further role in lipid accumulation through the secretion of inflammatory mediators and alterations in adipose tissue hormone (adipokine) production. Adiponectin is a large adipokine secreted by adipocytes in various polymers. Low levels of the high molecular weight (HMW) multimer of adiponectin are observed in patients with insulin resistance, obesity and NAFLD. During fasting conditions, high levels of circulating HMW adiponectin bind receptors on hepatocytes and potentiate the activity of key metabolic mediators Peroxisome proliferator-activated receptor alpha (PPARα) and 5' AMP-activated protein kinase (AMPK) which in turn inhibit lipogenesis and promote FA oxidation to generate substrates for the TCA cycle, thus promoting lipid clearance. The converse in insulin resistance is therefore likely to be steatogenic. High levels of tumour necrosis factor alpha (TNFα) and Interleukin 6 (IL-6) released from inflamed dysfunctional adipose tissue have also been shown in patients with obesity and correlate with NAFLD score. The role these mediators have on lipid accumulation is currently unclear.

A mechanism for TNFα and IL-6 mediated hepatic insulin resistance has however been described through upregulation of the suppressor of cytokine signalling 3 (SOCS3) with subsequent inhibition of insulin signal transduction.
Figure 1.3 Proposed Mechanisms of Lipid Accumulation in NALFD. Adipose tissue dysfunction and gut derived factors provide extrahepatic substrates and stimuli that promote lipid synthesis inhibit of FA oxidation. Upregulation of FATP2+5 and CD36 increase uptake of NEFAs from serum. The FA pool is further increased by an increase in DNL stimulated by high delivery of insulin, glucose, adiponectin and inflammatory mediators to the liver. Upregulation of vLDL secretion either reaches maximum rate or is altered by common (Pnpla3) or rare (Apob100, MTP) gene mutations. FATP = Fatty Acid Transport Protein, CD36 = cluster of differentiation 36, AdipoR2 = adiponectin receptor 2, Sreb1c = Sterol regulatory element-binding protein 1c, Chrebp = Carbohydrate-responsive element-binding protein, Tnfr = tumour necrosis factor receptor, ER = endoplasmic reticulum, Ces = carboxylesterases, Mtp = microsomal triglyceride transfer protein, vLDL = very low density lipoproteins.

**ADIPOSE DYSFUNCTION**
- Insulin resistance
- Increased lipolysis
- Suppressed adiponectin secretion
- TNFα + IL-6 secretion

**GUT DERIVED FACTORS**
- Endotoxamia
- Gut signalling molecules
De novo lipogenesis

An alternative source of hepatic lipids is de novo lipogenesis (DNL) in which excess carbohydrate is synthesised into new FAs, contributing to the hepatic fatty acid pool and providing substrates for triglyceride formation. DNL is the second most important source of triglycerides in NAFLD, accounting for around a quarter of hepatic fat 77. Under normal physiological conditions, DNL is precisely regulated at the transcriptional level with insulin and glucose levels activating transcription factors Sterol regulatory element-binding protein (SREB-1c) and Carbohydrate-responsive element-binding protein (ChREBP) respectively. Overexpression of SREB-1c, its trans activating co-enzyme (LXRα) and target genes (Fatty acid synthase; FAS and Acetyl CoA carboxylase; ACC1) have been demonstrated in NAFLD patients presumably in response to the hyperinsulinaemia and hyperglycaemia characteristic of T2DM 89. ACC1 generates the ‘molecular switch’ malonyl CoA in cytosol, which inhibits β-oxidation and provides a substrate for FA synthesis. Interestingly, overexpression of ChREBP in mice generates a hepatic steatosis with little or no effects on insulin resistance, suggesting that the interplay between hepatic lipids and insulin action is more complex in human T2DM 90. This would be consistent with the notion that hyperglycaemia alone is unlikely to be sufficient to drive hepatic steatosis as this is a recognised but uncommon finding in lean patients with T1DM. More recently, a novel mechanism of DNL in NAFLD has been described through endoplasmic reticulum (ER) dysfunction 91. In the context of ER stress, translated proteins are incorrectly folded during post translational modification. Accumulation of these misfolded proteins stimulates a cell signalling response to reduce protein synthesis and in severe cases active apoptosis, known as the unfolded protein response (UPR) which is a feature of chronic liver disease 92,93. In addition to this the UPR activates sterol regulatory element binding proteins which in turn stimulate FA synthesis 93.

Impaired fat export

Lipid export is a normal and essential function for hepatic fat balance and the delivery of triglycerides and cholesterol to other tissues including skeletal muscle, adipose tissue and cardiac muscle. Triglycerides, cholesterol and cholesterol esters are primarily exported from hepatocytes in vLDL particles. vLDLs are a lipoprotein particle with a hydrophilic lipid bilayer exterior housing a hydrophobic core of the TG and cholesterol load. Each particle is stabilised with one apolipoproteinB-100 polypeptide chain. vLDL assembly and lipidification occurs in the ER of hepatocytes assisted by a range of triglyceride hydrolase proteins including
Triglyceride hydroxylase (TGH), Microsomal triglyceride transfer protein (MTP), the carboxylesterase (Ces) enzymes and PNPLA3, the highly prevalent risk allele for steatosis. This process appears to function normally in NAFLD, and as expected vLDL excretion is higher in patients with steatosis compared to patients with normal hepatic lipid content. Interestingly the vLDL export rate in NAFLD plateaus at a hepatic fat content of 10% indicating that the liver attempts to normalise hepatic lipid content but is unable to do so beyond a certain level.

1.5.2 The Pathogenesis of NAFLD progressing to NASH and Hepatic Fibrosis

**Hepatocellular injury and apoptosis**

The ‘two hit hypothesis’ proposed by Day and James in 1998 suggested that the accumulation of hepatic triglycerides as a result of insulin resistance sensitised the liver to injury and could be deemed the ‘first hit’. The second hit is thought likely to be due to oxidative stress and lipid peroxidation which triggers the development of NASH. This suggests that triglyceride accumulation is harmful and pro-inflammatory which may not be the case. Sequestration of triglycerides within lipid droplets may actually be highly protective. Diacylglycerol acyltransferase 1 and 2 (DGAT1/2) enzymes catalyse the final step in hepatic triglyceride synthesis. Knock down of DGAT2 in mouse models reduces hepatic lipid load but worsens hepatocellular injury. Mice over expressing DGAT1 in adipocytes and macrophages are relatively protected from insulin resistance and systemic inflammation. Furthermore, mice fed a choline deficient diet (CDD) diet to induce severe hepatic steatosis by reduction of vLDL export, and humans with Apo B100 mutations with hepatic steatosis do not display insulin resistance. Thus, it may be that when the capacity to safely sequester lipid as triglyceride in both the liver and fat tissue is saturated, it is the generation of other non-triglyceride lipids which play an important role in hepatocellular injury, chronic inflammation and initiation of the wound repair process. The term lipotoxicity is used to describe the harmful effect of these non TG lipid species on hepatocytes and is accepted as a key process in the conversion to NASH. Non TG lipids including ceramides, cholesterol, FAs, diacylglycerol and acylcarnitines are all elevated in patients with NASH and are known to induce hepatocellular injury. Mechanistically this occurs through a number of pathways. For example, FA accumulation in hepatocytes activate Fas ligands and GSK-3β and induce hepatocyte apoptosis, a potent stimulator of stellate cell activation and fibrosis. Clearance of high levels of FAs through beta-, microsomal and peroxisomal oxidation generates reactive oxygen
species (ROS) and subsequently intracellular oxidative stress and hepatocellular injury. High levels of free cholesterol and FAs within cells generate ER stress activating the unfolded protein responses and subsequent apoptosis. Similarly, high levels of ceramides are a feature of NAFLD and induce both ROS generation and apoptosis.

In addition to lipotoxicity, a myriad of other contributory factors have been proposed to drive NASH progression including oxidative stress, genetic predisposition, mitochondrial dysfunction, gut microbiota and adipokine dysregulation (Figure 1.4). Thus, a more multifactorial multi-hit model has been proposed in which all of these factors contribute to triggering NASH, with the ‘safe’ sequestration of triglyceride within lipid droplets deemed a defence mechanism by which the liver attempts to buffer the toxic effects of lipid load.

**Inflammation**

A mixed cellular inflammatory infiltrate is a required histological feature of NASH diagnosis. Tempering this may ameliorate the fibrogenic wound healing response associated with progression to fibrosis and cirrhosis. Pivotal systems in the induction and propagation of the hepatic inflammatory response include adipose tissue dysfunction and alterations in the adipose secretome, cytokine release from injured hepatocytes, lipotoxicity and Kupffer cell (KC) activation. KCS are specialised macrophages lining the walls of liver sinusoids which play a central role in the hepatic inflammatory response (Figure 1.4). Due to constant hepatic portal delivery of blood from the gastrointestinal tract, the liver is a highly immunocompetent organ. KCS detect and respond to combinations of signalling molecules from the enterohepatic circulation and remove dangerous pathogens. This is done primarily through interaction with cell surface receptors known as Toll-like or Nod-like receptors (TLRs or NRL) of which there are various subtypes with ligand specific capacity. Depending of the summation of the signals received via TLR or NLRs, Kupffer cells reside in either a quiescent anti-inflammatory state (M2) or an activated proinflammatory state (M1) and hepatic inflammation in the liver is significantly affected by the ratio of this phenotype.

It is widely recognised that KC activation is important in NASH and it occurs through a number of mechanisms. Low levels of adiponectin from insulin resistant adipocytes drives a proinflammatory shift to M1 phenotype activating KCS. Injured hepatocytes release a cocktail of inflammatory signalling molecules known as danger associated molecular patterns (DAMPS), which activate KCS via TLR and NLR binding. FAs induce KC activation through
TLR-2 or TLR-4 via a TLR4 ligand known as fetuin-A, and FA, cholesterol and ceramide lipotoxins can activate KCs through the generation of oxidative stress\textsuperscript{112}. Finally, alterations in the gut microbiome flora may significantly influence NASH progression through changes in the composition of signalling molecules passed through the enterohepatic circulation\textsuperscript{113,114}. 

Figure 1.4. A summary of the multifactorial mechanisms by which hepatic steatosis progresses to NASH and the subsequent laying down of ECM by activated hepatic stellate cells in the generation of hepatic fibrosis.
**Fibrosis**

Although not a required histological finding for the diagnosis of NASH, the presence of fibrosis is the strongest predictor of overall mortality and liver transplantation. Cirrhotic livers have a 4-7 fold increase in collagen levels representing a disproportionate increase in Type 1 collagen, laminin and proteoglycan components of the extracellular matrix. To some extent fibrosis is reversible, however prolonged fibrotic damage leads to tissue remodelling from which resolution is limited. An extensive literature base is consistent with the activation of hepatic stellate cells (HSCs) as the critical player in fibrosis deposition. It is as yet unclear why some patients with NASH develop severe fibrosis and others are relatively spared and this remains a key clinical question. In the context of liver injury HSCs undergo transformation from predominantly quiescent, vitamin A-storing cells to a myofibroblast like cells with fibrogenic properties. Furthermore, HSCs play an important role in vascular tone and contribute to portal hypertension. HSCs are activated by many of the same pathways that are implicated in steatosis and inflammatory generation. A potent activator of HSCs is the profibrogenic cytokine TGFβ which is secreted predominantly by activated KCs but also be damaged hepatocytes. Other key processes include the differentiation of hepatic progenitor cells into profibrotic ‘reactive ductules’ with secretion of profibrogenic cytokines. Germ free mice and mice deficient in TLR4 are protected in surgical models of fibrosis suggesting endotoxaemia also plays a role. Relative hypoxia in injured liver induces hypoxia inducible factors (HIF) and vascular endothelial growth factor (VEGF), which stimulates HSCs and parallels fibrogenesis. Blocking VEGF inhibits fibrosis in animal models and along with TGFβ present key therapeutic targets in chronic liver disease.

1.6 Current and Future NAFLD Therapies.

There are currently no acceptable pharmacological interventions for the treatment NAFLD. Potential therapies include targeting the metabolic drivers of insulin resistance and hyperglycaemia as well as limiting hepatocyte inflammation and fibrogenesis.

**Weight loss Intervention**

By far the most preferable approach to the treatment of NAFLD is lifestyle modification. Increasing activity levels, weight loss and a reduction in dietary fats are advised. Weight loss of 5-10% improves multiple facets of the metabolic syndrome including levels of hepatic fat. Crucially, weight loss of around 10% results in a significant improvement in levels of
lobular inflammation, hepatocellular injury and fibrosis demonstrating NASH reversibility

Unfortunately, the beneficial effects of dietary intervention are typically transient with a recurrent in obesity and insulin resistance. Surgical intervention for obesity is increasing globally with procedures such as gastric banding, sleeve gastrectomy and Roux-en-y gastric bypass demonstrating significant improvements in all features of the metabolic syndrome in addition to almost complete reversal of NASH. Furthermore, the weight loss effects of bariatric surgery are more sustained.

**Pharmacotherapy trials**

As described in 1.5, insulin resistance in peripheral tissues is a key pathogenic component of NAFLD though FA, insulin and glucose delivery to the liver. Thus the effects of available treatments for T2DM have been examined for efficacy in NAFLD. The PPARγ agonists pioglitazone and rosiglitazone are used for the treatment of T2DM and are effective in reducing glucose levels, improving insulin sensitivity and reducing hepatic fat content probably though the promotion of peripheral adipocyte differentiation. However, randomised double blind control trials of rosiglitazone and pioglitazone for the treatment of NASH show improvement in steatosis but no improvement in fibrosis score. Given the undesirable side effects of weight gain with increased risk of fractures and bladder cancer, the use of these drugs for NASH cannot be supported. Trials to examine the efficacy of agonists of the more liver specific PPARα/δ in the context of NASH are underway (www.clinicaltrials.gov NCT01694849).

Farnesoid X receptor (FXR) is a nuclear receptor which is directly bound and activated by bile acids following food ingestion, translocates to the nucleus and regulates transcription of a number energy and inflammatory processes in liver, adipose and skeletal muscle. FXR mediated pathways include cholesterol and triglyceride synthesis, cytokine production and gluconeogenesis. Following positive results in mouse studies, human administration of the synthetic bile acid and RXR ligand obeticholic acid improved insulin sensitivity with some improvement in serum markers of liver fibrosis although critically there was no histological analysis. Interim analysis from a more in depth human trial of this drug suggests some improvement in NASH histology but long term follow up is awaited. Similarly, the glucagon-like peptide-1 (GLP-1) agonist liraglutide has recently been shown to improve NASH and fibrosis scores in a small cohort of patients and this warrants expanded studies.
1.7 Modelling NAFLD *in vitro* and *in vivo*

The recapitulation of NAFLD in rodents and *in vitro* culture systems has been employed in an attempt to understand pathophysiology and screen novel therapeutics. The ideal is to replicate all the histological and metabolic characteristics of the human spectrum including obesity, dyslipidaemia, hyperglycaemia, insulin resistance, adipose dysfunction and a chronic inflammatory state. No model performs this implicitly and as such caloric excess through the use of high saturated fat diets can induce obesity, steatosis and insulin resistance and with protracted feeding can induce some elements of NASH particularly in combination with high sucrose \(^\text{139-142}\). Restriction of dietary choline generates severe hepatic steatosis and with additional restriction of methionine inducing severe inflammatory hepatitis \(^\text{99,143,144}\). These mice do not display other features of the metabolic syndrome with the exception of one report suggesting intrahepatic insulin resistance \(^\text{144}\). Alternative approaches include murine genetic manipulation to increase appetite (Ob/Ob, Db/Db), increase lipid synthesis (Pepck-Nsrebp-1c, aP2-NSREBP-1c), reduced VLDL export (MTT\(^{-}\)) or reduced fatty acid oxidation (Ppara\(^{-}\), AOX\(^{-}\)) \(^\text{145}\). Each model represents part but not all of the human phenotype and thus the choice of intervention should be tailored to the individual research hypothesis.

*In vitro* culture of hepatocytes is most easily done using immortalised hepatocellular carcinoma lines from rat or human tumours such as HepG2, HepRG or Hep3B cells. It is known however that malignant transformation requires profound changes in metabolic function limiting the applicability of these lines in NAFLD research \(^\text{146}\). Alternatively, primary hepatocytes can be isolated from human explanted liver or mouse tissue although these have a highly limited lifespan and cannot be passaged or expanded in culture. Immortalisation of human hepatocytes by viral or small molecules is another potential source of *in vitro* hepatocytes but results in highly limited functionality and variable karyotype \(^\text{147}\). There has thus been intense interest in the generation of hepatocytes from pluripotent cell populations either embryonic in origin or reprogrammed from human somatic cells (induced pluripotent stem cells (iPSCs)) as potential source of both in vitro modelling material and bioartificial liver for transplantation \(^\text{147-150}\). Cells derived from these sources demonstrate an excellent cellular morphology, albumin production, expression of hepatocyte specific markers and gluconeogenesis and, although considered ‘foetal’ in maturity under current protocols, have enormous therapeutic potential \(^\text{148,151,152}\). As with murine modelling, recapitulating all featured of NAFLD *in vitro* remains challenging and has predominantly been characterised in
cancer cell lines. Steatosis is typically induced with long chain fatty acids (palmitate, oleate) or hyperglycaemia with some evidence of impaired insulin transduction\textsuperscript{106,153}. More recently, the use of high energy substrates in combination with medium chain fatty acids in HepG2 cells has also been reported to induced steatosis and mitochondrial dysfunction but not insulin resistance\textsuperscript{154}. Thus far there is no published literature characterising NAFLD models in ES derived hepatocyte populations.

1.8 Epigenetics and NAFLD

Increasing evidence suggests that epigenetic regulation plays an important role in both the susceptibility of patients to hepatic fat deposition and NAFLD progression. In this section I will discuss the definition and mechanisms of epigenetic regulation, evidence for a role in NAFLD pathogenesis and translational opportunities in this field.

1.8.1 The Epigenetic Concept

The precise definition of the term epigenetics is continually debated and has no absolute consensus. The developmental biologist Conrad Waddington coined the term ‘epigenotype’ in 1942 to describe ‘the processes involved in the mechanism by which the genes of the genotype bring about phenotypic effects’\textsuperscript{155}. Studying drosophila wing development, the original concept focussed on how isogenic cells are able to differentiate into multiple, highly varied cell types within the same complex organism. Essentially, Waddington was predicting the discovery of developmental gene network regulation. Today epigenetic research is a recognised sub-discipline of biology that encompasses a different remit. The epigenetic term now describes mechanisms that commit cells to and subsequently maintain cells within a specific lineage through mitotic and/or meiotic division. A widely quoted definition from Auther Riggs in 1996 defines epigenetics as “The study of changes in gene function that are mitotically and/or meiotically heritable and that do not entail a change in DNA sequence”\textsuperscript{156}. As such it also encompasses the maintenance of normal tissue function beyond embryogenesis and the adaptation of cells and cell lines to internal and external stimuli. Thus epigenetic mechanisms are postulated to contribute to the development of diverse transcriptional profiles that allow the maintenance of multiple tissue types in complex multicellular organisms and these profiles are subject to potential adaptation in response to changes in the intra and extracellular environment.
Mechanistically, putative epigenetic processes include chemical modification of DNA itself or the modification of scaffolding proteins governing its superstructure. More specifically, methylation of DNA cytosine residues and modification of nucleosome histone tails are thought to be functional epigenetic marks by a wide consensus. However recent developments, especially the discovery of methylcytosine demethylases, challenge this view as the system appears to be more dynamic than originally conceived. Other putative mechanisms that some deem as epigenetic in nature include RNA interference and higher order chromatin structure and nuclear organisation.\textsuperscript{157,158}

1.8.2 DNA Methylation - Distribution, Regulation and Functional relevance.

In the 1970s, it was first proposed that DNA methylation (5mC) may play a functional role in gene regulation. In vertebrates, DNA methylation represents the chemical addition of a methyl group to the 5\textsuperscript{th} carbon atom of the cytosine ring. It is the most intensively studied epigenetic mark and predominates on cytosine-phosphate-guanine (CpG) dinucleotide sequences in the 5\' to 3\' direction. Mammalian DNA is globally methylated with around 70 % of all genomic CpGs displaying the methyl mark.\textsuperscript{159}

DNA methylation patterns are most plastic during early embryogenesis with a wave of demethylation during zygotic cleavage followed by restoration of the ‘methylome’ by \textit{de novo} methylation shortly after embryo implantation.\textsuperscript{160-162} The addition and maintenance of DNA methylation is performed by DNA methyltransferases (DNMT) enzymes. DNMT1 is responsible for the maintenance of methylation whereas DNMT3a and DNMT3b perform \textit{de novo} methylation of unmodified bases.\textsuperscript{163} The importance of this process is highlighted by observing the phenotypic effect of DNMT deficiency and also loss of genes that bind methylcytosine \textit{in vivo}. Loss of DNMT1 is embryonic lethal in mice and Xenopus, correlating with global DNA demethylation.\textsuperscript{164-166} Loss of DNMT3a and DNMT3b in mouse ES cells and early embryos are also embryonic lethal with impaired \textit{de novo} methylation but no effect on maintenance of established imprinted methylation patterns.\textsuperscript{163} In humans, heterozygous mutations in DNMT3b are associated with ICF syndrome, characterised by immunodeficiency, centromeric defects and facial anomalies. Furthermore mutations in proteins which bind to methylated DNA \textit{in vivo} result in adverse effects perhaps best characterised in humans by MeCP2 mutations in human Rett syndrome.\textsuperscript{167}
As well as playing an essential role in embryonic development, 5mC has been proposed to be important in the maintenance of transcriptional silencing of imprinted genes and repeat elements and in the maintenance of the tissue specific transcriptome. Although most extensively described in the field of cancer biology, aberrant DNA methylation patterns are reported in a plethora of human diseases including mental health disorders, vascular disease, multiple sclerosis and autoimmunity. However, it is unclear to what extent these findings are causal rather than reflective of disease state and it is not known whether our understanding of 5mC regulation can be utilised for patient benefit through novel therapeutics or biomarkers to facilitate a personalised medicine approach.

1.8.3 CpG Islands as landmarks in mammalian genomes

The vertebrate genome is relatively devoid of CpG dinucleotides, containing approximately 25% the number that might be expected through random allocation. This may be due the mutagenic properties of methylated cytosines which can be deaminated to thymine and subsequently lost through rounds of replication. There are however, regions of high CpG density generally devoid of methylation known as CpG islands (CGIs) which may have been protected from deamination due to the maintenance of the demethylated state. The absolute criteria for defining a CGI varies, however a proposed pragmatic definition requires a G+C content greater than 55% with an observed/expected CpG ratio of 0.65. Typically around 1000bp in length, 72% of known genes contain a CGI region within or directly upstream of the annotated transcriptional start site (TSS) and these genes include almost all ubiquitously active genes as well as a proportion of developmental gene sets and tissue specific markers. The mechanism by which CGIs associate with transcription is unclear but may involve recruitment of histone methyltransferases and promotion of a permissive chromatin state.

Around half of the identified CGIs exist outside gene bodies and intragenic regions and are termed ‘orphan’. Unlike TSS associated CGIs, orphan CGIs are almost entirely methylated in contrast to the 3% of TSS CGIs which are significantly methylated. Although the exact function of orphan CGIs remains unclear, around 40% are associated with RNA markers of transcriptional initiation (RNAPII ChIP or Cap analysis gene expression analysis) indicating a potential role in the regulation of intragenic transcription. 5mC function in these regions has been described as ‘non-canonical’ as it appears to differ significantly from promoter CGI methylation (discussed below).
1.8.4 Methylation Within and Out with promoter Sequences

It is widely accepted that promoter CGI methylation associates with transcriptional silencing although the extent to which this is functional rather than mere association continues. Proposed mechanisms include the exclusion of transcription factors by methyl binding proteins, or the recruitment of co-repressors inducing a transcriptionally repressive state. Much of the evidence for this comes from studying the processes of de novo methylation in embryogenesis, X chromosome inactivation and mono-allelic repression of imprinted genes.

In early embryogenesis, pluripotency genes Oct3/4 and Nanog are silenced and concomitantly methylated, a finding that persists in all differentiated cells preventing re-activation of the transcript. During the random process of X chromosome inactivation, CGIs associated with inactivated genes on the inactivated X chromosome (X\textsubscript{i}) are hypermethylated in comparison to the activated X chromosome (X\textsubscript{a}) whereas active genes are hypomethylated on both X\textsubscript{i} and X\textsubscript{a}. It should be noted however that marsupial DNA undergoes Xi inactivation without accompanying CGI hypermethylation. DNMT1 knock out ES cells and embryos demonstrate defects in Xist expression and X-inactivation. Finally, imprinting control regions responsible for silencing autosomal ‘imprinted’ genes and gene clusters are typically differentially methylated in contrast to the majority of promoters in the mammalian genome.

Nevertheless, the relationship between methylation and transcriptional activity is not linear. During development, methylation of the Oct3/4 and Nanog genes typically occur after transcriptional silencing suggesting a maintenance role for methylation rather than a direct effect and demethylation of these genes does not lead to their reactivation in somatic cells. Similarly with X-inactivation, methylation can be temporally distinct from transcriptional silencing with some genes such as Hprt methylated late following transcriptional silencing and others such as Mtm1 more concurrently.

In differentiated tissue, outside the X\textsubscript{i} chromosome and imprinted gene sets, functional promoter methylation is relatively infrequent and related to CpG density. Illingworth et al performed an elegant sequence of experiments using CXXC (non-methylated CGI) and methyl binding domain (methylated CGI) affinity purification and sequencing to map CGIs across the mammalian genome in multiple tissues. This work demonstrated that only 6-8\% of human promoter CGIs are methylated to any significant extent, thus the significant majority of genes, whether active or silent, are unmethylated. This is supported by high resolution genome wide methylation profiling of chromosomes 6, 20, 22 and 21 demonstrating an
inverse correlation of transcriptional activity and methylation does occur, but in fewer than
10% of CGIs. Weber et al combined methyl DNA immunoprecipitation and hybridisation to
promoter arrays, with RNA ChIP to examine the relationship between methylation and
expression genome wide in human fibroblasts and germ cells. They again demonstrate that
an inverse correlation of transcriptional activity and promoter methylation does occur but in
less than 10% of CGIs and this is related to CpG density. Low CpG density promoters (CpG
ratio ~0.2 observed/expected) were generally methylated and demonstrate no correlation to
transcriptional activity. Promoters with high or intermediate (CpG ratio ~0.8 and ~0.5
observed/expected respectively) have a strong negative correlation with transcriptional
activity however high CpG dense promoters are almost exclusively unmethylated. It is
likely therefore that promoter methylation is an additional transcriptional regulatory
mechanism laid down following the establishment of a tissue specific transcriptome by other
means. Throughout adult life, the methylation profile is maintained to prevent aberrant
activation in all but a subset of intermediately CpG dense promoter regions in which
methylation is the key regulatory factor.

While this appears to be the case of somatic tissue, promoter methylation may play a more
prominent regulatory role within the germline. Hackett et al identified a core of germline
specific genes which appear to be entirely dependent on methylation for the initiation and
maintenance of a gene silencing. Further, this core set is enriched for genes with a
functional role in suppressing repeat elements that are subsequently demethylated and
activated thought development and as such may act in a ‘genome defence’ capacity.

As stated, 50% of CGIs lie outside the promoter in either intragenic or intergenic regions.
The function of these CGIs is unclear, however there does appear to be a clear association
between genic 5mC levels and transcriptional activity in mammals. These results are
inconsistent and are complicated by the presence of another cytosine species (5’hydroxymethylectosine, 5hmC) in gene bodies, which confounds any technology requiring
bisulphite DNA treatment. If genuine, genic 5mC may prevent aberrant genic
transcriptional initiation or facilitate transcription by preventing binding of proteins such as
CTCF. Methylation of CGIs in other regions such as CpG shores, repeat elements and
distal enhancers may also have key regulatory roles but are beyond the remit of this chapter.
5hmC and Tet proteins and DNA demethylation pathways

5hmC was originally described in T phage over 50 years ago but has since been largely ignored due to an inability to isolate the base from vertebrate DNA. In 2009 Kriaucionis et al unexpectedly identified significant amounts of 5hmC in Purkinje neurones using 2D thin layer chromatography, and validated this using high pressure liquid chromatography and mass spectrometry. At approximately the same time, Tahiliani and colleagues utilised comparative genomics to deduce that Ten eleven translocase 1 (Tet1), a fusion protein in myeloid malignancy, could oxidise 5mC to 5hmC in an iron and α-ketoglutarate dependant manner.

Thus far, three Tet isoforms (1-3) have been identified in mammals each with a distinct tissue distribution. Each isoform contains a carboxy-terminal catalytic domain with a cysteine rich core typical of Fe\(^{2+}\) and 2-oxoglutarate-dependent dioxygenases. Tet1 and Tet3 contain an N-terminal CXXC domain typical of chromatin-associated proteins however Tet2 became detached from its CXXC domain during an evolutionary chromosomal inversion event. Interestingly, the isolated Tet2 CXXC domain became the separate gene IDAX (also known as CXXC4), which has been reported to target Tet2 for caspase mediated destruction in ESCs and human myeloid cancer cells.

In addition to the generation of 5hmC, Tet proteins can further oxidise cytosine bases to 5-carboxylcytosine (5caC) and 5-formylcytosine (5fC). 5caC and 5fC are present in almost undetectable amounts in ESCs and can be removed from the genome by thymidine glycosylase (TDG), resulting in the regeneration of unmodified cytosine though base excision repair (BER) mechanisms. Thus, the predominant theory is that through successive oxidation steps terminating in BER, cytosine residues can be actively demethylated by Tet enzymes. Alternative possible demethylation pathways include passive demethylation over successive cell divisions. The relative inability of DNMT1 and its binding partner UHRF-1 to bind hemi-hydroxymethylated DNA results in disruption of 5mC maintenance by DNMT1 of the newly synthesised daughter strand and progressive dilution of 5mC levels. Passive mechanisms do not however account for the replication independent demethylation observed in the paternal prozygote and primordial germ cells suggesting that an active enzymatic process must exist. Finally, active conversion of 5hmC to cytosine by the AID/APOBEC deaminase family has been proposed but remains controversial.
The precise biological function of both the 5hmC modification and the Tet dioxygenase enzymes is as yet unclear. Evidence to support a role in transcriptional regulation, cell fate and thus cell phenotype comes from three sources.

1) GWAS studies examining the association of gene mutations resulting directly or indirectly from loss of Tet function.
2) Tet isoform deficiency characterisation in mice and cell culture models.
3) Global and high-resolution 5hmC profiling in distinct healthy cell populations, in the context of 1) and 2) and in comparison with ChIP analysis of histone marks and transcription factor binding sites.

Figure 1.5: Proposed role for Tet proteins and 5hmC in the active (solid arrows) and passive (dashed arrows) demethylation of DNA resulting in restoration of unmodified cytosine. TDG = Thymidine glycosylase, BER = base excision repair. Reproduced with permission from Haung and Rao 216.
Disease associated mutations

Mutations in Tet enzymes have been reported in a variety of solid and haematological malignancies. Tet2 mutations are common in premalignant and malignant myeloid haematological conditions, with over 700 mutations now identified and are associated with a worse prognostic outcome. These are typically loss of function mutations centred on the catalytic domain. Within solid tumours, Tet mutations are significantly less common but have been reported in endometrial, colon, bladder and lung cancers. Tet1 expression is reduced in human breast, colon and hepatocellular carcinoma, renal and gastric tumours and associates with poor prognosis. Strikingly, re-expression of Tet1 in mouse colon cancer xenografts suppresses tumour progression and overexpression of Tet1 induces apoptosis in multiple cancer cell lines highlighting the importance of the Tet1 tumour suppressor role.

Indirectly, Tet proteins can be affected though the accumulation of ‘oncometabolites’. Somatic mutations in the genes Isocitrate dehydrogenase 1 (Idh1) and Isocitrate dehydrogenase 2 (Idh2) are common in human glioma and acute myeloid leukaemias, the mutant isoform generating 2-hydroxyglutarate, a competitive inhibitor of Tet function in vitro. Human loss of function mutations in the citric acid cycle (TCA) enzymes succinate dehydrogenase (Sdh) and Fumarate Hydratase (Fh) lead to cellular accumulation of the respective substrates succinate and fumarate which may in turn competitively inhibit Tet activity along with other α-KG dependant dioxygenases. The observed increase in renal cell carcinomas and phaeochromocytomas/paragangliomas in these patients may in part be a consequence of these effects.

Tet deficiency phenotypes

The effects of loss of Tet enzymes in mice and ESCs support a role for the generation of 5hmC in vitro in the control of cell fate and embryonic development. ESCs deficient in Tet1 show a mild reduction in 5hmC levels and transcriptional dysregulation in developmental gene sets but maintain pluripotency and are able to support the birth of live mice. Tet1−/− mice vary significantly in birthweight and although they are smaller on average, male mice recover to comparable size at 12 weeks of age. Interestingly Tet1−/− mice are exhibit grossly normal postnatal development and are fertile with crossed homozygous mice able to generate small but viable litters suggesting moderately preserved gametogenesis. Consistent with the relatively high levels of Tet1 expression in brain, three groups report an adverse cognitive...
phenotype in Tet1 deficiency with impaired memory function\textsuperscript{235–237}, suppressed hippocampal development\textsuperscript{237} and promoter hypermethylation in key neurodevelopmental gene sets\textsuperscript{235,237}. Furthermore, with aging Tet1\textsuperscript{-/-} mice display an adverse tendency to develop B cell lymphoma subtypes\textsuperscript{238}.

Loss of Tet2 both constitutively\textsuperscript{239} and in bone marrow only\textsuperscript{240} confers expansion of the haematopoietic compartment, with myeloid malignant transformation consistent with observations in human disease. In young, Tet2\textsuperscript{-/-} deficient mice, acute concurrent bone marrow specific deletion of Tet3 incurs near total loss of 5hmC with the rapid development of an aggressive acute myeloid leukaemia. Remarkably, constitutive double knock out Tet1\textsuperscript{-/-} and Tet2\textsuperscript{-/-} mice can be viable, fertile and grossly normal in around 40% of cases with the remainder exhibiting high levels of mid gestational abnormalities and death within 2 days of birth\textsuperscript{241}. Analysis of Tet1\textsuperscript{-/-} Tet2\textsuperscript{-/-} combined KO mice demonstrate hypermethylation in imprinted gene sets with reduction but not absence in 5hmC levels in multiple tissues suggesting a role for Tet3 in generating 5hmC during development. Tet3 is enriched in the paternal prozygote in early embryogenesis, is required for demethylation of the paternal genome constitutive mutation of Tet3\textsuperscript{-/-} subsequently embryonic lethal\textsuperscript{242}.

**Tet proteins may intersect metabolism and the epigenome**

Tet protein activity is highly sensitive to fluctuations in cofactors and chelating ions including Fe\textsuperscript{2+}, α-ketoglutarate, Zn and ascorbic acid\textsuperscript{243–245}. Interestingly, Tet function is intrinsically linked to glucose metabolism both through the competitive presence of stimulatory and inhibitory cofactors and secondly through post translational modification by glucose metabolites. The TCA cycle metabolite α-ketoglutarate is an essential co-factor for normal Tet function\textsuperscript{230,246}, however others (succinate, fumarate and 2-hydroxyglutarate) are competitive inhibitors of Tet activity\textsuperscript{247,248}. The relative availability of these metabolites in the context of metabolic dysregulation may therefore directly affect Tet function. Secondly, Tets are post-translationally modified by the addition of N-acetylglucosamine (GlcNAc) to residues in Tet regulatory regions by the enzyme O-linked N-acetylglucosaminyltransferase (OGT)\textsuperscript{249,250}. The uridine diphosphate-GlcNAc substrate is generated by the nutrient sensing hexokinase pathway, which is reliant on the availability of other key cellular energy intermediates ATP, glucose and Acetyl-CoA\textsuperscript{251}. Taken together, it has been proposed that Tet proteins may act as a metabolic ‘sensor’ regulating cellular activity in response to changes in energy supply and demand\textsuperscript{252}.
5hmC distribution has been examined at base pair or high resolution in human \(^{253,254}\) and mouse \(^{233,255,256}\) ES cells, mouse liver \(^{257}\), human foetal and adult liver \(^{258}\), mouse adipocyte cell lines \(^{259}\), mouse frontal cortex \(^{260}\) and hippocampus \(^{261}\). Interestingly, 5hmC is not present in DNMT triple KO cells confirming that 5mC is almost certainly the source of all 5hmC \(^{262}\). 5hmC profiles differ markedly between ESCs and differentiated tissue \(^{253,255,257}\) and differentiated tissues differ markedly from each other \(^{263}\). Indeed 5hmC profiles alone are adequate to independently cluster DNA samples into their tissue of origin suggesting 5hmC patterns are a sensitive marker of, and may contribute to, cell phenotype \(^{263}\).

In both human and mouse ESCs, there is a high degree of consensus of 5hmC enriched regions between studies and profiling techniques \(^{253,255,256,262}\). 5hmC is widely present across gene-rich regions but relatively devoid over intragenic regions and repeat elements. In both ESC and all differentiated tissues studies, 5hmC is enriched within intragenic regions and in the overwhelming majority of cases correlates highly with transcriptional activity, particularly around exon borders, and associates strongly with the H3K4me3 and other active histone modifications \(^{233,253–255,264,265}\). The mechanism for this striking correlation is unclear but suggests a functional role in facilitating transcript elongation. Work in the Meehan laboratory recently demonstrated a dynamic staged enrichment of hepatic genic/proximal 5hmC levels in upregulated genes in response to the non-genotoxic carcinogen phenobarbital \(^{264,266}\). These landmark studies not only revealed the plasticity of the hydroxymethylome in differentiated tissue but also introduced 5hmC profiling as possible future biomarker of disease state. In addition to gene bodies, 5hmC is generally enriched at enhancer regions as determined by H3K4me1 and H3K27ac histone marks and these regulatory regions become rapidly more enriched as ESCs differentiate \(^{254,267–269}\). In contrast to 5mC, 5hmC is typically enriched within CpG rich-gene promoter sequences and subsets of promoter transcription factor binding sites with almost total depletion of 5hmC directly over transcriptional start sites \(^{253,254,256,270–272}\). In contrast to gene body regions, high levels of promoter 5hmC enrichment is more common in genes with low levels of gene expression \(^{233,255,262,265}\).

Attempts to identify 5hmC binding partners continue. Utilising primer sequences of varying CpG density and cytosine species as ‘bait’, Iurlaro et al found few 5hmC protein interactions, although these include RPL26, a P53 regulator associated with the congenital red cell aplasia syndrome Diamond-Blackfan anaemia \(^{273,274}\). One study suggested that the methyl binding
protein MeCP2 binds 5hmC with a biologically significant affinity and that this is abolished with the R133C mutation frequently seen in Rett syndrome\textsuperscript{275}. However \textit{in vitro} studies of the MeCP2 MBD suggests reduced 5hmC binding affinity of 5hmC compared to 5mC by an order of magnitude\textsuperscript{276}.

In summary, 5hmC exists as a chemical DNA modification within in the mammalian genome at significant levels. Its distribution is cell type specific, precisely maintained and genic levels display a striking correlation with transcriptional activity\textsuperscript{266,277,278}. Thus it is reasonable to suggest that 5hmC has a distinct unique biological function that is yet to be elucidated. Proposed mechanisms for this include the following (Figure 1.6).
Figure 1.6: Possible mechanisms of 5hmC function in transcriptional regulation. A) MBD containing proteins are able to specifically recognise 5mC and recruit transcriptionally repressive histone methyltransferases or acetylases. Reports of significantly reduced affinity of MBDs to 5hmC suggest exclusion of these proteins may promote a more transcriptionally permissive state. Alternatively, conversion of 5mC to 5hmC may allow TF binding and transcriptional initiation. B) Active demethylation may directly relieve nucleosome compaction and allow promoter TF access. C) Replacing methylation in gene bodies may allow uninterrupted transcript elongation by RNA polymerase enzymes perhaps by reducing DNA duplex melting temperature. Adapted from Pastor et al and Laird et al.
1.8.7 Cytosine Modifications in NAFLD

A putative role for the regulation of cytosine modifications in NAFLD pathogenesis progression is emerging from both genome wide and targeted examination of methylation status in human subjects summarised below.

<table>
<thead>
<tr>
<th>Study</th>
<th>Disease states</th>
</tr>
</thead>
<tbody>
<tr>
<td>Murphy et al 282</td>
<td>NAFLD mild fibrosis (n = 52) vs advanced fibrosis (n = 38)</td>
</tr>
<tr>
<td>Ahrens et al 283</td>
<td>Control (n = 28), obese normal histology (n=27), steatosis (n = 22),</td>
</tr>
<tr>
<td></td>
<td>NASH (n = 25)</td>
</tr>
<tr>
<td></td>
<td>Post bariatric surgery samples (obese normal histology (n=7),</td>
</tr>
<tr>
<td></td>
<td>steatosis (n=10) NASH (n=6))</td>
</tr>
<tr>
<td>Nilsson et al 284</td>
<td>normoglycaemic obese (n=60) T2DM obese (n=35)</td>
</tr>
<tr>
<td>Horvath et al 285</td>
<td>Control versus obese with normal histology (n = 141 total)</td>
</tr>
</tbody>
</table>

Table 1.1 Summary of published data sets of high throughput methylation studies in obesity and NAFLD liver samples.

Murphy et al examined single based resolution methylation data between patients with mild vs advanced NAFLD based on histological fibrosis score 282. They report that the majority of differentially methylated sites in advanced NAFLD are hypo- rather than hypermethylated and this was particularly evident over CGI shores. Furthermore they observed a correlation between hypomethylation and increased gene expression in a number of pathologically relevant genes and genes sets including inflammatory chemokines, inflammatory response proteins and a variety of genes involved in fibrosis and tissue repair. Simultaneously Ahrens et al published a similar study but in less severe, minimally fibrotic disease examining a disease spectrum from obese patients with normal histology to NASH with significant inflammation and steatosis but with a fibrosis score averaging only 1 283. Crucially, a subset of this population at each NAFLD stage was also re-biopsied and profiled 6-9 months later following bariatric surgery and dramatic weight loss. They report a relatively small number of CpGs (467 of 450000) with a significant differential methylation in their cohort but do demonstrate that these sites are sufficient to separate normal liver histology from NASH by clustering analysis. Of the 274 genes annotated to those sites only 9 genes exhibited an associated differential expression and this change was only determined when including a CpG methylation difference of 5% and an expression change of log2 (0.2), the biological
significance of which may be questioned. However, despite small changes there was a striking hierarchical correlation between NAFLD stages and identified gene sets. Ahrens et al also state that the methylation changes are reversible following bariatric surgery. It should be noted however of the 113 methylation sites that change significantly following bariatric surgery, only 4 of these sites were differentially methylated between NAFLD subtypes in the original analysis suggesting a more tenuous relationship. Horvarth et al interrogated the same data set utilising an algorithm known as the ‘epigenetic clock’, an analysis of DNA CpG methylation which correlates with chronological age in multiple tissues. They describe that livers from obese individuals exhibit a striking acceleration in ‘epigenetic aging’ such that the methylation state of obese patients was similar to lean patients of more advanced age. Nilsson et al performed a near identical experiment in obese patients with T2DM versus normoglycaemic obese patients. In a larger cohort (control = 60, T2DM = 35) they again found small number of CpG sites differentially methylated (251 of 450000 CpGs), which through statistical correction and comparison with the above Ahrens et al study they attribute to specifically T2DM rather than NAFLD per se. Consistent with the findings of Murphy et al, this study found that the majority of differentially methylated regions were hypomethylated however, there was no clear association with transcriptional change when examining genes in close proximity to these CpG sites with equal number of genes showing direct and inverse correlations with methylation change. Interestingly the authors did describe a reduction in folate levels in a subset of their study population and suggest relative folate deficiency as contributory to their methylome findings.

The summation of these in depth analyses of human methylation levels is that site-specific methylation appears to change in diseases of overnutrition although these changes are associated with very few alterations in transcriptional state. Furthermore, the exact sites are not consistent between data sets perhaps due to differing hepatic pathologies examined. These studies are limited in that bisulphite-associated technology is unable to differentiate between 5mC and 5hmC species. In addition, the heterogeneity of the human population with respect to diet, pharmacological exposure and geographic location may render these studies underpowered. Finally, this human work is purely observational and a functional role for cytosine modifications in metabolic liver disease has yet to be examined.
Summary, hypothesis and thesis aims:

Despite the difficulties in accurate quantification, it is indisputable that the prevalence of NAFLD is accelerating rapidly in both the developed and developing world. Epidemiological data clearly displaying the associated negative health outcomes in patients with NAFLD demonstrates the profound public health issue NAFLD has become, the burden of which will continue to grow. Our understanding of NAFLD pathology has continued to expand through a combination of human observational studies with mouse dietary and transgenic data and *in vitro* cell culture work. These suggest a complex interplay between the liver and other metabolically active tissues. Despite this there are currently no effective pharmacological therapies to limit NAFLD progression and thus novel therapeutic targets are required.

The observation that DNA methylation changes in human NAFLD occur consistently and appear to be dynamic, coupled with the more recent finding that Tet enzymes respond to fluctuations in energy metabolites, has lead us to hypothesis that:

1) Tet function may play an important role in NAFLD pathogenesis or progression.
2) This may be due to fluctuations in key TCA cycle metabolites
3) Hydroxymethylation profiling may provide a useful biomarker of disease state.

Therefore, the aims of my thesis studies are:

1. To assess and describe the effect of overnutrition on Tet enzyme expression and activity in mouse and human models. This will be done by assessing perturbations in Tet transcript level and 5hmC profiles following high fat diet feeding.
2. To examine fluctuations in TCA cycle metabolites in the context of any observed effect on Tet activity.
3. To determine if Tet1 function is important in disease pathogenesis by examining the response of Tet1 deficient animals to obesogenic dietary intervention.
CHAPTER 2 – MATERIALS AND METHODS

2.1 Materials

All chemicals were purchased from Sigma (Dorset, UK) unless specifically stated otherwise. All sterile needles (Microlance) and syringes (Plastipak) were from Becton Dickinson (Oxford, UK).

Cells culture dishes and sterile pipettes were by Corning (Tewksbury, USA)

Room temperature was 22+/− 2 °C

2.2 Buffers And Solutions

**DEPC treated water:** Distilled water (1L was mixed with 1ml diethylpyrocarbonate (DEPC) (1% v/v). The solution was left to stand for 12 hours prior to autoclaving.

**10x Western Blot Running Buffer:** Tris Base 121g, sodium dodecyl sulphate 10g, HEPES 239g. Adjusted to 1L and autoclaved.

**Western Blot Transfer Buffer:** Tris Base 3g, Glycine 14.4g, 100% methanol 100ml. Adjusted to 1 litre with ddH2O.

**CHAOS Buffer:** guanidinium thiocyanate 4.5M, N-lauroylsarcosine, 50mM EDTA pH8, 0.2% antifoam A, 01.M β-mercaptoethanol.

**RIPA lysis buffer with protease inhibitors:** 150mM sodium chloride, 1x triton x-100, 0.5% sodium deoxycholate, 0.1% SDS, 50mM Tris pH8.0. Dissolved in deionised water. 10ml aliquots were stored at -20°C. Roche Complete protease inhibitor cocktail was added as per manufacturer instructions.

**10x TBE buffer:** Tris Base 108g, boric acid 55g and 40mls of 0.5M EDTA solution (pH 8.0). Adjusted to 1 litre total volume in distilled water and autoclaved before use. Typically used at 0.5X concentration.

**1M Na-Phosphate (mono and dibasic):** 1 molar solutions of mono and dibasic sodium phosphate salt were made up separately in distilled water, combined in a 39:61 ratio and adjusted to pH 7.0 with 4M NaOH.
**DNA Immunoprecipitation Buffer**: 100 mM Na-Phosphate pH 7.0 (mono and dibasic), 1.4 M NaCl, 0.5 % Triton X-100, made up to 10 mls in distilled water fresh prior to each procedure and 0.22μm sterile filtered.

**Proteinase K Digestion Buffer**: 50mM Tris – HCl pH 8.0, 10mM EDTA, 0.5% SDS, made up to 10ml prior to each procedure in distilled H₂O and 0.22μm sterile filtered.

**LPO Treatment**: Sodium L Lactate (Sigma cat. number L7022) was made up in 1ml aliquots of 1M solution and stored at -20°C. On the day of use, 11mg of sodium pyruvate was added to a thawed 1ml aliquot and kept on ice. Octanoic Acid (Sigma cat. number C2875) 865.2mg was added to 60ml of double distilled water and adjusted to pH 7.0-7.9 with NaOH 4M solution with constant stirring. The solution clears at this point and is stored in aliquots at -20°C. The required volume of each solution was then added to media 0.22μM filtered and added to the culture dish. **LCMS Extraction Buffer** 50% methanol (Sigma 34966-1L), 30% acetonitrile (Sigma 34967-250ML), 20% water (Sigma 39253-1L-R).

**2.3 Animal Models**

2.3.1 Animal Diet and Husbandry

All experiments were carried out under UK Home Office license (PPL (70/7874) C57BL/6J mice (Charles River, Tranent, UK) or B6;129S4-Tet1tm1.1Jae/J mice (Jackson Laboratories, Maine, USA) were maintained under controlled conditions. A 12 hour light cycle (07.00h to 19.00h) and 12 hour dark cycle was implemented throughout. The temperature was maintained at 22°C +/- 2°C. Diets were obtained from Research Diets, New Brunswick, NJ, USA (D12328, D12331) or DYETS.inc Bethlehem, US (control, CDD, MCDD). Compositions of each diet can be seen in table (table 2.3.1).
Table 2.3.1 Composition of rodent diets

Order reference codes are in brackets.

<table>
<thead>
<tr>
<th></th>
<th>Control Diet (518574)</th>
<th>Choline Deficient Diet (518753)</th>
<th>Methionine Choline Deficient Diet (518810)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kcal/gram</td>
<td>4.3</td>
<td>4.3</td>
<td>4.3</td>
</tr>
<tr>
<td>Kcal%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein (as L-amino acids)</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Carbohydrate</td>
<td>55</td>
<td>55</td>
<td>55</td>
</tr>
<tr>
<td>Fat</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Choline Bitartate (g/kg)</td>
<td>14.48</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>L-Methionine (g/kg)</td>
<td>1.7</td>
<td>1.7</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Control Diet D12328</th>
<th>High Fat Diet D12331</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kcal/gram</td>
<td>4.07</td>
<td>5.56</td>
</tr>
<tr>
<td>Kcal%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>16.4</td>
<td>16.4</td>
</tr>
<tr>
<td>Carbohydrate</td>
<td>73.1</td>
<td>25.5</td>
</tr>
<tr>
<td>Fat</td>
<td>10.5</td>
<td>58.0</td>
</tr>
<tr>
<td>Sucrose (g/kg)</td>
<td>835</td>
<td>175</td>
</tr>
<tr>
<td>Maltodextrin 10 (g/kg)</td>
<td>170</td>
<td>170</td>
</tr>
<tr>
<td>Coconut oil, Hydrogenated (g/kg)</td>
<td>40</td>
<td>333.5</td>
</tr>
<tr>
<td>Soyabean oil (g/kg)</td>
<td>25</td>
<td>25</td>
</tr>
</tbody>
</table>

Animals were maintained on diet for the duration of the experiment as described in each chapter.
2.3.2 Venesection

Animals were allowed to move freely on the top of the cage. The distal 1mm of the tail was snipped off with fine scissors and blood massaged from the wound. Blood was either applied directly to a glucometer strip (Accu-check, Roche, Burgess Hill UK) or collected by capillary action into an EDTA tube (Microvette® CB 300 K2E; Sarstedt, Numbrecht, Germany) to a volume of 25μl. Tubes were spun at 5000rpm for 10 minutes and the supernatant removed and stored at -80°C.

2.3.3 Oral glucose tolerance testing and serum measurements

Following 6 hour fast from 0800 to 1400 hours, 2g/kg sterile filtered glucose solution (40%w/v) was administered by intraperitoneal injection using an insulin syringe (BD Medical, Oxford, UK). Venesection was by single tail snip and blood was collected at 0, 15, 30 60 and 90 minute time points. Mice were then returned to the social group with access to diet and water. Serum insulin concentration was measured using the Mercodia mouse insulin ELISA kit (Mercodia, Uppsala, Sweden). 5μl serum was used per well. Whole blood glucose was measured using the Accu-check glucose meter (Roche, Burgess Hill) for studies with C57Bl/6J mice on high fat diet and with 2 μl of serum using the ‘glucose infinity assay’ (Thermoscientific, UK) in Tet1 knock out studies.

2.3.4 Insulin Tolerance Testing

Mice were fasted from 0800 to 1200 followed by intraperitoneal injection of Actrapid insulin (Novo Nordisk, Gatwick, UK) at a concentration of 1mIU per/g diluted in 0.9% saline solution. Blood glucose measurements were made at baseline, 15, 30 and 60 minute time points.

2.3.5 CO2 Asphyxiation

Mice were placed in a sealed box with increasing CO2 flow until breathing stopped followed by removal and cervical dislocation. Death was confirmed by a lack of corneal reflex.

2.3.6 Tissue Harvesting

For consistency, timing of animal cull was standardised for each experiment. In C57Bl/6 high fat diet model experiments, mice were all culled between 0800 and 1030 hours. The Tet1 knock out mice were culled fasted at 1400 to 1600 hours.
Immediately after culling, mice were dissected rapidly to preserve RNA integrity. Blood was harvested by cardiac puncture. Liver, kidney, spleen and adipose depots (mesenteric, right retroperitoneal, right flank subcutaneous and right epididymal) and right quadriceps muscle were dissected, weighed where appropriate, snap frozen on dry ice and stored at -80°C. Two aliquots of the right lobe of liver were frozen separately for DNA/RNA extraction. The left lobe of the liver was fixed in 10% Formalin solution for 24 hours.

2.3.7 Ten Eleven Translocase 1 Global Knock Out Mouse Breeding

B6;129S4-Tet1tm1.1Jae/J heterozygote mice (4 pairs) were purchased from Jackson Laboratories, Maine, US. Heterozygote mice were interbred to generate homozygous knock out and wild type progeny.

2.3.8 Hepatic Triglyceride concentration.

Hepatic triglyceride content was determined using the triglyceride colorimetric assay kit (Cayman chemical, UK). Approximately 200mg of frozen hepatic tissue was homogenized in 1mL of the diluted Standard Diluent with protease inhibitor and centrifuged for 10 min at 10000g at 4 degrees. A 1:10 dilution of the supernatant with the Standard diluent was made and 10uL was loaded in duplicate to the kit assay dish. To this was added 150uL of the diluted enzyme buffer and measured the absorbance at 530nm.

2.3.9 S’Adenomethionine Quantification by MALDI

Matrix-assisted laser desorption/ionization (MALDI) was outsourced to the BHF core mass spectrometry unit at Queens Medical Research Institute, University of Edinburgh. Briefly, tissue embedded in gelatine solution 10%-w/v and the cryostat (Leica Microsystems Inc, Bannockburn, IL, USA) was used to slice cryosections (10 µm). Liver sections were thaw mounted onto a conductive indium tin-oxide (ITO)-coated glass slides (Bruker Daltonics, Bremen, GmbH). Tissue sections were then stored in a vacuum desiccator (RT, 1h) and then at -80°C. SAMe quantification was performed using a 12T SolariX MALDI-FTICR-MS (Bruker Daltonics, MA, US) employing a Smartbeam 1 kHz laser, with instrument control using SolariX control v1.5.0 (build 42.8), Hystar 3.4 (build 8) and FlexImaging version 3.0 (build 42).
2.3.10 Indirect gas calorimetry

Animals were single housed under standard conditions from weaning (4 weeks). Food intake was measured weekly for 6 weeks. At baseline and following a 7 week dietary intervention, animals were transferred to the TSE Phenomaster System (Bad Homberg, Germany) for 5 days. Animals were allowed to acclimatise for 72 hours before indirect gas calorimetry and total energy expenditure was measured over a 24-hour period. Data were analyses on the TSE Phenomaster software version 5.0.8 (Bad Homberg, Germany).

2.4 HISTOLOGY

Sections of fresh liver were fixed in 10% v/v formalin solution for 24 hours prior to transfer to 70% ethanol. They were then embedded in paraffin solution. Paraffin sections (5μm approximately) were cut using a microtome and stained with Picosirus red, and Haematoxylin and Eosin by the University of Edinburgh Histology department as follows:

2.4.1 Tinctoral Stains

**Haematoxylin and Eosin**

Haematoxylin and eosin (H&E) staining allows visualisation of the hepatic architecture, staining nuclei blue and matrix pink. Liver sections were cut as above followed by dewaxing in xylene for 10 minutes and subsequently rehydrated through incremental alcohol solutions for 5 minutes each (100%, 100%, 70%, 50% v/v ethanol). Sections were then dipped in Harris Haematoxylin for 30 seconds and subsequently washed in running tap water (1 minute) and then Scotts tap water (5 seconds). After a further wash in tap water, sections were then washed with acid-alcohol for 3 seconds and stained with Eosin. Sections were then dehydrated in increasing concentrations of alcohols (reverse of above), cleared in xylene and mounted using DPX.

**Picosirus Red (PSR)**

Picosirius red staining allows visualisation of new collagen fibrils and therefore an estimation of hepatic fibrosis. Dewaxing and rehydrating of liver sections was performed as per H&E staining. Sections were then placed in Picosirius red/fast green solution for two hours, washed in tap water and subsequently rehydrated in increasing alcohols, cleared in xylene and fixed as before.
2.4.2 Histological analysis and NAFLD scoring

NAFLD scoring was performed blinded by both the author and a consultant pathologist using a validated NAFLD scoring system \(^{24}\). For quantification of picosirius red stain, slides were scanned on the Axioscan.Z1 (Zeiss, Cambridge, UK) slide scanner using default settings. Image analysis was performed on ImageJ software (http://imagej.nih.gov/ij/).

2.5 Human Embryonic Stem Cell Derived Hepatocytes.

2.5.1 Materials and Media

All media products were purchased from Life Technologies, Paisley, UK unless stated otherwise.

**Plastic Culture Materials**

Cell culture dishes and sterile pipettes were by Corning (Tewksbury, USA). hESCs were cultured in 6 well plate format. Differentiation and experimentation was carried out in either 6, 12 or 24 well format.

**Preparation of matrigel plates**

1ml aliquot of matrigel (Corning, Tewksbury, USA) was allowed to thaw at 4°C overnight. It was then diluted with the addition of 17mls of cold KO-DMEM media (using ice-cold pipettes and 0.5-1ml added to each well). Plates were left at 4°C overnight and then brought to room temperature for 2 hours before use.

**Conditioned Media**

Conditioned Media with basic human fibroblast growth factor was added on the day of use to a final concentration of 10mg/ml (both R&D systems, Abingdon, UK).

**hESC maintenance media preparation**

Human Pluripotent Stem cell media – mTeSR1 (Stem Cell Technologies, Cambridge, UK) was defrosted, combined with mTeSR supplement and warmed to room temperature before use.
Primming Medium preparation
RPMI 1640 containing 1x B27 supplement, penicillin 50U/ml penicillin and 50mg/ml streptomycin was combined and kept at 4°C. On the day of use 50ng/ml Wnt3a (R&D systems, Abingdon, UK) and 100ng/ml Activin A (Peprotech, New Jersey, USA) were added prior to administration to the culture layer.

Hepatic Differentiation Medium preparation
KO-DMEM with 20% KO-SR serum replacement, Glutamine (1mM), MEM non-essential amino acids (1%), β-mercaptoethanol 0.1 mM (all from Life Technologies), 1% dimethyl sulfoxide 1% (DMSO, Sigma). Solutions were combined, sterile filtered and kept at 4°C.

Hepatic Maturation Media preparation
Maturation media was prepared by combining HepatoZYME (Thermo Fisher scientific) containing 1% Glutamax supplemented with 10 μM hydrocortisone 21-hemisuccinate (Sigma-Aldrich), 10 ng/ml hepatocyte growth factor (Peprotech, Rocky Hill, NJ, USA), and 20 ng/ml oncostatin M (Peprotech). Products were mixed and warmed to room temperature before use.

LPO High energy substrate
1M Lactate (Sigma L7022) was made up and stored in 1ml aliquots at -20°C. Sodium pyruvate (Sigma p5280) was added fresh to the media on day of use. Octanoic acid (Sigma C2875) was added to water to make 0.1M stock and the pH was adjusted to 7-8 with 4M NaOH followed by storage at -20°C. The required amount of each reagent was added to culture media and 0.22μm filtered prior to covering cells. Cells were incubated with LPO for 48 hours.

2.5.2 Hepatocyte Differentiation Protocol
Human H9 ESCs cells were resuscitated from frozen in warmed conditioned media and maintained in co-culture with an irradiated mouse embryonic fibroblast (MEF) feeder cell layer (vhBIO, 1 vial per 2 wells of a six well culture plate). Cells were maintained at 37°C and 5% CO2. Prior to differentiation, cells were gradually transitioned into mTeSR1 media during which feeder cells were no longer used. Cells were fed daily with 2 mls of media and split in a 1:3 ratio when achieving 70% confluence.
Differentiation of ESCs was performed as previously described. H9 human embryonic stem cells were maintained in mTESR1 and split 1:3 onto matrigel coated plates. After 24 to 48 hours at a confluence of 40-60% the media was changed as below to initiate the differentiation process (Fig 2.5.1). Cells were exposed to LPO between days 20 and 22.

<table>
<thead>
<tr>
<th>DIFFERENTIATION DAY</th>
<th>MEDIA</th>
<th>Conditioned Media MEFs</th>
<th>mTeSR1</th>
<th>Priming Media</th>
<th>Differentiation Media</th>
<th>Maturation Media</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-3</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>48</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>3-9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2.5.1 Schematic representation of the the maintenance and differentiation of hESCs into hepatocytes.

2.5.3 Assessment of hepatocyte differentiation

Maturation of hepatocytes was assessed by morphological assessment (Figure 5.1), relative mRNA level of ESC specific (OCT3/4, NANOG) and hepatocyte specific (ALP, HNF4A) markers (2.6, 2.8.2, Figure 5.1 D). In addition, Cyp1A2 and Cyp3A4 activity and Albumin and al secretion into medium was also assessed.

Cytochrome P450 activity

CYP3A and CYP1A2 activity was measured on day 21 of maturation using pGlo technology (Promega) and carried out according to the manufacturer’s instructions. Briefly, cells were incubated with 20µl (cyp1A2) or 40µl (cyp3A4) luciferase P450 substrate for 5 hours at 37°C before collection of media. 50µl of media was then incubated with 50µl of luciferase detection reagent for 20 minutes and luminescence measured on luminometer (Promega). Media in a
well with no cells was used as a negative control and subtracted from final readings. Activity is expressed as relative light units per millilitre of media per mg of protein (Pierce BCA assay).

**Albumin and αfetoprotein**

Albumin and αfetoprotein (AFP) production was assessed at day 21 of maturation using commercially available ELISA kits (Alpha Diagnostic International). Samples were run in triplicate and measured on a FLUOSTar Omega multi-mode microplate reader (BMG Labtech). Protein production was expressed as nanogram of protein per milliliter of medium per cm² of convalescent cells.

2.5.4 Fluorescence microscopy

**BODIPY 493/503 lipid marker.**

BoDIPY 493/503 (Life Technologies) was dissolved in a 1:1 ethanol:DMSO mixed solvent solution to a stock concentration of 0.5mg/ml. A working solution was made with a 1:1000 dilution of stock solution in PBS. Cells were fixed in 4% paraformaldehyde at room temperature for 20 minutes, washed twice with PBS and incubated in darkness with working solution for 20 minutes at RT.

Cells were counterstained with Hoesch 3342 stain (Life Technologies) for 20 minutes in darkness, mounted with ProLong gold antifade reagent (Life Technologies) and left overnight at 4°C. Cells were then visualised with a Zeiss Axio Imager 2 Light Microscope (Cambridge, UK) and fluorescence quantification was performed using imageJ (http://imagej.nih.gov/ij/).

**Mitosox Staining.**

1 vial of Mitosox reagent (Life Technologies) was dissolved in DMSO 13μl and then diluted 1:1000 in Hanks buffered salt solution (Life Technologies) to make a working solution. Cells were washed once with PBS before being covered with working solution (660mls for one well of a 24 well plate) and incubated at 37°C for 45 minutes. Cells were washed twice with PBS before counterstained with Hoesch 3342 stain (Life Technologies) for 20 minutes in darkness. Cells were then washed twice with PBS then visualised with a Zeiss Axio Imager 2 Light
Microscope (Cambridge, UK). fluorescence quantification was performed using imageJ (http://imagej.nih.gov/ij/).

2.6 Molecular Biology

2.6.1 Expression Analysis

RNA extraction

Hepatic RNA was extracted from liver using the Qiagen RNeasy Minikit (Crawley, UK) with Qiazol (Life Technologies) as per the manufacturer’s instructions in the ‘RNeasy Lipid Tissue Handbook’. 20-25mg of liver or 100mg of adipose tissue was homogenised in 1ml Qiazol for 40 seconds and left at room temperature for 5 minutes. 200 µl chloroform was added and the tube shaken vigorously for 15 seconds prior to centrifuging at 14000rpm for 15 minutes at 4°C. Supernatant was removed, and to it added one volume 100% ethanol followed by mixing by pipetting. 700µl was then placed in an RNeasy spin column within a 1.5ml Eppendorf tube and centrifuged at 10000rpm. The flow through was discarded. This was repeated with the remaining lysate/ethanol mix. 700µl buffer RW1 was added placed on the column and centrifuged at 10000rpm. 500µl buffer RPE was added to the column and centrifuged at 10000 for 15 seconds, the flow through was discarded. A further 500µl of buffer RPE was added to the column and centrifuged for 2 minutes at 10000rpm. To elute RNA, 50 µl of RNase free water (Sigma Dorset UK) was added directly to the column membrane and centrifuged at 10000rpm. The eluted liquid was then re-passed through the column in the same manner.

For RNA extraction from cell culture the protocol was identical, however, rather than homogenising tissue, 1ml of Qiazol reagent was added to the cell layer for 5 minutes before scraping off the cell layer and transfer by pipetting to a 1.5ml Eppendorf tube.

RNA Quantification and Integrity

RNA integrity was assessed using the Agilent Bioanalyser 2100 with nanochip 6000 kit as per manufacturer’s instructions (Agilent Technologies, Santa Clara, US). Briefly, 550µl of RNA nano gel matrix was centrifuged at 1500g for 10 minutes at room temperature in the provided spin filter. To a 65µl aliquot of this, 1 µl of blue RNA dye concentrate was added, vortexed and centrifuged at room temperature for 13000g for 10 minutes. An RNA 6000 nano chip was loaded onto the priming station and 9 µl of the gel dye mix was added to the marked well and dispersed using the syringe provided. 1µl of each RNA sample and 1µl of nano 6000 RNA
ladder were loaded in the appropriate wells. Visualisation of gel, electrophoreogram and calculation of RNA integrity number (RIN) was performed using Bioanalyser software. RNA was quantified with the Qubit 2.0 Fluorometer (Invitrogen, Paisley UK) using RNA “broad range” standards as per manufacturer’s instructions.

**Synthesis of complementary DNA (cDNA) by reverse transcription**

Following quantification of RNA, dilutions of 5μg per 50μl were made. The following reaction mixture was made to eliminate residual genomic DNA: 8μl (800ng) RNA, 1μl RQ1DNase and 1 μl 10xDNAsse reaction buffer (Promega, Southampton, UK) with 1μl recombinant RNAses inhibitor (Life technologies). DNase reaction mixture was incubated for 30 minutes at 37°C, 1μl DNase stop solution (Promega) was added to each sample and further incubated at 65°C for 10 mins.

cDNA was reverse transcribed directly from this using the high capacity cDNA reverse transcription kit (Life Technologies, Paisley, UK). Reverse transcriptase (1μl), primer mix (1μl), 10 x reaction buffer (2μl) and a dNTP mix (0.8μl) was added to each sample and made up to 20μl total with RNase free water (Sigma, UK). This was incubated as follows

- 25°C for 10 mins
- 37°C for 120 mins
- 85°C for 5 mins

in a thermocycler (Techne Prime, Bibby Scientific, Birmingham UK). cDNA products were stored at -20°C.

**Quantitative PCR (qPCR) of Gene Expression**

Transcript abundance was quantified using the Pfaffl method to allow incorporation of assay efficiency for accurate relative transcript quantification. Primers (details in section 2.8) were designed to match intron-spanning boundaries wherever possible using the Roche Universal Probe Library (UPL) (https://lifescience.roche.com). Where a satisfactory assay was not available using the UPL system, a commercially designed assay was used (TaqMan® Gene Expression Assays; Life Technologies). 2μl aliquots of each sample was pooled to generate a serial dilution standard curve (1:4 to 1:128) for assay
efficiency calculation. cDNA was diluted 1:20 to fall within the curve. A water and non-reverse transcriptase control were used routinely. Assay composition was as below.

<table>
<thead>
<tr>
<th>Component</th>
<th>Roche UPL</th>
<th>ABI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water (µl)</td>
<td>2.9</td>
<td>2.5</td>
</tr>
<tr>
<td>Quanta 2X Master mix (µl)</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Primers (100µM) forward and reverse (µl)</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>UPL Probe (Roche, Welwyn, UK) (µl)</td>
<td>0.1</td>
<td>-</td>
</tr>
<tr>
<td>ABI Primer probe mix (µl)</td>
<td>-</td>
<td>0.5</td>
</tr>
</tbody>
</table>

qPCR plates were centrifuged at 1500rpm for 2min, to ensure all samples and mastermix were at the bottom of the well. The lightcycler 480 (Roche) was used for thermocycling and fluorescence quantifications.

<table>
<thead>
<tr>
<th>Step</th>
<th>Time (secs)</th>
<th>Temp (°C)</th>
<th>Cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denaturation</td>
<td>300</td>
<td>95</td>
<td>1</td>
</tr>
<tr>
<td>Denaturation</td>
<td>10</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>Annealing</td>
<td>30</td>
<td>60</td>
<td>50</td>
</tr>
<tr>
<td>Elongation</td>
<td>0.1</td>
<td>72</td>
<td></td>
</tr>
</tbody>
</table>

The LC480 Software (Roche) monitored fluorescence during the PCR and plotted amplification curves, fluorescence against PCR cycle number. The crossing point (Cp) was calculated as the maximum point of the second derivative of the amplification curve. Triplicates were deemed acceptable if the standard deviation of the crossing point was less than 0.3 cycles. Negative controls were deemed acceptable if the Cp of any signal was greater than 10 cycles higher than that of the lowest standard. Each assay was deemed acceptable if the calculated reaction efficiency was between 1.8 and 2.1. The mRNA ‘concentration’ of unknown samples were interpolated from the standard curve and expressed relative to the abundance of a reference gene. A minimum of three reference genes were analysed and the mean of the two genes with the lowest variance between treatment groups as used. One-way analysis of variance was used to ensure transcript quantities were not significantly different between groups. Typically, stable genes included Gapdh, Ppia, Tbp and b-Act. 18s was found to be excessively abundant and required a further 1:32 dilution.
RNA labelling prior to micro array analysis was performed using the Illumina TotalPrep RNA amplification kit (Illumina, Cambridge, UK) as per manufacturer’s instructions. We optimised some specific variables in the protocol as follows. 500ng of starting RNA was used with a RIN number of at least 8.5 per sample as determined by Agilent nanochip (Agilent Technologies, Santa Clara, US). Due to an absence of a calibrated 0.5ml tube thermocycler, we use non-stick 0.2ml tubes (BIOplastics, Landgraaf, the Netherlands) and transferred into 1.5ml tubes when required for cDNA binding steps. The final PCR step (in vitro transcription of cRNA) was performed for 14 hours. Lastly, the elution volume of cRNA following cRNA purification in the final step of the protocol was reduced to 100ul rather than the recommended 200ul of nuclease free water to increase yield concentration.

Microarray Expression Analysis

Hybridisation of biotinylated RNA to Illumina mouse WG_6 version 2.0 beadchip microarrays (Illumina, Cambridge, UK) was outsourced to the University of Edinburgh Genomics Core (Western General Hospital, Edinburgh, UK). Illumina Mouse WG-6 v2 or Illumina HT_12 beadchip arrays were used with Illumina Whole Genome Gene Expression Direct Hybridisation Assay (WGGX). Imaging was performed and analysed using the Illumina HiScan platform and genotypes were called automatically using GenomeStudio Analysis software version 2011.1.

RNA sequencing

Total RNA samples were assessed on the Agilent Bioanalyser with the RNA 6000 Nano Kit (Agilent Technologies, catalogue #5067-1511) for quality and integrity of total RNA, and then quantified using the Qubit RNA BR kit (catalogue # Q10210). Samples were also assessed for DNA contamination using the Qubit DNA HS assay Kit (catalogue # Q32851) both Life Technologies. Libraries were prepared from 500ng of each total-RNA sample using the TruSeq Stranded Total RNA with Ribo-Zero Human/Mouse/Rat kit (catalogue #RS-122-2201, Illumina, US). Libraries were quantified by PCR using the Kapa Universal Illumina Library Quantification kit (catalogue # KK4824) and a selection were assessed for quality using the Agilent Bioanlyser with the DNA HS Kit. Sequencing was performed using the NextSeq 500/550 High-Output v2 Kit (150 cycles) on the NextSeq 550 platform. The 6 samples were combined into 1 pool and run across one flow cell.
2.6.2 Epigenome Profiling

**DNA extraction**

DNA extraction was performed using a modified version of the Qiagen DNeasy Blood and Tissue kit (Crawley UK) protocol. Tissue was homogenised for 2 x 20 second bursts in 360mls of ATL buffer, 40ul of proteinase K (Quiagen) was added and the homogenate incubated at 55°C overnight. 20ul of RNase A (Purelink, Ambion, UK) was added and incubated at room temperature for 2 minutes. 400µl buffer AL was added to each sample and vortexed. 400 µl 100% ethanol was then added and vortexed. The solution was applied to a DNeasy spin column and centrifuged at 8000rpm for 1 minute. The flow through was discarded. The spin column was then placed in a new centrifuge tube and 500µl buffer AW1 added, followed by centrifuging at 8000 rpm for 1 minute; the flow through was discarded. 500 µl buffer AW2 was applied to each column and centrifuged for 14000 rpm for 3 minutes; the flow though and collection tube were discarded. The column was then placed in a 1.5ml eppendorf tube and DNA eluted from the column with 200 µl buffer AE centrifuged at 8000rpm for 1 minute twice to make a final volume of 400ul.

**DNA Quantification and Integrity**

DNA was quantified with the Qubit 2.0 Fluorometer (Invitrogen, Paisley UK) using DNA “broad range” standards as per manufacturer’s instructions for DNA samples. DNA integrity was determined by resolving extraction products on a 1.5% agarose gel and staining with gel red (Cambridge Bioscience, Cambridge UK) 0.005% final concentration.

**DNA Sonication**

DNA was sonicated using a Bioruptor sonicator (Diagenode, Ougree, Belgium). 10µg genomic DNA was diluted to 450µl in TE buffer and sonicated at 4°C, 30s on 30s off for 28 cycles. This was paused every 6 cycles, with samples vortexed and centrifuged briefly to encourage complete sonication of the sample. The quality of the sonication was assessed by running DNA in a 1.5% agarose gel in 0.5xTBE buffer for 90 minutes at 70 volts. Acceptable sonication was deemed as fragments between 100 and 600 base pair length with a mean of 250 to 300 base pairs and a uniform appearance between sonicated samples.
Hydroxymethylcytosine DNA Immunoprecipitation

2.5 μg of DNA was diluted up to 450ul in TE buffer in a 1.5ml microcentrifuge tube (Eppendorf DNA LoBind cat. 0030 108.051) and then denatured by placing in boiling water in a heat block for 10 minutes. Samples were immediately cooled on ice for 5 minutes. 50ul of 10xIP buffer was then added and vortexed briefly. 50ul ‘input’ was then removed and placed at 4°C until the DNA clean up step below. 1ul of anti 5hmC antibody (Active Motif 39769) was then added the remaining sample and incubated for 3 hours at 4°C.

40ul of Dynabeads protein G (Invitrogen, Paisley, UK) was prewashed with BSA 0.1% in PBS for 5 minutes on rotating wheel at room temperature. Beads were then collected on a magnetic rack, supernatant removed, centrifuged at 2000rpm for 5 seconds, bound to rack and supernatant remove again. Washing was repeated twice. The beads were then resuspended in 40ul of 1xIP buffer and added to the DNA/antibody mixture for 1 hour at 4°C on a rotating wheel. Beads were then collected on the magnetic rack, the supernatant discarded. Beads were then spun at 2000rpm for 10 seconds, collected and residual supernatant removed. Beads were then washed three times with 1ml of cold 1xIP buffer in the same manner as above. Beads were then resuspended in 250ul of digestion buffer, 20ul of proteinase K 20mg/ml (Roche, UK) was added and incubated in a thermoshaker at 1000rpm, 55°C overnight. Following this, tubes were spun at 2000rpm for 5 second, beads collected on a magnetic rack and the supernatant with enriched DNA fragments removed. Both enriched fraction and input samples were then purified using the Qiagen Qiaquik PCR purification kit (Qiagen, UK) with elution in 22μl of water.
Quantitative PCR of Immunoprecipitated Products

10μl of the eluted DNA from immunoprecipitation was diluted 1:5 in water and was used in the following reaction mixture.

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume (μl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SYBR green Master Mix (Roche, Burgess Hill, UK)</td>
<td>5</td>
</tr>
<tr>
<td>Forward Primer (100 μM)</td>
<td>0.02</td>
</tr>
<tr>
<td>Reverse Primer (100 μM)</td>
<td>0.02</td>
</tr>
<tr>
<td>Water</td>
<td>1.96</td>
</tr>
</tbody>
</table>

Estimation of abundance was determined as described above in section 2.6.1. Primer sequences can be found in section 2.8.

Whole Genome Amplification (WGA).

WGA was performed using the Sigma SEQXE DNA amplification kit as per the manufacturers instructions. 10μl of immunoprecipitation product and input sample were used. 18 cycles of DNA amplification was performed in a standard PCR machine (Techne Prime, Bibby Scientific, Birmingham UK). For adaptor removal – all samples were quantified by nanodrop, the lowest DNA concentration was taken and volume required for 2.1μg calculated. This volume was then used for all samples in the adaptor removal step. Final elution was in 30μl of water.

Ion Torrent Proton Sequencing

Proton sequencing was outsourced to the Wellcome Trust Clinical Research Facility at the Western General Hospital, Edinburgh, UK. All kits and hardware for this procedure were from Life Technologies (Paisley, UK) unless specified. Briefly, samples were quantified using the Qubit dsDNA HS kit and approximately 100ng of DNA was used to generate a DNA library from each sample using the Ion Xpress Plus Fragment Library Kit. During this process, DNA fragments are end repaired and then ligated to Ion specific barcode adaptors before being amplified (8 cycles) and twice purified using the Agencourt AMPure XP PCR clean up kit (Beckman Coulter, High Wycombe, UK) which size selects fragments approximately 100–250 bp in length. Libraries were then quality controlled using the Agilent Bioanalyser DNA
HS kit (Agilent, Santa Clara, US) and pooled in equimolar pairs prior to template preparation using the Ion PI™ Hi-Q™ OT2 200 Kit and sequencing on the Ion Torrent semiconductor sequencer using the Ion PI™ Hi-Q™ Sequencing kit and an Ion PI™ Chip Kit v3. For consistency, each sample was sequenced on a PI chip with its own input.

2.6.3 Liquid Chromatography Mass Spectrometry

5ug of frozen tissue was ground up in a mortar and pestle in 200ul of pre cooled LCMS extraction buffer (2.2). Samples were then spun at 16000g 4°C for 5 minutes and supernatant collected and stored at -80°C. LCMS analysis was outsourced to The IGMM Mass spectrometry laboratory, Institute of Genetics and Molecular Medicine, University of Edinburgh.

2.6.4 Genotyping of TET1 Knock Out Offspring.

Genotyping was performed using identification ear notches as follows. 200μl NaOH 50mmol was added to the sample and placed in heat block for 20 minutes followed by cooling on ice for 5 minutes. Samples were vortexed thoroughly and to them was added 50μl of 200mM Tris HCl pH8.0. Polymerase ‘mastermix’ used was KAPA2G Green HS Mastermix (KAPA Biosystems, London, UK). Primer sequences can be found in section 2.8.5.

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume (μl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KAPA2G HS Green Mastermix</td>
<td>6.25</td>
</tr>
<tr>
<td>WT Reverse primer 100uM</td>
<td>0.06</td>
</tr>
<tr>
<td>Mutant Reverse Primer 100uM</td>
<td>0.06</td>
</tr>
<tr>
<td>Common Forward Primer 100uM</td>
<td>0.06</td>
</tr>
<tr>
<td>DNA Lysate</td>
<td>2</td>
</tr>
<tr>
<td>Water</td>
<td>3.57</td>
</tr>
<tr>
<td>TOTAL</td>
<td>12</td>
</tr>
</tbody>
</table>

PCR was performed on a thermocycler (Techne Prime, Bibby Scientific, Birmingham UK) as follows.
<table>
<thead>
<tr>
<th>Temperature (°C)</th>
<th>Time</th>
<th>Cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>94</td>
<td>2 min</td>
<td>1</td>
</tr>
<tr>
<td>Start Cycle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>94</td>
<td>20 sec</td>
<td></td>
</tr>
<tr>
<td>65 (-0.5°C per cycle)</td>
<td>15 sec</td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>10 sec</td>
<td></td>
</tr>
<tr>
<td>End Cycle</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Start Cycle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>94</td>
<td>15 sec</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>15 sec</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>10 sec</td>
<td>28</td>
</tr>
<tr>
<td>End Cycle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>2 min</td>
<td>1</td>
</tr>
</tbody>
</table>

Products were run on a 3% agarose gel at 70V for 1.5 hours. Mutant bands were determined at ~650bp with wild type bands assumed at ~300bp.

### 2.7 Bioinformatic and Statistical Analysis

#### 2.7.1 Microarray Gene expression.

Background subtraction, between array adjustments, quartile normalisation and linear modelling for differential expression was performed using R statistics (http://www.r-project.org) and Limma package for bioconductor (www.Bioconductor.org). Plots were created using ggplot2 package and heatmap2 also from bioconductor. A gene ontology analysis was also performed in R using GOstats and Pathview packages also from Bioconductor.

#### 2.7.2 RNA sequencing analysis

NextSeq 550 sequencing data was analysed using BaseSpace (https://basespace.illumina.com/). Basespace apps ‘Tophat’ and ‘Cufflinks Assembly’ were used to perform alignment to the mm10 genome and differential expression analysis. Plots were generated using R version 3.2.1 and packages ‘ggplot2’ and ‘gplots’ (heatmap.2 function) both from Bioconductor (http://bioconductor.org/). Gene ontology analysis with
statistical over representation testing was performed using the PANTHER
http://pantherdb.org/.

2.7.3 Epigenome profiling.

Raw sequence data analysis from the semiconductor sequencer was processed using Ion
Torrent Suite software version 4.0.2 (Life Technologies, Paisley, UK). This quality controls
the reads and uses the TMAP aligner to generate binary alignment map (BAM) files of
adequate compression for export. BAM files were then converted to BED files and then
subsequently to sorted wiggle files using the GALAXY server at the Human Genetics Unit,
Western General Hospital, UK. I then performed a sliding window analysis across the entire
genome, essentially calculating the average DNA modification scores in a non-overlapping
set of 150bp windows.

To normalise between data sets, I then performed total read normalisation to account for
variations in the sequencing reactions. This allows the user to subtract the values sequenced
in the input from the immunoprecipitation, essentially giving an epigenetic score over each
window present only in the IP sample and allowing for comparison between datasets.
Resulting wiggle files were then visualised using the Integrative Genomics Viewer (IGV)

Average patterns of 5hmC were plotted across length normalised genes. In short, gene
coordinates were converted to a percentage of length (TSS = 0%, TES=100%) extended to
cover 25% upstream and downstream of these co-ordinates. Genes of less than 2 basepairs in
length were removed from the dataset. Processed 5hmC patterns (average levels of reads
across 150bp windows) were then obtained using the “sliding window over length normalised
features” on the GALAXY server.

The mean normalized read score of 5hmC over annotated regions of the genome were
calculated using annotation coordinates from the mm9 ENCODE build of the mouse genome,
which can be downloaded from the UCSC genome bioinformatics resource
(https://genome.ucsc.edu). Annotation co-ordinates for enhancer regions wee also
downloaded from the ENCODE project via the Ludvig Institute for Cancer Research which
can be found at http://chromosome.sdsc.edu/mouse/download.html. Enhancer regions in this
database are predicted via the presence of H3K4me1 but absence of H3K4me3 on ChIP-
sequencing.
Boxplots of normalized reads per region, scatter plots of differentially hydroxymethylated windows and regions, and correlations with gene expression were generated using R scripts for ggplot2 from Bioconductor.

2.7.4 Null Hypothesis Testing Statistical Methods

Statistical comparisons between groups of cells or animals were performed using SPSS version 21 (IBM, Illinois, USA). Data sets were routinely tested for normality of distribution and outliers using histograms, boxplots and the Shapiro Wilks test. Data with outliers or of non-parametric distribution was either log_{10} transformed or a non-parametric test used as required.
## 2.8 Primer Sequences

### 2.8.1 Gene Expression Quantitative PCR Analysis Mouse

<table>
<thead>
<tr>
<th>Transcript</th>
<th>Forward</th>
<th>Reverse</th>
<th>UPL probe number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gapdh</td>
<td>gggtcctataataacgctgc</td>
<td>ccatttgtctacggaagca</td>
<td>52</td>
</tr>
<tr>
<td>Ldha</td>
<td>ggcactgacgcagacaag</td>
<td>tgcacccctgtgagctactg</td>
<td>12</td>
</tr>
<tr>
<td>Scd1</td>
<td>gtcctctctgcaagccctac</td>
<td>cagagcctggtcaagtctag</td>
<td>34</td>
</tr>
<tr>
<td>Aacs</td>
<td>cagtgtatgtgagatgagg</td>
<td>acacagccaggctcaagtg</td>
<td>1</td>
</tr>
<tr>
<td>Fasn</td>
<td>tcaacaatccacatgg</td>
<td>gttgtgaagtgcaggttag</td>
<td>1</td>
</tr>
<tr>
<td>Lpl</td>
<td>ctcgctctcagtgccctac</td>
<td>aggcctgtgtgctgctt</td>
<td>95</td>
</tr>
<tr>
<td>Mlxlpl</td>
<td>gacaccctgcaacggacctt</td>
<td>gttgtgctttgtcctcctt</td>
<td>104</td>
</tr>
<tr>
<td>Acs1l</td>
<td>aaagatggctgtttacacag</td>
<td>cgtaaatctcaaggtgcccatt</td>
<td>46</td>
</tr>
<tr>
<td>Ces1d</td>
<td>cctctacccgcctatgtg</td>
<td>cctctgttgtggaagagcc</td>
<td>89</td>
</tr>
<tr>
<td>CES1f</td>
<td>aagcctgtgatccctgtcat</td>
<td>ttgtggtaatcacaagcaa</td>
<td>5</td>
</tr>
<tr>
<td>Ces3b</td>
<td>ctggagcatgattgtctca</td>
<td>atgtgctgtgacaatccact</td>
<td>3</td>
</tr>
<tr>
<td>Ces1g</td>
<td>cccctcttgagttctctga</td>
<td>cctctgggttggtagcac</td>
<td>1</td>
</tr>
<tr>
<td>Pdk1</td>
<td>gcacattggaagcataaatcc</td>
<td>cctagcgttcctataacccatct</td>
<td>22</td>
</tr>
<tr>
<td>Sc4mol</td>
<td>tggtgaatgtgttttgtgt</td>
<td>gcgggtgagaggaatac</td>
<td>2</td>
</tr>
<tr>
<td>Fdps</td>
<td>gacagtgggctggtgatg</td>
<td>gcgggtgagaggaaggtc</td>
<td>1</td>
</tr>
<tr>
<td>Gene</td>
<td>Forward Primer</td>
<td>Reverse Primer</td>
<td>Catalogue Number</td>
</tr>
<tr>
<td>--------</td>
<td>-------------------------</td>
<td>-------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Lss</td>
<td>tgtctggctccaggtccat</td>
<td>gacgagtgctcgggatg</td>
<td>3</td>
</tr>
<tr>
<td>Hsd17b7</td>
<td>gtgcagatggatgctagcag</td>
<td>aggcaggattccagcattc</td>
<td>6</td>
</tr>
<tr>
<td>Mvd</td>
<td>ctgaatggtaagcaggag</td>
<td>gagtgtccccgtctctgt</td>
<td>1</td>
</tr>
<tr>
<td>Sqle</td>
<td>catgagtctcgggaagcag</td>
<td>tgaagcacaacaccttaaaactt</td>
<td>53</td>
</tr>
<tr>
<td>Beta Actin</td>
<td>ctaaggccaaccgtaaaag</td>
<td>accagaggcatacagggaca</td>
<td>64</td>
</tr>
<tr>
<td>Tbp</td>
<td>gggagaatcatggaccagaa</td>
<td>gatgggaatccaggatga</td>
<td>97</td>
</tr>
<tr>
<td>Cyp17a1</td>
<td>Life Technologies (Paisley, UK ) Taqman assay catalogue number # 4331182</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ppia</td>
<td>Life Technologies (Paisley, UK ) Taqman assay catalogue number # 4331182</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.8.2 Gene Expression Quantitative PCR Analysis Human

<table>
<thead>
<tr>
<th>TRANSCRIPT</th>
<th>Forward</th>
<th>Reverse</th>
<th>UPL probe number</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALB</td>
<td>gtgagggtgctcatcggtt</td>
<td>gagcaaggcaatcaacacc</td>
<td>7</td>
</tr>
<tr>
<td>OCT3/4</td>
<td>caattgccaagctcctga</td>
<td>agatggctgtttgctgaat</td>
<td>7</td>
</tr>
<tr>
<td>HNF4A</td>
<td>agcaacggacagatgttga</td>
<td>tcagaccctgagccacct</td>
<td>27</td>
</tr>
<tr>
<td>NANOG</td>
<td>atgcctcacaggagagtgtg</td>
<td>cagggctgtctgaataagc</td>
<td>69</td>
</tr>
<tr>
<td>PLIN1</td>
<td>agagcgcagtagcctgg</td>
<td>ttggcagctgaactgg</td>
<td>1</td>
</tr>
<tr>
<td>PLIN2</td>
<td>tcagectccattctactgctcacc</td>
<td>cctgaatitctgattggcact</td>
<td>72</td>
</tr>
<tr>
<td>PCK1</td>
<td>agatggaggaagagggcact</td>
<td>gttcagfagagccscaacc</td>
<td>41</td>
</tr>
<tr>
<td>PCK2</td>
<td>cgaagctcccccaagtacaa</td>
<td>gtctctactctgccacact</td>
<td>20</td>
</tr>
<tr>
<td>G6PD</td>
<td>aacagagtgggccctttctca</td>
<td>ggaggctgcatctgtctact</td>
<td>5</td>
</tr>
<tr>
<td>FASN</td>
<td>categgetccaccaagtc</td>
<td>gctatgggaagtgcagttgg</td>
<td>1</td>
</tr>
<tr>
<td>LXR</td>
<td>ggctatgtgcagaagcte</td>
<td>cggctatggggtcctgtggt</td>
<td>44</td>
</tr>
<tr>
<td>ACACA</td>
<td>tcacaactgctgctccaaac</td>
<td>attttctgccccagctccac</td>
<td>1</td>
</tr>
<tr>
<td>SREBF2</td>
<td>gtgggccacttgaggcaag</td>
<td>gacagcagtgcgcagacatta</td>
<td>2</td>
</tr>
</tbody>
</table>
2.8.3 hmeDIP DNA Quantification Mouse

<table>
<thead>
<tr>
<th>GENE</th>
<th>Forward</th>
<th>Reverse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gapdh</td>
<td>ccaacctctccctcccagtttcc</td>
<td>cctataataccggaagtgcagc</td>
</tr>
<tr>
<td>Actin Promoter</td>
<td>atgtacaggaatatgcctcgg</td>
<td>cttaagtgcgctagataccac</td>
</tr>
<tr>
<td>H19 Genic</td>
<td>gccaagagagaagaaggaga</td>
<td>gaatgttaaggagactgagg</td>
</tr>
<tr>
<td>Tex19.1 Genic</td>
<td>gggagatatgtaaatgagctgg</td>
<td>catcctacctccctgactgag</td>
</tr>
<tr>
<td>Lss genic</td>
<td>acttgtctctagatgcctta</td>
<td>ggtgagcttcattgccagttgg</td>
</tr>
<tr>
<td>Sc4mol genic</td>
<td>taccctccccacacaataacaaac</td>
<td>acgtaatctctctgtgggca</td>
</tr>
<tr>
<td>Mvd genic</td>
<td>cattgaggttaggagatcgg</td>
<td>gagtcttgtgctgctgagcgg</td>
</tr>
<tr>
<td>Sqle genic</td>
<td>cggtattggtggagagttta</td>
<td>tgtacagagggagacaggga</td>
</tr>
<tr>
<td>Hsd17b7 genic</td>
<td>gagctgggcagagaaaaaca</td>
<td>ccaataagaatcgacagagg</td>
</tr>
<tr>
<td>Fdps genic</td>
<td>ttgcacactccctccagcttag</td>
<td>gaagccctccatatcggttgg</td>
</tr>
<tr>
<td>Dhcr7 genic</td>
<td>cgttgcgtgttttgtagatg</td>
<td>gagataattgggtgtaga</td>
</tr>
<tr>
<td>Cpy17a1 genic</td>
<td>ccacacctgggattaaaca</td>
<td>gcaggaagtctgtaggaat</td>
</tr>
</tbody>
</table>
### 2.8.4 hmeDIP DNA Quantification Human

<table>
<thead>
<tr>
<th>GENE</th>
<th>Forward</th>
<th>Reverse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gapdh Promoter</td>
<td><code>ggctactageggttttcg</code></td>
<td><code>aagaagatgcggctgactgt</code></td>
</tr>
<tr>
<td>Line-1 PA1</td>
<td><code>aaatggtgctgggaaaactg</code></td>
<td><code>gccattgcttttgggtttt</code></td>
</tr>
<tr>
<td>UBIAD1 genic</td>
<td><code>ctcttcctcctcctctctcct</code></td>
<td><code>catccaggaaccacagtctct</code></td>
</tr>
<tr>
<td>H19 genic</td>
<td><code>gatctcgccctagtgtgaa</code></td>
<td><code>gtgatgtgagcctgcact</code></td>
</tr>
</tbody>
</table>

### 2.8.5 Tet1 Knockout mouse genotyping primers

| Wild type Forward | `tcagggagctcatggagacta` |
| Mutant Forward    | `aactgattccctctctact`   |
| Common Reverse    | `tttaagcatggctggagtc`   |
CHAPTER 3: EVALUATION OF CHOLINE AND METHIONINE DEFICIENT MICE: NEW INSIGHTS INTO THE MECHANISM OF STEATOSIS AND INSULIN RESISTANCE

3.1 Introduction

The accumulation of hepatic triglycerides and other lipid species, known as hepatic steatosis, is highly prevalent in western countries and is increasing rapidly in the developed world. Examination of liver histology in apparently healthy liver donors suggests a prevalence of steatosis of 12-18% in European countries with around 31% of the unselected United States population found to have hepatic steatosis based on the gold standard 1H-MRS liver lipid quantification. This has two major clinical consequences. Firstly steatosis associates strongly with insulin resistance and T2DM and is widely accepted to be the hepatic component of the metabolic syndrome. Secondly, a subpopulation (3-16%) of patients with NAFLD will progress to an inflammatory steatohepatitis and are at high risk of irreversible fibrotic liver disease and hepatocellular carcinoma.

Faithful recapitulation of human fatty liver disease in tractable animal models is essential for the exploration of relevant molecular pathways and screening novel therapeutics in the prevalent disease. Accurately modelling the stages of human NAFLD in order to address the clinical, biochemical and histological features of human disease represents a challenge. Strategies in rodent models include dietary excess of saturated fats and carbohydrates, genetic manipulation of satiety (ob/ob or db/db mice and Zuker fa/fa rats) and genetic manipulation of lipid metabolism (Ppara -/- and Acox -/- mice) and genetic manipulation of lipid metabolism (Ppara -/- and Acox -/- mice) and genetic manipulation of lipid metabolism (Ppara -/- and Acox -/- mice) and genetic manipulation of lipid metabolism (Ppara -/- and Acox -/- mice) and genetic manipulation of lipid metabolism (Ppara -/- and Acox -/- mice) and genetic manipulation of lipid metabolism (Ppara -/- and Acox -/- mice) and genetic manipulation of lipid metabolism (Ppara -/- and Acox -/- mice) and genetic manipulation of lipid metabolism (Ppara -/- and Acox -/- mice). Impaired methionine metabolism and suppressed S-adenosylmethionine (SAMe) production has been repeatedly documented as a feature of chronic liver disease, regardless of aetiology including patients with NAFLD. SAMe is a universal methyl-donor and an essential substrate for the methylation of many biological active molecules including DNA, phospholipids, proteins and biogenic amines. Restriction of methyl donors in rodents rapidly and reproducibly results in a liver injury histologically similar to hepatic steatosis (choline deficient diets: CDD) and NASH (methionine-choline deficient diets: MCDD) and as such these models have been used widely in mechanistic and therapeutic studies. These observations suggest an overlapping mechanism that remains incompletely understood and warrants further research. Furthermore, the degree to which the liver injury in these models reflects the global transcriptional changes found in human disease has not been reported.

I therefore set out to further define the cause of liver pathology in both CDD and MCDD mice by mapping pathways of lipid, cholesterol and one carbon metabolism and secondly to
evaluate their potential usefulness as models of transcriptional dysregulation in human NAFLD. To address these aims I have performed a detailed bioinformatics analysis of the hepatic transcriptome in mice maintained on CDD and MCDD and compared these with published human NAFLD datasets.

3.1.1 HYPOTHESIS

Methyl donor deficiency induces hepatic injury histologically and transcriptionally representative of human NAFLD and is therefore a suitable model for epigenetic studies.

3.1.2 AIMS

1) Generate and characterize the effects of choline and methionine/choline combined deficiency on mouse liver.
2) Examine the effect of these diets on the mouse hepatic transcriptome with respect to glucose, lipid and methyl donor metabolism.
3) Perform a comparative cross species transcriptional analysis between each model and published human NAFLD data sets.
4) Ultimately determine the suitability of this model for the study of epigenetic dysregulation in metabolic liver disease.
3.2 Methods

3.2.1 Animals

All experiments were carried out under a UK Home Office license and with local ethical committee approval. Adult male C57BL/6J mice (Charles River, Tranent, UK) were maintained under controlled conditions as described in 2.3.1. Mice were maintained on control (C), CDD or MCDD diets (Dyets, Bethlehem, PA) for 4 weeks and then killed and tissues collected and used for histology (2.4) or snap-frozen and stored at -80°C. Diet composition can be found in 2.3.1.

3.2.2 Histology staining, triglyceride and SAMe quantification

Livers were removed and sections were fixed in methacarn solution (methanol:chloroform:glacial acetic acid; ratio 6:3:1) and mounted in paraffin blocks prior to staining with haematoxylin and eosin or picosirius red (2.4). Image analysis for fibrosis content was performed in ImageJ (http://imagej.nih.gov/ij/) and hepatic triglyceride concentration was determined (2.3.8). SAMe was quantified from liver sections using Matrix-assisted laser desorption/ionization (2.3.9).

3.2.3 Gene Expression

RNA was extracted from snap frozen liver tissue using the RNasy kit (Qiagen, Manchester, UK) as per 2.6.1, quantified, DNase treated and reverse transcribed (2.8.1). RNA (n= 4 per group) was biotin labelled and submitted for transcriptional analysis at the Wellcome Trust Clinical Research Facility, Western General Hospital, UK (2.8.1). Quantitive real time PCR validation (2.6.1) was performed throughout each treatment cohort on key mediators in pathways of lipid transport (Lipoprotein lipase, Lpl), de novo lipogenesis (Stearoyl-coA desaturase-1, Scd1; Acetyl-CoA Synthetase, Accs; Fatty acid synthase, Fasn; MLX Interacting Protein-Like, Mlxipl; Acyl-CoA Synthetase Long-Chain Family Member 1, Acs1l), fibrogenesis (Matrixmetaloprotein 12, Mmp12), tryglyceride export (Carboxyesterases; Ces1d, Ces1f, Ces1g, Ces3b) and the insulin signaling mediator 3-phosphoinositide-dependent protein kinase-1 (Pdk1) synthesis. Gene expression is presented relative to the mean of three housekeeping genes (Glyceraldehyde-3-Phosphate Dehydrogenase, Gapdh; Cyclophylin, Ppia; lactate dehydrogenase A, Ldha).
3.2.4 Transcript analysis

RNA labelling was performed on 500ng RNA using the Illumina Total Prep RNA amplification kit (2.6.1) and subsequently hybridised to Illumina Mouse-ref6 expression bead arrays (2.6.1). Intensity data were generated using a HiScan array scanner (Illumina, San Diego, USA) and analysed using iScan Illumina software. Data analysis and generation of plots were performed in RStudio (http://www.rstudio.com) with R version 3.12. Data import, quality control, normalisation and between array adjustment was performed using Lumi package and differential expression was determined using Limma package (Bioconductor.org). Unsupervised clustering was performed using Euclidean distance. Where multiple probes mapped to the same gene, the median result was used. Data have been uploaded to EBI-ArrayExpress, accession number E-MTAB-3943.

3.2.5 Pathway Analysis

Gene Ontology and pathway enrichment was performed using the GOstats package (Bioconductor.org). Investigation of fat handling was performed by interrogating relevant pathways of lipid metabolism and insulin signalling from the Kyoto encyclopedia for genes and genomes (KEGG) module database (http://www.genome.jp/kegg/module.html). To further understand the molecular mechanisms of this dietary phenomenon I mapped differentially expressed genes from KEGG module sets ‘Lipid Metabolism’, ‘Fatty Acid Metabolism’, ‘Sterol Biosynthesis’ and ‘Terpenoid Backbone Biosynthesis’, and KEGG pathways ‘Fat Digestion and Absorption’ and ‘Insulin Signalling Pathway’ and ‘Glycerolipid Metabolism’. In addition the family of carboxylesterase (Ces) genes were analysed due to their recently discovered role in triglyceride hydrolysis. Furthermore, to dissect the link between lipid and one carbon metabolism, relevant mediators were analysed and mapped to known biochemical pathways.

3.2.6 Cross species comparison

A comparison of transcriptional data from both CDD and MCDD was made with published human expression sets of simple steatosis (GSE48452), NASH (GSE49541) and HCC (GSE60502). Data sets were retrieved from the ArrayExpress archive (http://www.ebi.ac.uk/arrayexpress/). Predicted gene orthologues were determined using homologene via the Hugo Gene Nomenclature Committee server (http://www.genenames.org).
3.2.7 Statistics

Animal model and qPCR statistical analysis was performed using SPSS statistics (Illinois, USA). Data were routinely analysed for outliers, normalisation and sphericity where required. Non-parametric data were either log transformed or a non parametric test used as indicated.
3.3 Results

3.3.1 CDD and MCDD phenotype
After 4 weeks, mice fed CDD gained significantly less weight than mice on control diet whereas mice fed MCDD lost weight from the start of the dietary intervention in keeping with previous reports (Figure 3.1). Total hepatic triglyceride content in mice on CDD and MCDD was significantly greater than mice fed on control diet however there was no difference between treatment groups (Figure 3.1). Both CDD and MCDD induced profound hepatic steatosis (Figure 3.2). MCDD but not CDD induced an increase in new collagen deposition as detected by picosirious red staining (figure 3.2).
Figure 3.1. Body weight and hepatic lipid content. (A) Mice on CDD gain less weight whereas those on MCDD lose weight compared with controls. B) Hepatic liver triglyceride content was increased on both CDD and MCDD diets. * P < 0.05 ** P < 0.01 (one-way ANOVA with Tukey post hoc analysis). Error bars = +/- SEM
Figure 3.2 Liver histology from control, CDD and MCDD mice. A) H+E = Hematoxylin and Eosin staining showing marked macrovesicular steatosis in both CDD and MCDD mice (black arrows). PSR = Picosirius Red stain staining for new collagen formation showing increased periportal and interstitial fibrosis in MCDD animals (black arrows). B) Quantified fibrosis content (PSR positive staining as a percentage of entire image). * P < 0.05 (one way ANOVA with Tukey post hoc analysis). Error bars = +/- SEM
3.3.2 Effect of CDD and MCDD on expression of lipid, glucose and one-carbon metabolic pathways.

Unsupervised clustering by Euclidean distance of the top 500 most variable transcripts between groups was sufficient to cluster mice by dietary intervention (Figure 3.3). There was a high degree of overlap in transcriptional derangement between diets. 816 (81.6%) of the 1004 genes differentially expressed in CDD, were also differentially expressed in MCDD. MCDD induced the differential expression of a further 1948 transcripts over and above choline restriction alone. In all, considerably more genes were upregulated than down regulated in both diets (Figure 3.3).

Genes which were suppressed by 2-fold or greater in both interventions were markedly over-represented in GO-terms for lipid, sterol, fatty acid and organic acid biosynthesis. Over-expressed pathways in CDD mice included ‘immune system process’ and ‘inflammatory response’. MCDD induced transcripts were over represented in pathways of ‘positive regulation of mitotic cell cycle’ and ‘negative regulation of cell cycle arrest’ (Figure 3.4).

Following examination of ontology annotations with the greatest enrichment in each diet I proceeded to examine transcripts of most extreme derangement with putative biological significance to chronic liver disease. In both CDD and MCDD mice, key upregulated genes included the immune mediators (Gpnmb, Ly6d, Ccl4); fibrosis mediators (Mmp12, Mmp13, Timp1, Cxcl10, Col1a1); and the detoxification enzyme Gsta1. The most significantly downregulated genes in both diets included lipid synthesis genes (Sqle, Elovl3, Elovl6, Aacs, Acly, Acss2, Acacb), the endopeptidase inhibitor Serpina4-ps1 and the multifunctional triglyceride metabolism enzymes Pnpla3 and Pnpla5. In addition, genes of interest up-regulated in MCDD included the mitotic proteins Cdc20, Nupr1, the metalloproteinase Adam32, the inflammatory mediator Slpi and the aldo-ketoreductase Akrb7. Global visualisation of transcriptional changes in each diet by volcano plot revealed a more significant derangement in MCDD mice both in the number of genes and the fold change from control (Figure 3.5).
Figure 3.3 Global Transcriptional Analysis of CDD and MCDD livers. (A) The transcript profiles of each dietary intervention were sufficiently consistent to cluster by Euclidean distance. (B) Venn diagram demonstrating high degree of overlap in dysregulated transcripts (≥2 fold change) in each group and the direction of transcriptional change.
Figure 3.4 Gene ontology analyses of dysregulated genes in dietary interventions. The top 8 over represented GO terms in CDD (hatched bars) and MCDD (solid bars) in upregulated (red) and down regulated (blue) gene sets. Adjusted P value determined using a modified Fishers exact test.
Figure 3.5 Volcano plots demonstrating degree and direction of transcriptional change in CDD (A) and MCDD (B) mice versus control animals. Dotted lines and red colour represent adjusted P values < 0.05 and two-fold differential expression change.
Validation of microarray findings was performed using quantitative PCR of functionally relevant genes in lipid uptake (Lpl), mediators in *de novo lipogenesis* implicated in hepatic lipid accumulation in human NAFLD (*Scd1, Aacs, Fasn, Mlxipl, Acsl1*) and hepatic fibrosis (*Mmp12*). Furthermore, due to recent interest in their role in triglyceride hydrolysis and lipid clearance I examined four members of the carboxylesterase (Ces) gene family \(^{296,297,300,301}\). Microarray findings and qPCR data were consistent in all cases (Figure 3.6).
Figure 3.6 CDD and MCDD induce changes in the hepatic expression of genes important in lipid metabolism and storage. Microarray analysis (A) and qPCR validation (B) of selected genes important in lipid transport, de novo lipogenesis, fibrogenesis, triglyceride hydrolysis and insulin signalling. (A) Adjusted P value < 0.05 (Benjamini–Hochberg) for all samples except those marked σ. (B) * = P < 0.05 versus control, one-way ANOVA with Bonferroni post hoc analysis. Error bars = +/- SEM.
**Lipid Pathway Analysis**

I then proceeded to examine pathways of lipid disposal to identify genes which might contribute to the abnormal lipid sequestration in these diets. Genes differentially expressed in relevant KEGG pathways in either group are depicted in Figure 3.7. Typically, expression changes were in the same direction in both groups but were greater in MCDD. Key pathways of hepatic fatty acid clearance (triglyceride synthesis, β-oxidation, microsomal (ω) oxidation and peroxisomal oxidation) were all relatively unaffected, with the exception of up regulated phosphatidic acid phosphatases *Pap2a* and *Pap2c* (which convert phosphatidic acids to diacylglycerol) and *Elov11* in MCDD.

Genes involved in fatty acid, cholesterol and terpenoid synthesis were suppressed in both interventions. Consistant with previous evidence of impaired VLDL secretion in these mice, key mediators of triglyceride hydrolysis and VLDL assembly (*Pnpla3* and the carboxylesterase enzymes: *Ces1d, Ces1f, Ces3b*) were markedly suppressed in MCDD, with *Pnpla3* also suppressed in CDD. Interestingly, the Ces1b isoform was clearly up-regulated in both groups in contrast to the other members of this class suggesting differing function. Finally, two components of hepatic insulin signal transduction (*Irs2* and *Pdk1*) were down-regulated in MCDD but not CDD mice.
Figure 3.7 Heatmap depiction of expression changes in KEGG pathways of lipid transport, lipid synthesis and degradation and insulin signaling. Up- and down-regulated genes in each dietary intervention versus control animals are demonstrated by colour key. Pathways of lipid and cholesterol synthesis are globally suppressed.
One Carbon Metabolism

Methionine and choline are both components of the one carbon metabolism cycle donating methyl groups to a variety of biologically active molecules \(^3\). I therefore proceeded to map the expression of transcripts important in one-carbon metabolism. Both diets induced broadly similar changes in genes associated with choline, methionine and phosphatidyl choline (PC) metabolism Figure 3.8. MCDD resulted in significant dysregulation of two pathways which contribute to the maintenance of the universal methyl donor SAMe. Firstly, the conversion of choline to PC (Chkb, Pcyt1a) and secondly the conversion of methioine to SAMe (Mat2a). I went on to examine the expression of enzymes involved in SAMe clearance and found them to be suppressed in both interventions (Mthfd1, Gnmt, Achy, Dnmt3b). Given these intriguing expression findings focussed on SAMe maintenance, I measured hepatic SAMe concentrations using matrix-assisted laser desorption/ionization (MALDI). This demonstrated a significant reduction in SAMe in MCDD mice with no change in CDD mice (Figure 3.8).
Fig. 3.8. CDD and MCDD induce changes in the hepatic expression of genes important in one-carbon metabolism. (A) Transcriptional changes in one-carbon metabolism enzymes induced by CDD and MCDD diets (* = adjusted P value < 0.05, Benjamini – Hochberg analysis). (B) SAMe levels as measured by MALDI analysis. (* = P< 0.01 one-way ANOVA with Bonferroni post hoc analysis). (C) Metabolic interactions between methionine cycle, phosphatidylcholine synthesis and DNA methylation. Hatched (CDD) and grey (MCDD) arrows depict expression changes in each pathway. PE=phosphatidylethanolamine, PC=phosphatidylcholine, SAH=S-adenosylhomocysteine, SAMe = S-adenosylmethionine. Adapted from Li and Vance 2008.
3.3.3 Comparison of CDD and MCDD with stages of Human NAFLD

I then proceeded to correlate the data sets with large published expression profiles of two stages and one complication of human NAFLD: simple steatosis, NASH with high risk of progression, and HCC (Figure 3.9)\textsuperscript{283,298,299}. There were very few common dysregulated genes between either dietary intervention and human simple steatosis with no statistically significant correlation (Figure 3.9 A-C). Similarly, there were very few common genes altered between CDD and NASH datasets, although those that were differentially expressed in both did demonstrate a weak correlation, including the immune regulator Lgals3 and the post-translational modification protein Golm1 (Figure 3.9 D and E). However, MCDD demonstrated a much greater transcriptional similarity with NASH. 41 (12.3\%) of the NASH transcripts were also dysregulated in the MCDD livers with a highly significant correlation in the direction of transcriptional changes (P<0.001) (Figure 3.9 F). Functionally important genes common to both data sets included the lipid synthesis protein lipid synthesis protein Acsm3, fibrosis mediators \textit{Col4a2}, \textit{Col6a3} and \textit{Dcn}, negative inhibitors of Wnt signaling \textit{Bicc1} and \textit{Dkk3} and the oxidative stress buffer \textit{Gpx7}. When comparing both diets to HCC, there were again very few (90 transcripts (2\%)) commonly dysregulated genes in CDD and the HCC data sets (Figure 3.9 G and H). However, the greatest parity in all data sets was found between the MCDD model and human HCC (9.8\% of HCC dysregulated transcripts, P <0.001). Enriched GO terms within the common upregulated gene set between MCDD/HCC included \textquoteleft GO:0045931 - positive regulation of mitotic cell cycle\textquoteright. Enriched down-regulated common pathways included GO:0006631 “fatty acid metabolic process”, GO:0006629 “lipid metabolic process and GO:0008610 "lipid biosynthetic process". 
Figure 3.9. Cross-species comparison of CDD and MCDD transcriptional changes with human stages of NAFLD. Venn diagrams showing common and distinct dysregulated gene orthologues between CDD and MCDD livers and human hepatic steatosis (A), human NASH (D) and human HCC (G). Log2 fold change in common dysregulated orthologues between human hepatic steatosis (B+C), human NASH (E+F) and human HCC (H+I) on vertical axes. This is plotted against the corresponding transcript in CDD and MCDD. Trend line and p value calculated by linear regression analysis.
3.4 Discussion

The precise mechanism of hepatic injury in rodent methyl donor deficiency is incompletely understood. An accumulating body of evidence in human NAFLD suggests some common mechanism(s) which if understood could lead to novel therapeutic interventions and the identification of individuals vulnerable to disease progression\(^{292,303–305}\).

The predominant view is that sequestration of hepatic lipids in both diets is due to impaired export of triglyceride laden VLDL lipoproteins from hepatocytes. This may be due to an inability to synthesise phosphatidylcholine (PC), the major component of the VLDL lipid bilayer, with the liver most affected due to natural PC loss through biliary secretion\(^ {306}\). My sequence of experiments is consistent with the concept that impaired lipid export is the primary deficit. Furthermore, I introduce the concept that PC production is further sacrificed in an attempt to maintain SAMe levels and that suppression of the ER associated Ces enzymes may contribute to impaired VLDL assembly and lipid clearance.

Three key studies support impaired VLDL secretion in methyl donor deficient rodents. First, delayed clearance of radiolabelled palmitate is observed in the livers of CDD fed rats\(^ {307}\). Secondly, reduced VLDL, triglyceride and PC secretion is observed in hepatocytes isolated from CDD rats and finally reduced serum triglyceride accumulation has been demonstrated in the context of the peripheral lipase inhibitor typoxalol in MCDD mice\(^ {99,143,306}\). In addition, increased accumulation of radiolabelled oleic acid and increased incorporation of \(^{14}\)C into hepatic triglycerides suggests enhanced lipid uptake and/or de novo lipogenesis may also occur\(^ {99,143}\).

The evidence for impaired VLDL secretion appears consistent for MCDD rodent diets, however for choline deficiency alone studies are conflicting. The possibility that differing mechanisms exist for lipid accumulation in CDD and MCDD diets is supported by the observation that the zonal distribution of lipid droplets differs considerably as steatosis evolves\(^ {308}\). Two studies have demonstrated no reduction in serum triglyceride accumulation in choline deficiency alone following typoxalol administration\(^ {99,143}\). Furthermore, \textit{ex vivo} studies in isolated rat hepatocytes incubated with or without choline demonstrated no reduction in apolipoprotein secretion or PC synthesis\(^ {294}\). PC synthesis may be preserved in CDD due to the presence of an ancillary pathway, present only in liver, whereby PC can be directly synthesised from phosphatidylethanolamine (PE) by the enzyme phosphatidylethanolamine
N-methyltransferase (PEMT) using methionine as a methyl donor. Indeed, ~30% of rodent hepatic PC is synthesised in this way.

My transcriptional analysis of lipid synthesis in CDD and MCDD mice strongly suggests an appropriate compensatory response to excess hepatic lipid content, with a clear suppression of key mediators of fatty acid synthesis and elongation and cholesterol synthesis. This supports the concept that impaired lipid clearance rather than impaired de novo lipogenesis is responsible for the hepatic fat accumulation in both diets. Expression of the Ces enzymes (Ces1d, Ces1f, Ces3b) was markedly suppressed in MCDD and the liver predominant Ces isoform, Ces3b, was also suppressed in CDD. Ces enzymes are important regulators of VLDL lipid packaging and assembly in the hepatic endoplasmic reticulum (ER), and as such a reduction in expression would be expected to result in reduced hepatic lipid clearance (Figure 3.10). Liver specific knock out of Ces3b (also known as triglyceride hydrolase) in mice results in reduced circulating VLDL triglycerides and cholesterol levels on a standard chow diet and also altered hepatic lipid droplet morphology. Furthermore, liver specific knock-down of Ces1 in mice results in increased hepatic triglyceride content whereas Ces1 overexpression reduces hepatic triglyceride content and plasma glucose levels. Whilst it is unclear why the expression of these genes is suppressed in the presence of an increased hepatic lipid load (notably in MCDD), I suggest that these models may present an opportunity for investigating the function of these important therapeutic targets.
Figure 3.10 Transcriptional dysregulation mapped to pathways of lipid transport and metabolism in hepatocytes. Arrows demonstrate direction of transcriptional change in CDD (hatched) and MCDD (grey) diets. Pnpla3 and Ces suppression suggest impaired packaging of lipid into VLDL particles on the surface of the ER (black star).
The expression of Patatin-like phospholipase domain containing 3 (Pnpla3) was also suppressed in both CDD and MCDD models. The human PNPLA3I148M variant is strongly associated with human NALFD; humans homozygous for the PNPLA3I148M allele are reported to have ~73% more hepatic triglyceride when compared with matched heterozygote controls, and both in vivo and in vitro studies suggest that this is due to impaired triglyceride hydrolysis and VLDL export. Thus, Pnpla3 suppression may also contribute to triglyceride accumulation in methyl donor deficiency.

Detailed analysis of the interacting pathways involved in choline, methionine and PC metabolism in these models provides further insights into the mechanisms by which the murine liver responds to dietary choline and methionine deficiency. Whilst choline has a major role as a substrate for PC synthesis, the essential amino acid methionine is also necessary for the methylation of a large variety of substrates including DNA, proteins, lipids and for the synthesis of polyamines, and is crucial for normal hepatocyte function. Both substrates are important for the maintenance of hepatic SAMe levels, which are normally tightly regulated to maintain normal hepatic function. The transcriptional changes with choline and methionine deficiency strongly suggest a drive to maintain hepatic SAMe concentrations. Up-regulation of methionine adenosyl-transferase (Mat2a) which synthesises SAMe from methionine with a simultaneous downregulation of Chkb and Pcyta which consumes choline to generate PC suggests a drive to maintain SAMe levels. Furthermore, the concurrent down-regulation of Gnmt and Ahcy (which metabolise SAMe and SAH respectively) may act as a further buffer to prevent SAMe clearance. We subsequently found reduced levels of hepatic SAMe in MCDD mice in agreement with other studies suggesting an inability to maintain SAMe levels with severe deficiency of both substrates despite these compensatory changes. Therefore, the summation of the observed transcriptional changes in MCDD mice is directed at maintaining SAMe concentrations at the expense of PC synthesis, with the potential to result in decreased VLDL synthesis. In support of this theory, the deleterious effects of MCDD diets can be rescued by the administration of SAMe.

Whilst there are some clear biological and histological similarities between human NAFLD/NASH and the hepatic injury induced by methyl donor deficiency in rodents, the global metabolic phenotype is quite different. In humans, NAFLD is closely associated with obesity and insulin resistance, whereas in rodents, CDD results in profound hepatic steatosis without insulin resistance and MCDD causes an inflammatory steatohepatitis with fibrogenesis and significant weight loss with an increase in peripheral insulin sensitivity.
Our comparative analysis suggests that CDD and MCDD induce a hepatic transcriptional phenotype which is markedly dissimilar to human NAFLD in terms of lipid handling. Whereas human NAFLD is associated with an upregulation of genes important in de novo lipogenesis (FASN, MLPXL, ACACA, SREB-1c), this is either unchanged, or indeed the reverse is observed in mice maintained on CDD/MCDD diets. Furthermore, although some findings in human NASH support the concept that hepatic steatosis may result at least in part in from an inability to synthesise PC, no genes in the lipid pathways of interest were commonly dysregulated in the steatosis or NASH datasets. Finally, although SAMe depletion is a feature of human NAFLD and correlates with severity of disease in NAFLD biopsies, and oral SAMe preparations are currently under review as a treatment for chronic liver disease, mediators of SAMe metabolism were not altered in the human steatosis or NASH datasets. Some enzymes important in SAMe metabolism were dysregulated in the human HCC data (upregulated CHKα (LogFC 0.70); down-regulated MAT1A (LogFC -2.23), BHMT (logFC-2.90) and PEMT (LogFC -1.44)), however these did not reflect the changes seen in the mouse model apart from a similar change in the expression of GNMT (LogFC -2.37).

To my knowledge this is the first study to perform a global analysis of the transcriptome in these widely used models of human NAFLD and gives new insight into how the murine liver reacts to choline and methionine/choline deficiency. Furthermore, this is the first time that the CDD/MCDD transcriptome has been compared to the human disease these diets attempt to model. However, there are weaknesses to this study. Firstly, I have looked exclusively at transcription with no analysis of translated protein. Furthermore, I have yet to determine why Ces enzymes are suppressed in these models. Ces enzymes are targets of IRE1α-mediated mRNA decay which in turn are activated by ER stress and this is one possible mechanism. Alternatively, the nuclear farnesoid X receptor (FXR) is a key regulator of glucose and lipid homeostasis and is an upstream regulator of Ces expression. These potential mechanisms require further exploration. Logical future experiments would include examining how IRE1α hepatic conditional knock out mice respond to methionine and choline restriction. Alternatively, the use of an oral FXR inhibitor could be employed.

3.5 Conclusion

In conclusion, my data suggest a novel mechanism that may contribute to methyl donor deficient liver injury involving impaired VLDL particle assembly due to suppression of key
triglyceride hydrolysis proteins. Furthermore, although these diets are extensively employed to model human NAFLD, the global transcriptional profile associated with the development of liver disease in CDD and MCDD mice is highly dissimilar to that present in human hepatic steatosis. However, MCDD is transcriptionally representative of a premalignant inflammatory steatohepatitis. My findings suggest that although there are common features in the pathology induced by methyl donor deficiency and human disease (e.g. decreased VLDL export, reduced hepatic SAMe concentrations), the precise mechanisms leading to these effects may be very different. In light of these findings I elected to pursue an alternative murine model of metabolic liver disease for further epigenetic studies.
CHAPTER 4: HEPATIC TET EXPRESSION AND 5hmC DYNAMICS IN MURINE NAFLD

4.1 Introduction

Hepatic steatosis is the excessive accumulation of fats in the form of triglycerides and other lipid species in the liver and is strongly associated with overnutrition and obesity. This excess accumulation of lipids through as yet undefined mechanisms results in a blunted response to insulin action contributing to T2DM and a vulnerability to progressive end stage liver disease. Collectively these diseases place a huge burden on health resources. A detailed understanding of this process is required to address two key clinical needs. Firstly, alleviation of hepatic insulin resistance to reduce glucose levels and treat T2DM. Secondly, to obtain useful biomarkers for disease progression such that health resources might be concentrated on those with poor prognostic outcome.

The liver is a central organ in metabolic control with a key regulatory function in glucose and lipid metabolism. As such, hepatocytes must be able to perform rapid and stable changes in gene expression in reaction to the fluctuating delivery of and demand for these nutrients. One mechanism by which this may occur is through dynamic regulation of transcription that may include perturbations of chromatin structure in response to molecular stimuli. There is an accumulating body of evidence that dysregulation of DNA methylation (5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC)) may have a role in the generation and progression of NAFLD. Methyl donors are centrepieces of one-carbon metabolism that have a key role in transmethylation reactions, and thus in epigenetic and epigenomic regulations. In mouse studies, dietary restriction of methyl donors or genetic impairment of methyl donor metabolism causes liver injury similar to NAFLD. Furthermore, dietary supplementation of methyl donors does appear to protect rodents from high fat/sucrose diet induced hepatic steatosis. More recently, two human studies have examined global methylation dynamics across the NAFLD spectrum from hepatic steatosis to advanced NASH. Both studies utilised bisulphite converted DNA hybridised to the Illumina Infinium 450k array platform to give a single base pair resolution map of methylation at targeted CpG dinucleotides across large patient groups (n = 63 and n = 90). Both studies showed methylation changes in theoretically causative but different gene sets with an inverse correlation to gene expression. Furthermore, Ahrens et al demonstrated at least partial reversibility of these epigenetic changes following bariatric surgery and resolution of the liver injury. Taken together, there is now accumulating evidence in both human and mouse, that
dysregulation of the DNA methylation may play an important but as yet undefined role in NAFLD. This is an exciting new field of translational research in a global health problem.

5hmC is a recently identified cytosine modification that accounts for approximately 0.5% of cytosine species in hepatic DNA. The Ten Eleven Translocase (Tet1, 2 and 3) proteins catalyse the addition of molecular oxygen to 5mC to generate 5hmC perhaps as the first step in a demethylation pathway, and as such may facilitate the removal of 5mC from genomic DNA, permitting access of transcriptional machinery and regulatory proteins and stimulating transcription. Tet function is intrinsically linked to glucose metabolism through two defined mechanisms. Firstly, Tet enzymes form complexes with, and are post translationally modified by, O-linked N-acetylglucosaminyItransferase (OGT) with the addition of N-acetylglucosamine (GlcNAc) to serine and threonine residues in Tet regulatory regions.

The substrate of this reaction, UDP-GlcNAc, is the end product of the nutrient sensing hexokinase pathway, which is reliant on the availability of other key intermediates ATP, glucose and Acetyl-CoA for UDP-GlcNAc production. The second observed link between Tet proteins and metabolic balance is through direct interaction of Tet proteins with TCA cycle metabolites: α-ketoglutarate is an essential co-factor for normal Tet function however others (succinate, fumarate and 2-hydroxyglutarate) are competitive inhibitors. Cumulatively, these observations have led to the theory that Tet proteins may act as a molecular ‘sensor’ regulating cellular function in adaptation to energy supply and demand.

Despite a breadth of knowledge surrounding DNA methylation and NAFLD, the role of Tet proteins and 5hmC regulation in this process is lacking. We therefore set out to examine the role Tet mediated hydroxymethylation may play in the pathogenesis of NAFLD.
4.1.1 HYPOTHESIS

NALFD is associated with altered 5hmC at pathologically important genes involved in disease progression. Mechanistically this may be due to alterations in key metabolites directly affecting Tet function.

4.1.2 AIMS:

1) Generate a murine model of NAFLD that represents the clinical, metabolic and histological features of the disease using diet-induced obesity (DIO).
2) Reverse this phenotype in a parallel mouse cohort through conversion to a standard control diet.
3) Perform genome wide analysis of 5hmC enrichment in whole liver DNA in DIO mice and compare this with array based expression profiles.
4) Determine whether or not these DNA modification changes are reversible in the parallel dietary reversal cohort.
5) Analyse by mass spectrometry the relative whole liver concentrations of cellular metabolites known to affect Tet function as a possible mechanism for any observed change.
4.2 Methods

4.2.1 Animals

All experiments were carried out under a UK Home Office license and with local ethical committee approval. Adult C57BL/6J mice (Charles River, Tranent, UK) were maintained under controlled conditions as (2.3.1). Mice were maintained on control (C) or High Fat Diet HFD (2.3.1) for 17 weeks and then killed and tissues collected and used for histology (2.4) or snap-frozen and stored at -80°C. A second cohort of mice was maintained on identical diets, however after 17 weeks the HFD was converted to a control diet and mice were maintained on this for the remainder of the experiment (29 weeks total) (Figure 4.1).

Figure 4.1. Schematic representation of experimental design.

4.2.2 Histology staining and NAFLD Score

Livers were removed and sections were fixed in formalin 10% solution and mounted in paraffin blocks prior to staining with haematoxylin and eosin or picosirius red (2.4). NAFLD...
scoring was performed blinded using a validated human scoring system by a consultant pathologist.

4.2.3 Gene Expression

RNA was extracted from snap frozen liver tissue using Qiazol and the RNase kit (Qiagen, Manchester, UK) as per 2.6.1. RNA was quantified, DNase treated and reverse transcribed (2.8.1). Quantitative real time PCR (2.6.1) was performed on each treatment cohort for key genes in pathways of cholesterol synthesis (Methylsterol Monooxygenase 1, Sc4mol; Mevalonate Diphosphate Decarboxylase, Mvd; Farnesyl Diphosphate Synthase, Fdps; 7-Dehydrocholesterol Reductase, Dhcr7; Lanosterol Synthase, Lss; Cytochrome P450 family 17 subfamily A member 1, Cyp17a1; Squalene Epoxidase, Sqle). Gene expression is presented relative to the mean of two housekeeping genes (TATA Box Binding Protein, TBP and Beta Actin, ActB).

4.2.4 Transcript analysis

RNA (n = 4 control diet and n = 8 HFD fed) was biotin labelled and submitted for single colour transcriptional array analysis at the Wellcome Trust Clinical Research Facility, Western General Hospital, UK (2.8.1). RNA labelling was performed on 500ng RNA using the Illumina Total Prep RNA amplification kit (2.6.1) and subsequently hybridised to Illumina Mouse-ref6 expression bead arrays (2.6.1). Intensity data were generated using a HiScan array scanner (Illumina, San Diego, USA) and analysed using iScan Illumina software. Data analysis and generation of plots were performed in RStudio (http://www.rstudio.com) with R version 3.1.2. Data import, quality control, normalisation and between array adjustment was performed using Lumi package and differential expression was determined using Limma package (Bioconductor.org). Unsupervised clustering was performed using Euclidean distance.

4.2.5 5-hydroxymethylcytosine Slot Blot

Slot blot analysis was performed as described (2.6.2). Serial dilution of extracted DNA were spotted onto Hybond N+ nitrocellulose membrane (Amersham, Buckinghamshire, UK) using a slot blot manifold (HSI, GE Healthcare, Buckinghamshire UK), washed in 2xSSC buffer and dried overnight. Membrane was then probed with 1:10000 anti 5hmC antibody (Active Motif, cat: 39769) for 1 hour at room temperature followed by goat anti-rabbit IRDye 800CW (1:10000) for 1 hour at room temperature and subsequent imaging on the Li-cor Odyssey.
infrared Imaging system (both Li-cor, Cambridge, UK). DNA loading was confirmed by staining with 0.02% methylene blue in 0.3 M sodium acetate (pH 5.2) and destaining with ddH$_2$O.

4.2.6 5-hydroxymethylcytosine DNA Immunoprecipitation (hme-DIP)

DNA extraction was performed using a modified version of the Qiagen DNeasy Blood and Tissue kit (Crawley UK) protocol and RNase treated with RNase A (Purelink, Ambion, UK). DNA was quantified with the Qubit 2.0 Fluorometer (Invitrogen, Paisley UK) using DNA “broad range” standards as per manufacturer’s instructions for DNA samples. DNA integrity was determined by resolving extraction products on a 1.5% agarose gel and staining with gel red (Cambridge Bioscience, Cambridge UK) 0.005% final concentration. 10μg genomic DNA was sonicated (Bioruptor, Ougree, Belgium) to a mean base pair length of 275bp. Acceptable sonication was deemed as fragments between 100 and 600 base pair length with a mean of 250 to 300 base pairs and a uniform appearance between sonicated samples. 2.5μg of sonicated DNA was denatured for 10 minutes at 90°C in a heatblock before cooling for 5 minutes at 4°C. 50ul of 10xIP buffer (2.2) was then added and vortexed briefly. 50ul ‘input’ DNA was then removed and placed at 4°C until the DNA clean up step below. 1ul of anti 5hmC antibody (Active Motif 39769) was then added to the remaining sample and incubated for 3 hours at 4°C.

40ul of Dynabeads protein G (Invitrogen, Paisley, UK) was prewashed with BSA 0.1% in PBS and added to the DNA/antibody mixture for 1 hour at 4°C. Beads were then collected on the magnetic rack, supernatant discarded and washed three times with 1ml of cold 1xIP buffer (2.2). Beads were then resuspended in 250ul of digestion buffer (2.2) and treated with 20ul of proteinase K 20mg/ml (Roche, UK) in a thermostaker at 1000rpm, 55°C overnight. Beads were then removed by magnetic rack and both enriched fraction and input samples were then purified using the Qiagen Qiaquik PCR purification kit (Qiagen, UK) with elution in 22μl of water. Quantification of enrichment was determined using qPCR (2.7.2) and (2.8.3).

Raw sequencing data was quality controlled, filtered and aligned using the Ion Torrent suite software (Life Technologies, Paisley, UK) and then normalised to total reads in R using bespoke scripts (2.7.2). Relative hydroxymethylation levels per 150bp window were determined using the ‘sliding windows’ function on the Galaxy server at IGMM, Western Weneral Hospital, UK (see 2.7.2).
2.6.7 Liquid Chromatography Mass Spectrometry (LCMS)

5ug of frozen tissue was ground up in a mortar and pestle in 200ul of pre cooled LCMS extraction buffer (2.2). Samples were then spun at 16000g 4°C for 5 minutes and supernatant collected and stored at -80°C. LCMS analysis was outsourced to The IGMM Mass spectrometry laboratory, Institute of Genetics and Molecular Medicine, University of Edinburgh.

4.2.8 Statistics

Animal model and qPCR statistical analysis was performed using SPSS statistics (Illinois, USA). Data were routinely analysed for outliers, normalisation and sphericity where required. Non-parametric data were either log transformed or a non parametric test used as indicated.
4.3 Results

4.3.1 Mouse Phenotype

HFD fed animals gained significantly more weight than control animals with increased retroperitoneal and epididymal fat pad weight (Figure 4.2). Serum biochemistry and glucose tolerance testing revealed fasting hyperglycaemia, glucose intolerance, hyperinsulinaemia and insulin resistance as estimated by the HOMA-IR index (Figure 4.2). HFD feeding induced a variable effect on liver histology within the HFD population. Broadly, a macrovesicular steatosis with predominantly periportal distribution was seen with a minimal increase in hepatic fibrosis and a higher NAFLD score (Figure 4.3). Withdrawal of HFD prompted a striking and rapid weight loss in the HFD reversal group with convergence of weights of the ‘HFD reversal’ and ‘control reversal’ cohorts within a 2 week period. Interestingly, the weight loss in these animals then stabilised and weight followed a parallel trajectory to the control animals until termination of the experiment. At 29 weeks, there was a reversal in the metabolic derangement in the HFD fed reversal animals with comparable glucose tolerance and insulin sensitivity. Furthermore, there was a considerable improvement in liver steatosis, although the level of lobular inflammation remained (Figure 4.3).
Figure 4.2 HFD induces excess weight gain (A) with increased gonadal and retroperitoneal fat pad weight (B) after 17 weeks. HFD fed mice also demonstrated fasting hyperglycaemia and glucose intolerance (C) with an increase in insulin resistance as estimated by HOMA-IR index (D). * P < 0.05, ** P < 0.001 Student T test. Error bars = +/- SEM.
Figure 4.3: Histological analysis of 17 week HFD animals. DIO induced a macrovesicular steatosis (black arrows) with only minimal change in fibrosis (A). There was a significant variation in the histological injury induced by DIO as determined by validated NAFLD scoring.
system (B)\textsuperscript{24}. Reversal of HFD resolved steatosis in all but one mouse. H+E = haematoxylin and eosin, PSR = picosirius red. Black line in scatter plot = mean per group.

4.3.2 Derangement and Reversal of Key Genes in Cholesterol Synthesis and Lipid Transport.

Transcriptional analysis of over 45K transcripts was performed using Illumina MouseWG-6 v2.0 Expression BeadChip, which allows for parallel processing of 6 cRNA samples. Control (n = 4) and HFD (n = 8) were analysed on two WG-6 chips (2 controls and 4 HFD per chip). There was a significant batch effect between BeadChips necessitating between assay correction using the ComBat function (sva package, Bioconductor.org). Following this, analysis of the top 200 varying transcripts was sufficient to cluster individuals by Euclidean distance with minimal variance between individuals in each group (Figure 4.4). Assuming an adjusted p value of <0.05 and a transcriptional threshold of 1.5 fold, 43 transcripts were upregulated mapping to 32 annotated genes with only 6 down regulated transcripts mapping to 5 annotated genes (figure 4.4, See gene list and fold changes in table 1 Appendix 1). Of the upregulated transcripts the most enriched pathway was cholesterol biosynthesis with 8 of 32 genes directly involved in this process (Figure 4.4 B). Notable down regulated transcripts were the HCC associated tumour suppressor genes Txnip and Osgin 1 and the circulating protein Fgf21, analogues of which show real promise as novel therapies in the metabolic syndrome\textsuperscript{330-332}.

The Tet gene family has relatively low hepatic expression and as such were not detectable on the microarrays. qPCR analysis of Tet1 isoforms in both liver and adipose tissue demonstrated no transcriptional change in liver with only a trend to Tet1 suppression in adipose tissue (Figure 4.4).
Figure 4.4. Global and TET specific expression changes in response to 17 week HFD feeding. The top 200 genes were sufficient to cluster by dietary intervention with highly consistent changes between groups (A). Of the 32 upregulated genes in HFD 8 were directly related to cholesterol biosynthesis and transport (labelled) (B). (C) qPCR analysis of Tet transcript levels. Low total expression gave very high assay variability. There was a trend to suppression of Tet1 in adipose tissue with HFD. * P = 0.08, Error bars are 10-90th percentile.
4.3.3 High Fat Diet Induced Reversible 5hmC Enrichment in Genic Regions of Cholesterol Synthesis and Transport genes without changing global 5hmC content.

Global 5hmC content was first examined using immune ‘slot-blotting’. I modified this technique to utilise LI-COR infrared emitting secondary antibodies with intensity quantified on the Odyssey Imager (Both LI-COR, Cambridge, UK) reported to give a more accurate quantitative analysis \(^{263}\). Liver DNA from n=4 per group was analysed in serial dilutions with no differences identified between treatment groups (Figure 4.5).

I employed an antibody affinity based enrichment technique coupled with next generation semiconductor sequencing on the Ion Torrent platform (Life Technologies, Paisley, UK) as a reproducible and validated method of profiling 5hmC enrichment \(^{333}\). I performed sequential optimisation of antibody concentrations, incubation times and amount of DNA used. I concluded that exposure of 2.5μg of sonicated DNA to 1ul antibody (active motif Cat no 39769), for 3 hours at 4°C yielded good enrichment of positive controls with minimal retention of negative control fragments (Figure 4.6 A). Control (n=2) and HFD (n=4) samples were sequenced to a depth of ~30 million usable reads with a mean usable read length of 150bp and >98% read accuracy at read lengths >50bp.

Using annotation sites from the ENCODE database I then examined mean 5hmC content through each annotated region. As previously reported, I demonstrated relative 5hmC enrichment in gene bodies and liver enhancer regions with lesser enrichment in promoter sequences and minimal 5hmC reads across transcriptional start sites and intragenic regions \(^{257,333}\). Furthermore, this profile was highly reproducible with no global change in 5hmC profile between dietary interventions (Figure 4.6 B).

Sliding window analysis across all genes throughout the mouse genome showed a high degree of reproducibility between individuals with no significant difference between groups. However, when examining the difference in read counts across 150bp windows in only up and down regulated genes, there was a clear suggestion of 5hmC gain in genic regions of upregulated genes with 5hmC loss in down regulated genes (Figure 4.6).
Figure 4.5: Global 5hmC levels are unchanged between 17 week HFD and control animals. 5hmC ‘slot blot’ analysis of extracted DNA in serial dilutions, n=4 per group (A). Oligonucleotides containing each cytosine species are used as experimental controls. (B) Each blot normalised to methylene blue loading control and expressed relative intensity. (Error bars +/- 1 SEM).
Figure 4.6: hmC-DIP-seq. Quantitative PCR following DNA immunoprecipitation demonstrated enrichment of positive controls with minimal enrichment of negative controls (A). (B) Sliding window analysis of mean 5hmC levels normalised to all annotated genes in each group shows no global difference between diets. (C) Total normalised reads mapped to each annotated region suggests high levels 5hmC in liver enhancers, promoter regions and gene bodies. Error Bars = +/-1SD. * = P < 0.001 vs intragenic regions Mann-Whitney test.
Having established the reproducibility of the 5hmC-DIP-seq protocol and with global findings similar to those previously described \(^{334,335}\), I went on to test my hypothesis by examining changes in 5hmC levels in annotated regions of genes with a functional relevance to NAFLD pathogenesis. Interestingly, sliding window analysis of all upregulated genes demonstrated clear genic 5hmC enrichment in upregulated genes involved in lipid synthesis and transport versus induced genes involved in other biological processes (Figure 4.7 A). I then went on to examine the relationship between transcriptional change with mean difference in 5hmC-DIP-seq reads per 150bp window between control and HFD fed animals. There was a highly significant correlation between mean genic 5hmC enrichment and transcriptional change that appeared to favour genes of lipid synthesis and transport over and above other induced genes (Figure 4.7 B). No such association was seen with other annotated regions (promoter sequences, enhancers, TSS regions; data not shown). This finding was subsequently confirmed by boxplot analysis (fig 4.7 C).
Figure 4.7 A-C) 17 week high fat diet is associated with increased genic hydroxymethylation predominantly in genes of cholesterol synthesis and transport. A) sliding window analysis suggests preferential genic enrichment in genes of lipid synthesis and transport versus all induced genes. Scatter plot B) and boxplot C) demonstrate highly significant correlation between mean genic 5hmC levels and transcriptional change with significant discrepancy between 5hmC change in all induced genes versus those specifically involved in lipid synthesis and transport. Scatterplot correlation by linear regression analysis. (blue = down regulated genes, red = up regulated genes, yellow = lipid synthesis/transport genes) * = P < 0.05 Mann-Whitney Test.
4.3.4 Genic 5hmC Enrichment as a Marker of Transcriptional Activity

It has been proposed that, given the transcriptional correlation between expression and 5hmC enrichment, studies of this type could be used to further our understanding of gene activity and thus cell phenotype, over and above that found in microarray or RNA-seq transcriptional analysis. I therefore proceeded to examine gene ontology analysis of gene sets with at least a 1.5 fold increase in mean genic 5hmC with HFD intervention (Figure 4.8). When corrected for multiple testing, no pathways were significantly enriched, however the most over represented GO terms in genes gaining 5hmC included ‘response to external stimulus’ and ‘carbohydrate biosynthetic process’. To further confirm genic 5hmC as a marker of transcriptional activity I proceeded to demonstrate that these changes are reversible with a rescue of the disease phenotype (Figure 4.9). Microarray transcript levels of induced mediators of cholesterol synthesis were successfully validated by quantitative PCR. Reversal of the HFD phenotype resulted in complete reversal or significant amelioration of these transcriptional changes (Figure 4.9 A). Candidate peaks in genic regions of cholesterol synthesis mediators were then validated by hmC-DIP qPCR and this enrichment was similarly reversed with phenotype rescue (Figure 4.9).
Figure 4.8: Genic 5hmC levels as a marker of transcriptional activity. (A) Direct comparison of genic 5hmC levels between control and HFD treated animals. Dashed lines represent genes with 1.5 fold increased (red) and reduced (blue) genic 5hmC levels. Gene ontology analysis of enriched gene sets (B)

<table>
<thead>
<tr>
<th>GO Term</th>
<th>P Value</th>
<th>Genes</th>
<th>Fold Enrichment</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0009605 response to external stimulus</td>
<td>9.81E-05</td>
<td>NANO, TLR2, NANOG, LY6G6F, HPS6, LY96, NLA4A2, CHST2, SAA3, TLR2, GIA1, CXCR2, GPR77, CNGB1, NLRP3, SLC10A3, CD180, CHST1, FO5, CYP27B1, SAA2, SAA1, CXCL16, CNR2, CXCL15, CHRNA10</td>
<td>2.43</td>
</tr>
<tr>
<td>GO:0008033 tRNA processing</td>
<td>0.0052</td>
<td>HSD17B10, METTL1, TRMT12, TRMT5, TYW3, FDXACB1</td>
<td>5.32</td>
</tr>
<tr>
<td>GO:0048514 blood vessel morphogenesis</td>
<td>0.0405</td>
<td>ZFP36L1, KL5F, CTGF, EDN1, HEY2, GIA1, ADRA2B, NXX2-5</td>
<td>2.51</td>
</tr>
<tr>
<td>GO:0030154 cell differentiation</td>
<td>0.01711</td>
<td>NANO, EID1, GMPR2, GLDN, HPS6, EDN1, GIA1, EHF, CD1D2, ZFP3611, ADRA3, CDKN2C, GM11517, CTGF, UPK1A, HEY2, NXX2-5, CEBPA, NANO, H2-M3, RXFP1, LGALS3, STAP1, TBX2, FOXA1, NRA4A2, KRTDAP, EOMES, MIXL1, SFRP5, DLX6, KAZAD1, CFL1, NLE1, UBD, NEUROD1, FABP4, PRDM1</td>
<td>1.46</td>
</tr>
<tr>
<td>GO:0016051 carbohydrate biosynthetic process</td>
<td>0.0443</td>
<td>MGAT2, TPI1, B3GALT6, CHST14, PCK2</td>
<td>3.74</td>
</tr>
</tbody>
</table>
Figure 4.9 Genic 5hmC enrichment reverse with resolution of the phenotype. Quantitative PCR validation of transcript levels (A) and hmC-DIP enriched loci (B) in mediators of cholesterol biosynthesis in control and HFD mice with complete or near resolution following reversal of HFD. * P < 0.05 Student’s T test versus 17 week control diet fed animals.
4.3.5 Altered TCA Metabolites.

Given the findings of increased 5hmC in the genic regions of activated genes suggesting increased Tet activity, I proceeded to determine whether or not this could be due to changes in relevant TCA metabolites (Figure 4.10). LCMS analysis was therefore performed on four key metabolites with a putative effect on Tet activity\textsuperscript{230,247}. When normalised for intra-assay variability, there was a small significant reduction in fumarate levels in the HFD treated group. Paradoxically, 2-OH glutarate, another known Tet inhibitor and putative oncometabolite\textsuperscript{336} displayed a trend towards increase in HFD fed animals.

![Figure 4.10](image-url)

Figure 4.10. Fluctuations in relevant metabolites to Tet function in the context of HFD. * p < 0.05, independent student T test. All data normalised to median value per metabolite measured.
4.4 Discussion

Our data supports the hypothesis that high fat diet feeding induces 5hmC enrichment in pathologically important genes and in NAFLD pathogenesis. This occurs independently of a change in Tet1,2 and 3 expression. Thus far the only known source of 5hmC is through oxidation of 5’-methylcytosine by the Tet isoforms. It is reasonable to assume therefore the 5hmC enrichment observed is due altered Tet activity at target loci.

4.4.1 DIO model Performance.

Small fluctuations in micronutrients can have a profound effect on mouse phenotype and as such dietary intervention studies must be carefully controlled 337. A variety of high fat diet compositions are available which principally aim to deliver a high percentage of calorie content through either animal or plant based saturated or unsaturated fats. We chose the Surwit diet which provides 58% of its calorific value through saturated fats (predominantly hydrogenated coconut oil) with a control diet matched in all micronutrients but providing the majority of its calorific value through the disaccharide sucrose. This approach delivered the expected phenotype in male C57Bl/6J mice of visceral adiposity, fasting hyperglycaemia, glucose intolerance and insulin resistance as compared to control fed mice. Furthermore, the liver histology of these mice was typical of human NAFLD as scored by a consultant pathologist using a human validated scoring system 24. The variation seen in phenotype within the mouse population on HFD is well described and has been attributed to behavioural hierarchy, epigenetic variation, gut biota and lactation litter size 142,338. The reversal of this phenotype by conversion to control diet at 17 weeks was striking and occurred as previously described 339. Interestingly, the liver histology in the HFD reversal mice showed that all but one mouse had no identifiable steatosis suggesting considerable resolution for hepatic fat accumulation; however there was no change in the level of lobular inflammation suggesting that once the onset of murine NASH is established in this model, dietary restriction is not curative. Given the clinical, biochemical and histological similarity of this model to human disease, it deemed adequate to proceed with detailed transcriptomic and DNA modification studies.
4.4.2 The Murine NAFLD transcriptome

Expression changes in the DIO model were modest with a maximum 3 fold change in a relatively small number of genes (n=32). This however is also a feature of transcriptomic studies in human hepatic steatosis. Previous reports have described either absent or significantly differing expression changes in patients with hepatic steatosis versus control subjects which may reflect the heterogeneity of the populations sampled and differing diets between regions globally. The isogenic nature of our cohort in addition to the age and diet matched nature of the intervention allowed us to detect the dysregulation of a key pathway in metabolic disease, cholesterol metabolism in addition to inflammatory mediators (\textit{Cxc11, Antrx2, Psmb9}) and potential tumour suppressors (\textit{Txnip, Osgin1}) identified in human hepatocellular carcinoma.

4.4.3 Murine NAFLD Drives Genic 5hmC Enrichment within Key Cholesterol Synthesis Mediators

Aberrant cytosine methylation is a feature of carcinogenesis typically exhibiting globally hypomethylated DNA with paradoxical hypermethylation of tumour suppressor genes as well as mutations in a number of epigenome modifying enzymes. As such the majority of epigenetic research in adult disease is in the cancer biology field. The focus has predominantly been identification of biomarkers for aggressive disease subtypes and pioneering novel therapies such as the methylation inhibitors 5'azacytidine and decitabine which are approved for some haematological malignancies. In recent years there has been considerable interest in the conversion of 5mC to 5hmC through the Tet isoenzymes in carcinogenesis. This fuelled by the observation that Tet inactivation either through loss of function mutation, or though upstream effector mutation (\textit{IDH1} and \textit{IDH2} mutations) are frequently found in haematological cancers and glioblastoma tumours. Furthermore, a global reduction in 5hmC is a hallmark of many solid tumours. However, despite the described interaction between Tet activity and cellular metabolism, there are no published works examining this key regulatory process in the field of metabolic disease.

As described above, there are some observational data in human cohorts suggesting aberrant methylation patterns exist in human patients with differing stages of NAFLD. These were found over both gene body and promoter regions and are suggested to be partially reversible. There is however no understanding of the mechanism behind these changes, whether they represent changes in cell population composition or cell intrinsic alterations; whether or
not they are of functional importance or merely a correlative finding. Furthermore, the role of Tet activity in 5mC regulation has yet to be explored in this highly prevalent disease. This study is therefore designed to address this.

Our data support the hypothesis that high fat diet feeding modifies Tet activity in vivo at target regions without altering Tet expression. This is manifest as enrichment in genic 5hmC over induced genes with a functional role in cholesterol synthesis (Lss, Mvd, Dicer7, Msmo1, Sgle, Hsd17b7, Fdps, Cyp17a) and the apolipoprotein Apoa4. Interestingly, genes which can be assumed to be consequent of, rather than contributory to, lipid accumulation did not demonstrate genic 5hmC enrichment. These included oxidative stress response genes (Gst3), inflammatory mediators (Cfd, Cxcl1, Zfp36) detoxification enzymes (Aldh1b1, Cyp1a2, Cyp2a5). This finding clearly links Tet activity with lipid accumulation rather than as a non-specific bystander reaction. The further observation that 5hmC changes appear reversible following rescue of the mouse phenotype strengthens the case for a functional role. Finally, fumarate is a known competitive inhibitor of Tet activity and as such this reduction may represent one of many mechanisms for Tet mediated genic hydroxymethylation in the context of HFD. The trend towards an increase in 2OH glutarate is interesting and suggests a complex interplay between relative metabolite abundance and Tet activity.

Which Tet Isoform is likely to be the key mediator in this process? Tet1 expression is highest in ESCs and reduced rapidly with differentiation to low levels in differentiated tissue. Our qPCR analysis would support Tet2 and Tet3 as the dominant hepatic genotypes. However, analysis of adult mouse liver from Tet1 depleted animals shows marked global hypo-hydroxymethylation in both promoter and genic regions suggesting an ongoing requirement for Tet1 in the differentiated liver epigenome. With respect to the annotated regional activity of each isoform, evidence from Tet1 and Tet2 knock down in mESCs suggests that Tet1 acts predominantly over TSS regions with Tet2 depletion causing a greater reduction in genic exon boundaries in highly expressed genes. This however may not be the case in every cell type and thus any or a combination of the Tet isoforms may be responsible for the observed changes here. Further studies of tissue specific Tet1-3 depleted mice challenged with HFD may be informative. The trend towards Tet suppression in gonadal adipose tissue is intriguing. Interestingly, Tet1 suppression was recently noted in subcutaneous fat in patients with T2DM when compared to a monozygotic twin without T2DM. Furthermore the role of Tet1 in adipocyte differentiation has been suggested through interaction with Ppar in mouse...
NIH-3T3 cells. The role of Tet activity in adipocyte differentiation and stability in differing metabolic states warrants further research.

4.4.4 5hmC as a marker of tissue state

5hmC analysis has potential as a biomarker of disease. Global loss of 5hmC is well described in a number of solid tumour types including prostate, breast, colon, lung, liver, kidney and melanoma and correlates with poor prognosis in astrocytoma and melanoma case series. Locus specific 5hmC profiles are highly tissue specific and are sufficient to independently clusters samples by tissue type. In addition, single base resolution 5hmC profiling of pancreatic tumours identified a specific redistribution of 5hmC with enrichment over gene bodies and promoters of over expressed genes. Hahn et al studied 5hmC levels in mouse neurogenesis and demonstrated that 5hmC predominated in the intragenic regions of genes with the highest transcriptional activity and specific enrichment in activated genes critical for neural differentiation. Additionally, challenging mice with phenobarbital (PB), a non-genotoxic hepatic carcinogen, induced defined 5hmC changes in promoter and proximal genic regions that were highly consistent between individuals and were maintained or expanded with higher duration of exposure. These findings suggest that 5hmC profiling may be a future diagnostic tool to determine tissue of tumour origin, prognostic outlook and perhaps response to treatment however larger case series, other tumour types and multivariate analysis to validate this claim are thus far lacking.

These data suggest that 5hmC profiling may be of similar benefit in the field of metabolic liver disease. Our initial analysis suggests that 5hmC was enriched in genes with a functional role in cholesterol synthesis and transport, however subsequent review of genes enriched in genic 5hmC but not dysregulated on the array identified highly relevant functional pathways and mediators. These included ‘inflammatory response GO:0006954’ containing chemokines Cxcl15 and Toll like receptors Tlr2 and Tlr12, ‘Cell differentiation GO:0030154’ and ‘Carbohydrate biosynthetic process GO:0016051 comprising three of the top five enriched pathways. It is known that highly protracted feeding in these mice will induce a NASH phenotype and as such it is tempting to speculate that the epigenome signature is predicting disease progression. Many non-invasive stratification systems have been evaluated for NAFLD staging however biopsy with histological grading remains the gold standard. This however has limitations due to the lack of a universally agreed scoring system, biopsy sampling error and intra-observer variability. Studies suggest that up to 5% of patients...
with NASH develop end stage liver disease and display a significant increase in liver related mortality \(^{47,53,360}\). The ability to target and appropriately screen high risk individuals is therefore paramount.

### 4.4.5 Experiment Limitations

This study is restricted by the available technology for analysis of 5hmC modified DNA and the statistical and bioinformatics methodology for analysis of this data type. DIP offers a reproducible method of 5hmC analysis but has two significant limitations. Firstly, denaturation of the double helix is required to allow access of the 5hmC antibody, The final product therefore is single stranded DNA unsuitable for semiconductor sequencing and typically of very low quantity. Parallel whole genome amplification of both the immunoprecipitate and an input sample is therefore required which cannot be guaranteed to be uniform. To limit this we normalised each DIP to a mean input for all samples. Secondly, sonicated DNA for this process is in the region of 250bp and thus the resolution of the data output is restricted to this length making analysis of hydroxymethylated CpGs in close proximity impossible. Single based resolution 5hmC profiling is possible using either oxidative bisulphide sequencing (oxBS-seq) or Tet assisted bisulphide sequencing (TAB-seq) however these are prohibitively expensive \(^{268,361}\). Finally, a consensus on bioinformatic and appropriate statistical analysis in these methods is lacking and thus far a bioconductor package to streamline and standardise processing of this data type is unavailable making comparison between studies challenging.

### 4.5 Conclusions

To our knowledge this is the first study to examine the activity of Tets and the active generation of 5hmC in the context NAFLD. These data demonstrated clear 5hmC enrichment over gene bodies of cholesterol synthesis and transport mediators over and above other induced pathways and that these changes are reversible with rescue of the disease phenotype. The possibility that regulation of Tet activity plays a functional role in human hepatic steatosis now warrants further research.
CHAPTER 5: EXAMINATION OF THE NAFLD HYDROXYMETHYLOME IN ESC DERIVED HEPATOCYTES.

5.1 Introduction

In chapter 4 I demonstrated perturbations in the hepatic hydroxymethylome in response to high fat diet feeding in rodents. While informative there are limitations to this approach. Firstly, lipid metabolism in mice and humans differs considerably and as such the applicability of this model to human disease is unclear. Secondly, histologic analysis of these mice demonstrated a degree of lobular inflammatory infiltrate and thus examination of whole liver may be confounded by a change in cell population. Therefore I set out to validate my findings in a pure human hepatocyte single cell population in vitro.

The use of established hepatocellular carcinoma cell lines are unsuitable for this study. It has been widely reported that 5hmC is rapidly lost during somatic cell adaptation to culture, including in all tested hepatocyte cell lines, the reason for which remains elusive \(^{263,362}\). In addition, methylation and hydroxymethylation changes are a feature of hepatocellular carcinoma in vivo which would confound any analysis following NAFLD intervention \(^{227,345,363–365}\). Furthermore, the use of cancer cell lines in modelling diseases of overnutrition can be questioned as malignant transformation requires profound changes in metabolic function to sustain rapid cell proliferation with a limited blood supply \(^{146,366}\). It was therefore necessary to establish an alternative model.

The use of primary hepatocytes isolated from human tissue has been employed to model human disease and predict drug toxicity \(^{367}\). In these cells, 5hmC has only been assessed at a global level by immunofluorescence \(^{368}\). The purification of human hepatocytes can be technically challenging and the effect on the hydroxymethylome is unknown. Furthermore isolated hepatocytes give a short time window for intervention, since they rapidly lose functionality, although novel substrates show promise in extending this \(^{369}\). In addition to this, the hepatic phenotype and pharmacological history of the donor may confound 5hmC analysis \(^{266,281}\).

Embryonic stem cells (ESCs) possess the key properties of self-renewal and the ability to differentiate into any cell type. They are therefore a potentially limitless source of model tissue to study in vitro. It is now possible to generate human hepatocyte-like cells from ESCs, which in addition to hepatocyte morphology and a ‘hepatocyte-like’ transcriptome, possess cytochrome p450 activity, secrete albumin, and when grown on recombinant hepatic
extracellular matrix, show a degree of self-organisation into hepatic sinusoids. ESC derived hepatocytes (ESCheps) have been employed in the mechanistic analysis of drug induced liver injury, the study of hepatitis viral replication and the characterisation of inherited disorders of lipid metabolism. Furthermore, at terminal differentiation, ESCheps do not proliferate and as such may potentially retain normal 5hmC levels and warrant further exploration as a model in this study.

Recapitulating the complex multi-faceted aetiology of NAFLD in vitro is challenging. Exposure of cancer cell lines to long chain fatty acids - either saturated (typically palmitate) or unsaturated (typically oleate) have been extensively employed. Studies variably report an increase in steatosis, upregulation of cytokines, increased reactive oxygen stress and apoptosis signalling and interruption of insulin signal transduction. One study focussed on the use of the high energy substrates pyruvate and lactate to mimic energy excess in combination with octanoic acid (LPO). Octanoic acid was preferred as it enters the mitochondria in a Cpt1 independent manner fuelling β-oxidation. This approach not only generated hepatic steatosis but an acute increase in mitochondrial reactive oxygen species and perturbation of energy metabolites. As such, LPO combination treatment is an attractive model.

In this sequence of experiments, I therefore had the following aims.

5.1.1 AIMS:
1) To determine whether or not ESCheps are a suitable model to study TET function and 5hmC distribution in hepatocytes in vitro
2) To recapitulate features of human NAFLD using high energy substrate LPO intervention.
3) To examine whether or not the observations of genic 5hmC enrichment observed in mouse high fat diet feeding are replicated in a human single cell hepatocyte population.

5.1.2 HYPOTHESIS:
1) ESCheps will retain 5hmC and TET expression in the differentiated state.
2) ESChep exposure to a combination of high energy substrates and medium chain fatty acids will recapitulate features of human NAFLD and provide an in vitro model of disease processes.
3) NAFLD intervention in ESCheps will replicate the findings of genic 5hmC enrichment in genes driving the synthesis and transport of lipid species.
5.2 Materials and methods

5.2.1 ESC cell maintenance and differentiation

Female H9 ESCs were maintained on Matrigel™ in mTESR1 medium (Stem Cell Technologies, Cambridge, UK) and split at 70% confluence. Differentiation was performed as previously described. Briefly, cells were split 1:3 ratio onto matrigel coated plates in either 6, 12 or 24 well format. A schematic of the differentiation protocol is described in 2.5.2.

5.2.2 LPO treatment and NAFLD phenotyping.

Lactate (Sigma L7022), sodium pyruvate (Sigma p5280) and octanoic acid (Sigma C2875) were combined and delivered to cells as described (2.5.1). Paraformaldehyde fixed cells were stained with BODIPY 493/503 (Life Technologies) neutral lipid marker and then counter stained with Hoesch 3342 stain (Life Technologies) each for 20 mins at room temperature, in the dark with agitation. Live cells were incubated with MitoSOX reagent (Life Technologies) for 45 minutes at 37°C, washed twice with PBS. AntimycinA 25µM (Sigma, Dorset, UK) was used as a positive control. Both were visualised with a Zeiss Axio Imager 2 Light Microscope (Cambridge, UK) and images analysed using imageJ (https://imagej.nih.gov/ij/). For image analysis, 12 randomly taken 20x magnification mosaics of 9 images containing ~ 700 nuclei per mosaic were included per condition.

5.2.3 Expression analysis

Cells were washed with 1xPBS solution and subsequently covered with 1 ml of Trizol (Life Technologies, UK) for 5 minutes followed by scraping and transfer into a clean Eppendorf for storage at -80°C. RNA was purified using the RNaeasy spin column RNA purification kit (Qiagen, UK), including on column DNAase treatment with DNaseI (Qiagen, UK), and elution into 40μl of water (2.6.1). Expression analysis by quantitative PCR and microarray expression (2.6.1) were then performed as indicated.

5.2.4 5hmC Analysis.

Cells were washed with 1xPBS, scraped in a further 1 ml PBS and centrifuged at 1000 rpm for 5 mins. Supernatant was removed and cell pellets stored at -80°C. DNA was extracted (DNeasy, Qiagen) including treatment with PureLink RNase A 20µl (Thermo Scientific) and final elution using 2 separate elutes of 200µl and combining. Sonication, immunoprecipitation, whole genome amplification, semiconductor sequencing, slot/dot blot
and data analysis were performed as previously (2.6.2). Control DNA was amplified APC promoter sequences with only the annotated nucleotide present (active motif, cat 55008)
5.3 Results

5.3.1 Differentiation of hESC derived cells with hepatocyte like morphology and function.

At terminal differentiation, ESCheps are morphologically a stable sheet of cells in 2D layer with square or polygonal shapes and cytoplasmic lipid droplets (Figure 5.1 A). Over the differentiation time course, ESCheps displayed a progressive reduction in pluripotency markers (POU class 5 homeobox 1, OCT3/4 and Nanog homeobox, NANOG) to near undetectable levels with a concurrent increase in hepatocyte specific markers (albumin, ALB and hepatocyte nuclear factor 4 alpha, HNF4α) (Figure 5.1 D). At day 21, ESCheps secreted albumin and AFP into culture medium and displayed high levels of Cyp1A2 and Cyp3A4 activity as measured by luciferase assay (figure 5.1 B and C).

5.3.2 ESCheps retain 5hmC and display a mature hepatocyte TET expression profile

Using immunoblot analysis I replicated previous findings of loss of 5hmC signal in the established HepG2 hepatocellular carcinoma cell line \(^{263}\). Strikingly, ESCheps at terminal differentiation (Day 21) retain 5hmC levels at equal levels to mouse liver tissue and this effect does not require the addition of ascorbate as described in mouse embryonic fibroblast cells \(^{362}\). Indeed, at higher ascorbate concentrations the 5hmC levels were significantly reduced (Figure 5.3 B and C).

Through differentiation, TET1 expression reduced rapidly following treatment with Wnt3a, Activin A and RPMI media. There was a concurrent increase in TET2 expression with TET2 the dominant isoform in mature cells (Figure 5.3 D).
Figure 5.1 Differentiated hepatocytes demonstrate morphological, metabolic and transcriptional aspects of somatic human hepatocytes. A) H9 female ESCs (left) differentiate over 21 days to mature hepatocyte like cells (right) with a polygonal shape and cytoplasmic lipid droplets in a 2D layer (black arrows). B) ELISA demonstrating albumin and AFP secretion and C) Cytochrome P450 1A2 and 3A4 activity as determined by luciferase assay in day 21 differentiated cells. D) at day 21 cells have suppressed pluripotency marker expression (OCT3/4, NANOG) and significantly elevated hepatocyte markers (HNF4A, ALB). * P < 0.05 one-way ANOVA with Tukey post hoc test. Error bars = SEM.
Figure 5.2: A) Slot blot analysis of mouse liver, H9 ESCs, HepG2 cell DNA. B+C) Dot blot analysis of mouse liver DNA with hESCs exposed to a dose titration of ascorbate for 24 hours followed by 5hmC quantification by image intensity analysis. D) Expression of TET enzyme isoforms through the hepatocyte differentiation process. \( \Theta = P < 0.05 \) * versus mouse liver DNA one-way ANOVA with Bonferroni post hoc analysis. * = \( P < 0.05 \) one-way ANOVA with LSD post hoc analysis versus ESC expression. Error bars = SEM.
5.3.3 Modelling NAFLD in ESCheps

Optimisation of NAFLD induction in ESChep cells was performed using two separate concentrations of lactate (L), pyruvate (P) and octanoic acid (O) over two durations of exposure (48 hours and 96 hours). High dose (L:P:O = 10mM:1mM:2mM) but not low dose (L:P:O = 2mM:0.2mM:0.4mM) stimulated the expression of multiple putatively causal genes in NAFLD pathological processes including fatty acid synthesis (ACACA, FASN, SREB1c, LXR), gluconeogenesis (PCK, G6PD) and lipid vesicular transport proteins (PLIN1, PLIN2). Interestingly the effect of this gene induction was transient for some genes (G6PD, PCK1, LXR, PLIN1) with normalisation at 96 hours suggesting some loss of metabolic function with prolonged exposure (Figure 5.3 A). Furthermore, there was attenuation of the LPO induced upregulation of PCK1 with insulin treatment (Figure 5.3 B). I therefore proceeded with 48 hour high dose LPO exposure for subsequent experiments. When examining phenotypic effects, high dose LPO exposure for 48-hours induced an increase in cellular steatosis with macro and micro vesicles seen on BODIPY neutral lipid staining (Figure 5.4 A). On examination of mitochondrial function, there was a non-significant trend (p=0.058) to an increase in mitochondria superoxide detection by MitoSOX staining within the LPO treated group (Figure 5.4 B). It should be noted however that the mitochondria displayed (by microscopy) a perinuclear clustering pattern suggesting an apoptotic process in both groups during the experimental protocol.

5.3.4 LPO induces upregulation of NAFLD Pathways in ESCheps

Microarray expression analysis of control and 48 hour high dose LPO treated ESCheps validated the qPCR findings of PLIN2, PCK1 and G6PD patterns of expression. Other genes which were upregulated upon qPCR analysis appeared to be similarly upregulated on the arrays, but these trends did not reach significance on robust correction for multiple testing. Upregulated mediators of lipid synthesis and transport included Cell Death-Inducing DFFA-Like Effector C (CIDEC), Apolipoprotein A4 (APOA4), Acyl-CoA Dehydrogenase, Very Long Chain (ACADVL), 3-Hydroxy-3-Methylglutaryl-CoA Synthase 2 (HMGCS2), Apolipoprotein A5 (APOA5), Cytochrome P450, Family 2, Subfamily J, Polypeptide 2 (CYP2J2), Insulin-Like Growth Factor Binding Protein 1 (IGFBP1). Additional pathways of transcriptional upregulation included drug detoxification processes (Cytochrome P450, Family 1, Subfamily A, Polypeptide 1, CYP1A1; UDP glucuronosyltransferase 1 family, polypeptide A1, UGT1A1) and steroid hormone metabolism (aldo-keto reductase family 1,
members C2, C4, B15 and B10; AKR1C2, AKR1C4 AKR1B15, AKR1B10) (Figure 5.3 C). The full list of dysregulated genes can be found in Appendix 2.
Figure 5.3 Changes in mRNA expression with LPO intervention. (A) LPO intervention induces transcriptional upregulation of mediators of de novo lipogenesis, lipid vesicular transport and gluconeogenesis. LXR, PLIN1, PCK1 and G6PD transcriptional induction did not exceed 48 hours. (B) 10nM actrapid ameliorated the effect of LPO induced PCK1 induction indicating insulin sensitivity. (C) Volcano plot of gene expression data from ESCheps following 48 hours high dose LPO treatment with annotated genes of interest. * = P < 0.05 versus control one-way ANOVA with Tukey post hoc analysis. Θ = P < 0.05 * one-way ANOVA with Bonferroni post hoc analysis. Error bars = SEM
Figure 5.4 Fluorescence microscopy and image intensity analysis of BODIPY 493/503 neutral lipid stain (A) and MitoSOX mitochondrial superoxide indicator (B) in control and 48 hours high dose LPO treated ESCheps. * = P < 0.05 Student T test. Error bars = SEM. n = 12.
2.3.5 NAFLD modelling in ESCheps induces genic 5hmC enrichment in lipid synthesis and transport genes without affecting global 5hmC levels.

I went on to determine whether or not my previous findings in mice were replicated in ESCheps using the validated model of LPO exposure. 5hmC-DIP-Seq was performed in an identical way to mouse liver however alternative positive and negative control regions were employed (negative control = L1PA1 repeat element and GAPDH promoter, positive control regions = H19 and UBIAD1 genic loci) based on previous reports (Figure 5.5 A)\textsuperscript{253,263}. Sliding window analysis of all genes demonstrated a markedly different 5hmC profile to mouse differentiated liver tissue with high levels of 5hmC over TSS and proximal gene bodies and lower levels over promoters. However, when stratified by expression level the highest 20% of expressed genes depicted a profile more similar to differentiated tissue with a marked reduction directly over the TSS and higher levels over the promoter and gene body regions with a further loss directly over the transcriptional end site (Figure 5.5 B).

On direct comparison between control and LPO treated cell populations, sliding window analysis again demonstrated no global change in 5hmC on exposure to high dose LPO for 48 hours. Furthermore, analysis of 5hmC levels within and outwith all annotated gene regions demonstrated no significant correlation between all dysregulated transcripts and 5hmC change. However, when examining individual gene sets for specific processes, genes involved in the synthesis and transport of lipids showed a significantly higher enrichment of 5hmC within gene bodies when compared with genes involved in other cellular processes such as buffering oxidative stress, drug detoxification and gluconeogenesis (Figure 5.6 (GENIC)).
Figure 5.5 5hmC DIP sequencing of control and LPO treated ESCheps.  A) positive and negative control regions of 5hmC-DIP ESCheps analysed by qPCR.  B) Sliding window analysis of 5hmC-DIP sequencing of control ESCheps with profiles stratified by gene expression as determined by array probe intensity. Higher gene expression values imitate the differentiated mouse liver hydroxymethylome (Figure 4.6).  C) Sliding window analysis comparison of all genes in control and LPO treated cells (n = 3 per group).  Error bars = SD.
Figure 5.6 Change in 5hmC over annotated gene regions between control and 48 hours high dose LPO treated cells in indicated gene sets. Graphs represent the mean of 3 separate experiments. P values generated by Mann Whitney test.
5.4 Discussion

5.4.1 ESCheps are a novel model for examining hepatic 5hmC regulation in vitro

The meaningful investigation of 5hmC function in somatic tissue in vitro is challenging, principally because 5hmC is rapidly lost. Nestor et al demonstrated near complete loss of 5hmC within three days of mouse embryonic fibroblast and CD4+ T cell adaptation to culture conditions. Indeed, culture specific perturbations in the epigenome were observed over 20 years ago by Antequera et al who demonstrated CGI hypermethylation in cultured cell lines. In addition, it was then proposed that loss of normal in vivo demethylase activity may be the mechanism for this finding long before the discovery of TET enzymes or the demethylation pathway. The mechanism for this loss of TET activity is not known, however it precludes the use of established cancer cell lines in this study.

This chapter demonstrates that in differentiated ESCheps, 5hmC levels are present in comparable amounts to fresh liver tissue in contrast to the established hepatocyte cell line HepG2. The reasons for this observation are unclear. 5hmC loss is a feature of rapid cell proliferation and malignancy and is observed in various cancers. ESCheps are not derived from neoplastic tissue unlike cancer cell lines and at terminal differentiation, ESCheps demonstrate no cell division, which may allow TET isoenzymes to maintain 5hmC levels. However, ESCs maintain 5hmC in culture despite rapid self renewal and the reasons for this are poorly understood. This may be due to comparatively high TET expression, due to the micronutrient cofactor content of commonly used stem cell media or due to an as yet undefined stem cell specific property, which is retained in the ESCheps model.

We observed striking changes in the TET isoenzyme expression profile throughout differentiation that add new insight into this field. Mouse ESCs are generated from the blastocyst inner cell mass during early embryogenesis. At this developmental stage, TET1 and TET2 expression are high and a spike in 5hmC levels is observed. Similarly, in mouse ESCs 5hmC levels are higher than most differentiated tissues as are TET1 and to a lesser extent TET2 transcripts. TET 1 and 2 expression are maintained by the pluripotency marker OCT4 and in turn may maintain the expression of NANOG through promoter hydroxymethylation. In agreement with this, induction of pluripotency from mouse embryonic fibroblast cells is associated with a marked increase in Tet1 and Tet2 expression with upregulation of Oct3/4. I observed a rapid loss of TET1 expression following the addition of Activin A and Wnt3a factors, however I also saw concurrent induction, rather than loss, of TET2 as Oct3/4 and NANOG are suppressed. TET2 stability is regulated by IDAX, a
reported inhibitor of Wnt3a signalling\textsuperscript{381}. It may be that a reciprocal relationship exists, in that TET2 is a target of Wnt3a mediated transcriptional activation. Alternatively, a novel interaction may exist between TET2 expression and the Activin A/Tgfβ signalling pathway and this warrants further study. At terminal differentiation, TET2 was the dominant isoform with low TET1 expression in keeping with my mouse studies (Figure 4.4).

5.4.2 LPO model performance.

Faithfully recapitulating the many facets of NAFLD in cell culture systems is challenging as it requires the induction of hepatic steatosis, insulin resistance, mitochondrial dysfunction and elevated gluconeogenesis. Fatty acids have traditionally been employed with various reports of effects on steatosis, insulin signalling and activation of apoptosis pathways\textsuperscript{62,105,106,374,375}. However, their hydrophobic nature requires the use of a vehicle (such as ethanol or DMSO) and although there is representation of high FA delivery to the liver, other facets of over nutrition are not examined. Therefore, the use of high energy substrates that drive gluconeogenesis with the inclusion of an unsaturated fatty acid that is readily solubilized and CPT1 independent is an attractive option\textsuperscript{382}.

The hESC derived hepatocytes demonstrated an increase in neutral lipid staining on LPO exposure. This effect was ameliorated due to the presence of lipid droplets in untreated cells as is typical of this model\textsuperscript{151,383}. Furthermore, rather than the typical single lipid macrovesicle seen in human NALFD (Figure 1.1 A), LPO treated cells exhibit a plethora of micro vesicles throughout the cytoplasm. Consistent with reports in HepG2 cells\textsuperscript{382}, there was some evidence of impaired mitochondrial function as demonstrated by an increase in mitochondrial specific superoxides although this was difficult to measure due to evidence of apoptosis during the staining protocol in both groups. Crucially however the protocol did consistently result in high levels of transcriptional activation in multiple genes associated with NAFLD pathogenic molecular pathways including lipid droplet transporters PLIN2 and CIDEC\textsuperscript{384,385}, the nuclear factor LXR\textsuperscript{89,386} gluconeogenesis enzymes PCK1, G6PD and PDK4, apolipoproteins (APOA4, APOA5 and APOC3) and the cholesterol synthesis mediator HMGCS2. Further, PCK1 induction was suppressed with pulsed insulin confirming the presence of an intact insulin signalling pathway. These transcriptional effects were partially lost following 96 hours exposure perhaps due to exhaustion of metabolic reserve or ESChep de-differentiation seen with maturation beyond 23 days\textsuperscript{151}. The favourable TET expression profile and retention of 5hmC levels in the ESCheps model, coupled with the reproducible transcriptional changes
strongly suggest this is the first tractable cell culture model for examining 5hmC regulation in metabolic liver disease and is worthy of further investigation.

5.4.3 5hmC profiles in ESCheps is typical of ESCs not Differentiated hepatocytes.

Many studies have profiled 5hmC levels in embryonic stem cells using a variety of affinity based and single base resolution techniques demonstrate high levels of 5hmC over TSS, distal promoter and proximal gene body regions. Furthermore, greater levels of 5hmC are observed within gene bodies and promoters of highly expressed genes, with a relative reduction directly over annotated transcriptional start sites. Ivanov and colleagues published the only study examining 5hmc profiles in human liver. Although no sliding window analysis is depicted, they report only slight enrichment within gene bodies and much greater enrichment over CGIIs and CGI shores. Profiles generated from ESCheps in this study were highly reproducible between individuals, broadly in keeping with these findings (Figure 5.5 B and C). Furthermore, given that the global transcriptome of ESCheps bears greater similarity to primary hepatocytes than ESCs, this suggests at least a partial adaptation from an ESC-like to a human liver-like hydroxymethylome. Further analysis of H9 ESC and human liver expression data integrated with the 5hmC profiling from these cells may give an indication as to what extent the hydroxymethylome here is an accurate representation of human hepatocytes.

5.4.4 LPO intervention induces genic 5hmC enrichment in lipid synthesis and transport mediators.

In keeping with my findings in the mouse, LPO had no effect on global levels of 5hmC or TET expression. However, examination of annotated gene regions revealed evidence of 5hmC over genic region of lipid synthesis and transport mediators in the absence of enrichment over other annotated regions. Furthermore, in this system genic 5hmC enrichment occurred rapidly (within 48 hours) as would be required by hepatic regulatory mechanisms responding to the delivery of dietary nutrients by the portal blood flow. This is in keeping with findings of murine drug induced liver injury which demonstrated 5hmC changes after one day. The mechanism for this observation requires further investigation. Murine and ESChep NAFLD interventions did not alter Tet transcript levels and as such it is now important to examine Tet protein level and post translational modification states (phosphorylation and O-GlcNAcylation) of Tet proteins as well as fluctuations in co-factor metabolites.
5.4.5 Experimental limitations

Although ESCheps demonstrate many of the desired morphological, metabolic and transcriptional properties of human primary hepatocytes, the whole transcriptome is clearly not entirely representative. The retention of AFP secretion suggests that these cells are more foetal than adult in developmental stage and as such the results should be interpreted in this context. In addition, although the LPO phenotype induced expression of a variety of genes within important NAFLD pathogenic pathways, it is not yet possible to accurately replicate the complex interactions between hepatocytes and other liver cell types such as Kupffer and stellate cells in vitro. The progressive development of co-culture systems and the generation of 3D organoids from pluripotent cell populations as well as the use of adipokines, inflammatory factors and gut derived peptides may allow more accurate representation of this complex pathology in the future.

As with the mouse model I have described an association between changes in the genic 5hmC content and transcriptional induction in relevant gene sets. This study does not however determine to what extent these changes drive or are caused by alterations in transcriptional activity. Sequential knock down of TET 1-3 isoforms during the later stages of ESChep differentiation followed by LPO challenge may provide valuable information in this respect and represent the logical next experiment.

Conclusion

To my knowledge, this is the first time that the hydroxymethylome and TET expression profile have been investigated in the context of ESC derived hepatocytes. Furthermore, this is the first time that 5hmC regulation has been investigated in any disease process using an ESC derived tissue model. Using this novel system, I demonstrate broad replication of my findings in mice in a human, single cell population, suggesting that promotion of hepatic TET activity in the context of over nutrition is a conserved phenomenon relevant to the human population.
CHAPTER 6: TET1⁻/⁻ RESPONSE TO OBESOGENIC DIETARY INTERVENTION

6.1 Introduction.

Chapters 4 and 5 were consistent in the demonstration of genic 5hmC enrichment in upregulated mediators of lipid synthesis and transport in response to over nutrition without changes in Tet expression. It is unclear however to what extent these changes are the result of or play a functional role in transcriptional activation. To address this it was therefore logical to examine the effect of over nutrition in the context of Tet deficiency.

Tet enzymes are crucial for the maintenance of normal transcriptional activity through mammalian development and in adult complex organisms. This is demonstrated by the skewed differentiation of Tet1⁻/⁻ ESCs in teratoma assays and vulnerability of adult Tet1⁻/⁻ mice to lymphoid malignancy and memory disorders. The rapid initiation of myeloid malignancy in constitutive and bone marrow specific Tet2 knock out mice and the embryonic lethality of Tet3 deficiency support this observation and also suggest non-redundant function for each Tet isozyme.

Recent reports suggest a role for Tet regulation in both liver disease and metabolic regulation. In human adult liver disease, a down regulation of Tet1 mRNA expression has been demonstrated in HCC and concurrent loss of 5hmC is associated with poor prognostic outcome. Global analysis of 5hmC levels as determined by immunoblot or staining suggests a reduction in patients with primary sclerosing cholangitis with concurrent upregulation of Tet2 protein levels. In addition, mouse NIH 3T3 pre-adipocyte cells demonstrated active demethylation of PPAR response elements within promoters of adipocyte specific genes during differentiation. Furthermore, 5hmC enrichment at these regions was prevented by Tet1 and Tet2 knock down and the adipocyte differentiation mediator PPARγ co-immunostained and precipitated with Tet1 in this cell line. Finally, Tet1 mRNA levels are suppressed in the subcutaneous adipose tissue of patients with T2DM compared to their non-diabetic monozygotic twins. Given Tet1 is required for reliable cellular differentiation, it is tempting to speculate that Tet1 suppression impair the generation of functional adipocytes and potentiate peripheral insulin resistance.

Tet1⁻/⁻ mice are reported to be grossly normal in appearance with male mice being of equal size to wild type mice at 6 weeks of age. Heterozygote pairs are reported to produce knock out and wild type progeny at expected Mendelian inheritance ratios. Aged Tet1⁻/⁻ mice
exhibit lymphoid malignancy however this is only apparent after 1 year of age with mice remaining healthy up until that point \textsuperscript{238}. Tet1\textsuperscript{−/−} ESCs and Tet1 knock down ESCs display a reduction in 5hmC levels of ~40% as estimated by immunoblot analysis, although any subsequent effects in differentiated tissue is unknown. At present, a Tet1 transgenic animal with loxP insertion is not commercially available to allow tissue specific deficiency to be examined. However, the reported association of Tet1 with the metabolic regulator PPAR\textgamma, the relatively healthy phenotype of the constitutive Tet1\textsuperscript{−/−} mouse and the observation of Tet1 suppression in HCC, a key clinical endpoint in NAFLD progression, make interrogating the metabolic response of Tet1\textsuperscript{−/−} mice intriguing.

6.1.1 AIMS

1) To breed Tet1\textsuperscript{+/−} mice and generate Tet1\textsuperscript{−/−} and Tet1\textsuperscript{+/+} littermates. Confirm the absence of the functional Tet1 transcript in liver tissue and any compensatory response from other Tet isoforms.

2) Examine the effect of Tet1 deficiency on the hepatic hydroxymethylome

3) Characterise the effect of constitutive Tet1 deficiency on diet induced obesity, dysglycaemia, insulin sensitivity and NAFLD induction.

4) Examine the effect of Tet1 deficiency on the transcriptional activation of hepatic lipid synthesis and transport mediators.

6.1.2 HYPOTHESIS

1) Tet1\textsuperscript{−/−} will reduce 5hmC levels over defined annotated regions within the hepatic hydroxymethylome.

2) Loss of Tet1 will reduce the murine capacity to generate and transport lipid molecules in response to over nutrition. This will cause either a) Reduced hepatic lipid synthesis capacity with a reduction in NAFLD score and improved insulin sensitivity or b) reduced peripheral adipose expansion with increased hepatic lipid sequestration and exacerbation of NAFLD.
6.2 Material and Methods.

6.2.1 Animals

B6;129S4-Tet1tm1.Jae/J heterozygote mice were purchased from Jackson Laboratories, Maine, US and housed in standard conditions as described (2.3.1). Heterozygote mice were interbred to generate homozygous, knock out and wild type progeny. The first cohort (n = 10 Tet1+/-, n = 9 Tet1-/- males) were housed either individually or in pairs with littermate subjects. Animals were matured until 12 weeks of age before being commenced on high fat diet (HFD) (table 2.3.1) ad libitum for 17 weeks followed by glucose and insulin tolerance testing and humane killing (2.3). A validation experiment was performed on a second cohort (n = 4 Tet1+/+, n = 4 Tet1-/-) in single housed conditions to analyse dietary intake, indirect gas calorimetry and body composition by Time-Domain Nuclear Magnetic Resonance (TD-NMR) (2.3.7).

6.2.2 Genotyping.

Lysed ear notches were digested and 2μl amplified by Kapa Biosystems HS green mastermix (2.6.3) using described primers (2.8.5). Products were resolved on a 3% agarose gel at 70v for 1.5 hours.

6.2.3 Liver and serum measurements

Serum glucose was measured using a hexokinase assay (Thermoscientific, UK), insulin using an Insulin ELISA kit (Mercodia, Upsala, Sweeden) and liver triglycerides using a colorimetric assay kit (Cayman Chemistry). 10% formalin fixed liver tissue was paraffin blocked and stained for haematoxylin and eosin (H&E) as described (2.4.1) and histologically scored in blinded randomised order by a consultant pathologist (Dr Tim Kendall).

6.2.5 Expression analysis.

RNA was purified using the RNAeasy spin column RNA purification kit (Qiagen, UK), including on column DNAase treatment with DNaseI (Qiagen, UK), and elution into 40μl of water (2.6.1). Total RNA samples were assessed on the Agilent Bioanalyser with the RNA 6000 Nano Kit (Agilent Technologies, Germany) and quantified using the Qubit RNA BR kit and assessed for DNA contamination using the Qubit DNA HS assay Kit (Life Technologies). Libraries were prepared and quantified as described (2.6.1). Sequencing was performed using the NextSeq 500/550 High-Output v2 Kit (150 cycles) on the NextSeq 550 platform and
analysed on the Basespace platform (2.7.3) and R version 3.2.3. Gene ontology analysis was performed on the PANTHER classification system ([http://pantherdb.org/](http://pantherdb.org/)) using the Bonferroni over representation test within GO terms for biological function.

6.2.6 5hmC Analysis.

DNA was extracted (DNeasy, Qiagen) including treatment with PureLink RNase A 20µl (Thermo Scientific) and a final elution using 2 separated elutes of 200µl and combining. Sonication, immunoprecipitation, whole genome amplification, semiconductor sequencing and analysis using the Galaxy server and R version 3.2.3 was performed as described (2.6.2).

6.2.7 Energy homeostasis analysis

Animals were single housed under standard conditions from weaning (4 weeks). Food intake was measured weekly for 6 weeks. At baseline and following a 7 week dietary intervention, animals were transferred to the TSE Phenomaster System (Bad Homberg, Germany) for 5 days. Animals were allowed to acclimatise for 72 hours before indirect gas calorimetry was measured over a 24-hour period.
6.3 Results

6.3.1 Frequency of Tet1 progeny and effect on the hydroxymethylome.

Genotyping of wild-type, heterozygous and knockout progeny of Tet1+/− x Tet1+/− matings produced the expected amplicon bands of wild type (~300bp) and Tet1−/− (~650bp) alleles (Figure 6.1A). The absence of the Tet1 transcript in mutant liver was confirmed by qPCR analysis with a forward primer targeted to exon 4. There was no compensatory upregulation of the other Tet isoforms seen (Figure 6.1D). At weaning (4 weeks), the first 299 genotyped offspring did not fit Mendelian inheritance patterns with only 10.5% being homozygous knockout animals, 56.8% heterozygotes and 32.7% wild type animals (P<0.001 Chi-square test) (Figure 6.1C). Genotyping was not performed at birth, and as such it was not possible to determine if this was a consequence of a high incidence of early post-natal death. There was no observed gender bias (45.6 % male, p = 0.219 Chi-square test). I then examined the effect of Tet1 loss on 5hmC levels in adult liver. Sliding window analysis revealed a marked loss of 5hmC throughout annotated gene and proximal promoter regions as previously described (Figure 6.2) 345.
Figure 6.1 Tet1 genotyping and progeny. (A) Tet1 allele wild type and knock out constructs with resolved products of genotyping by PCR (B). Red, blue and green arrows represent positions of genotyping primers. (C) Schematic of genotype frequencies in adult progeny, 1 circle = 1%. (D) qPCR analysis of Tet isoforms in liver tissue by genotype. $\sigma = P < 0.01$ Chi-square test. * $P < 0.01$ Mann-Whitney-U test. Error bars = +/- SEM.
Fig 6.2 Effect of Tet1 loss on hepatic 5hmC levels. HmeDIP-sequencing of normal chow fed wild type and knock out mouse liver at 12 weeks of age (n=1/group).
6.3.2 Metabolic effect of Tet1 deficiency.

Animals were allowed to mature on normal diet until 12 weeks of age. At this stage, the mice were of normal morphology with no difference in body length. Following instigation of the high fat diet, both genotypes demonstrated weight gain over 17 weeks (Figure 6.3A). As expected the wild type animals gained weight in an identical fashion to C57BL/6 mice from earlier experiments (Fig 4.2). Strikingly however, after 5 weeks the genotypes diverged. Knock out animals on the high fat diet continued to gain weight in keeping with normal growth but at a rate indistinguishable from C57Bl6 mice on normal chow diet and significantly less so than wild type animals on the high fat diet. Following dietary interventions Tet1 KO animals were significantly more insulin sensitive than wild type litter mates, with lower fasting and stimulated insulin levels on intraperitoneal glucose tolerance testing (Figure 6.3C and D). There was no difference in fasting glucose or glucose tolerance between genotypes (Figure 6.3E). At termination of the experiment, knockout mice exhibited reduced subcutaneous and mesenteric fat pad weight although interestingly with similar gonadal depot weight (Figure 6.3B). Accordingly, TD-NMR analysis of body composition confirmed that while body composition was similar on standard chow, following a high fat diet intervention the WT mice exhibited a significantly higher percentage body fat than Tet1<sup>−</sup> animals (Figure 6.4C). Histological examination of wild type and knock out high fat diet fed animal liver revealed a highly protected phenotype in the knockout animals with steatosis of >5% in only 3 animals compared to steatosis >5% in all of the wild type animals. This was supported by lower levels of hepatic triglycerides in the Tet<sup>−</sup> animals (Figure 6.4 A, B-D).
Figure 6.3 Metabolic profiling of Tet1 deficient mice. (A) Weight gain of high fat diet fed animals (n=10 wild type, 9 knock out) of each genotype and comparison with the original C56bl/6 high fat diet fed cohort. (B) Normalised fat, liver and spleen weights following HFD feeding. Glucose (C) and insulin (D) levels following 2g/kg intra peritoneal glucose injection. E) Glucose concentrations on insulin tolerance testing. \( \sigma = P < 0.001 \) one-way ANOVA with Bonferroni post hoc correction vs Tet1\(^{-/-}\). * = P < 0.05 Student’s T test, \( \Omega = P < 0.01 \) \( \Omega \Omega = P < 0.001 \) Student T test with Bonferroni correction versus Tet\(^{-/-}\) animals. Error bars = +/- SEM.
Figure 6.4 Liver and body fat composition in Tet1 deficient mice. A) Kleiner NAFLD score, B) H+E histological analysis of high fat diet fed mice showing macrovascular steatosis in WT mice (black arrow) C) TD-NMR analysis and D) hepatic triglyceride levels in wild type (blue) versus homozygous knock out (red) animals. * = P < 0.05 ** P < 0.01 Student’s T test versus wild type animals. Error bars = +/- SEM. Boxplot whiskers = IQR.
6.3.3 Tet1 deficiency confers a protective hepatic transcriptome

I then went on to examine the effect of Tet1 deficiency on the hepatic transcriptome following HFD intervention. Unsupervised hierarchical clustering of the top 1000 variable transcripts was sufficient to separate the cohorts by genotype (Figure 6.5A). At a threshold of 20% change with an adjusted P value of <0.05, there were 435 up and 625 down regulated genes in the knock out animals. The most enriched down regulated GO terms were lipid synthesis, inflammatory and fibrosis pathways including ‘cholesterol biosynthetic process’, ‘positive regulation of phagocytosis’ and ‘extracellular matrix organisation’. Among the upregulated transcripts in Tet1 KO animals, the GO term ‘positive regulation of mesenchymal cell proliferation’ was highly enriched above all other terms encompassing genes Irs2, Foxp1, Myc, Hmgb1, Arhgap5, Prrx1 and Fgfr1 (Figure 6.5B and 6.6A).

I then examined genes identified in chapters 4 and 5 which showed 5hmC enrichment and transcriptional activation in liver from C57Bl/6 mice on a high fat diet, and LPO challenged ESCheps. The majority of these genes exhibited significantly lower expression levels in the Tet$^{−/−}$ animals, suggesting they were not activated by HFD intervention in contrast to wild type offspring (Figure 6.6A). In addition, there were significantly lower transcript levels of other mediators of de novo lipogenesis (Sterol Regulatory Element Binding Transcription Factor 2, Srebf2, Acetyl-CoA Carboxylase Beta, Acacb; Acyl-CoA Synthetase Short-Chain Family Member 2, Acss2), lipid synthesis (Aquaporin 4, Aqp4; 1-Acylglycerol-3-Phosphate O-Acyltransferase 2, Agpat2) and fibrosis mediators (Collagen Types 1a1, 1a2, 3a1, 4a1, 4a2, 4a5, 5a2, 5a3; matrix metalloproteinases Mmp12 and Mmp13) (Figure 6.6B).
Figure 6.5 Transcriptional dysregulation in Tet1-deficient mice on a high fat diet. Unsupervised hierarchial clustering (A) and gene ontology (B) of up (red) and down (blue) regulated genes in Tet1-deficient mouse liver following HFD feeding.
Figure 6.6 RNA sequencing analysis of Tet1−/− mice versus wild type controls. A) Scatter plot of mRNAs significantly dysregulated by 1.2 fold change or greater. Up- (red) and down- (blue) regulated genes are plotted. Dysregulated 5hmC enriched genes from C57Bl/6 and ESChep NAFLD models are indicated as yellow (down regulated) or green (unchanged) in Tet1 deficient liver tissue following HFD feeding. (B) mRNA transcripts of de novo lipogenesis, triglyceride synthesis and fibrosis significantly down regulated in Tet1−/− mice following HFD intervention.
6.3.4 Further investigation of the Tet1\(^+/-\) phenotype.

To determine the cause of the apparent protection from obesity and metabolic dysregulation in the Tet1\(^+/-\) males, I examined calorie intake and energy expenditure in Tet1\(^+/-\) and wild type litter mates on a high fat diet using single housed conditions and indirect gas calorimetry. There was no difference in the relative mass of food intake between genotypes over a 6-week examination period (Figure 6.7A). This was repeated in the single housed validation cohort (n=4). Data are normalised to body mass. Furthermore, there was no increase in total energy expenditure during active and inactive periods as determined by the respiratory exchange ratio (vCO\(_2\) expired/vO\(_2\) consumed) (Figure 6.7B).

Fig 6.7 Energy intake and expenditure of Tet1\(^+/-\) and Tet1\(^-/-\) single housed mice. Food intake (A) and indirect gas calorimetry (B) showing no difference between genotypes following 7 week high fat diet feeding. Error bars = +/- SEM
Discussion

To my knowledge this is the first time that deficiency of any Tet methylcytosine dioxygenase enzyme has been examined from the perspective of metabolic disease and my work provides new insights into the role of 5hmC regulation in energy homeostasis.

Lower than expected numbers of homozygous knock out progeny were observed in contrast to previous reports. This may be due to my observations being in a much larger cohort and also due to genotyping being performed at 4 weeks of age rather than birth. My findings suggest that either Tet1<sup>−/−</sup> gametes or Tet1<sup>+/−</sup> zygotes are less viable than wild type or heterozygote counterparts or alternatively, that Tet1<sup>+/−</sup> offspring have a higher incidence of intra-uterine or early post-natal death not detected by our standard husbandry practice.

Despite low mRNA expression in adult liver, loss of Tet1 has a profound effect on 5hmC levels in adult mouse liver tissue suggesting minimal redundancy between isoenzyme subtypes. I found a marked disparity in the response to high fat diet with a constitutive loss of Tet1 expression. Tet1 deficient mice demonstrated an obesity resistant phenotype with less weight gain due to lack of adipose deposition. Consistent with this there was also notable improvement in other features of the metabolic syndrome, including lower fasting and stimulated insulin levels, improved insulin sensitivity and a reduction in hepatic lipid accumulation and NAFLD score. The explanation for this is unclear.

In any organism, total body fat content reflects the balance of caloric intake and absorption against energy expenditure and thus it is useful to discuss the relative evidence for these possibilities in the observed experimental findings. My findings show that the differences in the phenotype on a high fat diet is not due to reduced appetite or food finding ability e.g. secondary to the previously reported memory dysfunction. There was no evidence of steatorrhoea in the cages suggestive of impaired fat absorption, although malabsorption of other nutrients cannot be excluded. Calorie loss through glycosuria is a possibility although this would not be expected unless the constitutive knock out caused severe renal dysfunction. Measurement of energy expenditure in small animals is technically complex. The product of all metabolic activity and thus all energy expenditure is either thermogenesis or movement. The measurement of heat generation in mammals is difficult, and thus it is easier to measure the components of glucose metabolism in the form of O<sub>2</sub> and CO<sub>2</sub> gas exchange also known as indirect gas calorimetry. The RER in both groups was low (~0.8) suggesting fatty acid oxidation as the dominant energy source. This is expected given the 58% saturated fat dietary...
composition. I did not observe any difference in energy expenditure between genotypes during the day-time or the active night period. Finally, the weight gain profile of wild type mice was identical to that of previous C57Bl6 mice indicating that wild type mice behave comparably to the validated C57Bl6 diet induced obesity model and are unaffected by any residual Sv129 background from the original chimera generation \(^{232}\). The knock out mice continued to gain weight on HFD feeding with no periods of sudden weight loss that would indicate systemic disease and I saw no evidence of malignancy in Tet1 deficient mice in the form of hepatosplenomegaly, lymphadenopathy or malignant infiltration on liver histology. Therefore, I have no clear evidence of reduced intake or absorption, increased thermogenic energy expenditure or illness to explain the lean phenotype.

Hepatic expression analysis revealed a remarkably protected transcriptome in \(\text{Tet}1^{-/-}\) mice on a high fat diet compared to wild type mice, with reduced expression of not only the majority of transcripts which were upregulated in both the mouse high fat diet fed and cell culture LPO models (Chapter 4 and 5), but also of many other mediators of lipid synthesis and fibrosis. Since Tet1 may have a role in modulating transcriptional activity in adipose, liver and muscle tissue, it is difficult to determine whether these hepatic transcriptional changes are primarily a consequence of hepatic Tet1 deficiency or are a result of the global reduction in body fat.

Observations point to an important functional association between the regulation of cytosine modifications and adipose tissue function. In a remarkable series of experiments, Kim et al demonstrate that obesity derived inflammatory mediators induce DNMT1 expression, hypermethylation and chromatin reorganisation of the adiponectin promoter with subsequent suppression of adiponectin expression in mouse white adipose tissue (WAT) and cultured cells \(^{392}\). They further show that the DNMT1 inhibitor RG108 ameliorates the metabolic syndrome in \(\text{db/db}\) mice in an adiponectin-dependant manner. Thus, modifying cytosine modifications in adipose tissue can improve metabolic phenotype. Fujiki et al report a possible role for Tet1 in the normal differentiation of adipocytes through a direct interaction with Peroxisome proliferator-activated receptor gamma (Ppar\(\gamma\)) and demethylation of Ppar\(\gamma\) DNA binging loci \(^{353}\). Ppar\(\gamma\) is a ligand activated nuclear receptor highly expressed in adipose tissue. Activation of Ppar\(\gamma\) stimulates recruitment of the co-activators Ppar\(\gamma\) coactivator 1 alpha (Pgc1a) and retinoid X receptor (RXR) and drives adipocyte differentiation, expansion and fatty acid uptake \(^{393}\). In humans, mutations in PPAR\(\gamma\) result in familial lipodystrophy characterised by reduce body adipose deposition with subsequently sequestration of fat within liver tissue causing profound hepatic steatosis \(^{394}\). Similarly, adipose specific Ppar\(\gamma\) knock out mice
exhibit hepatic steatosis with reduced body fat mass. Unlike these examples of reduced adipose depot expansion, the HFD fed Tet1−/− demonstrate a clear reduction in hepatic lipids suggesting this is not the only mechanisms involved. However, some degree of adipose dysfunction may be contributory to the Tet1−/− lean phenotype and serum assessment of adipokines, WAT histology, WAT maturity markers and free fatty acid levels may be informative in exploring this further.

An alternative possibility for my observations may be a skewed differentiation of WAT to ‘beige’ adipose tissue or expansion of brown fat (BAT) depots. Brown fat is a specialised adipose tissue found in mice and humans in specific depots such as cervical and intrascapular regions. In addition, ‘brown like’ (also known as beige) adipocytes, exist within WAT depots and exhibit a similar but not identical transcript prolife to BAT. Both BAT and beige cells consume energy though thermogenesis by uncoupling ATP production from oxidative phosphorylation. In addition, BAT stores may play an essential role in glucose and triglyceride clearance. Expansion of BAT depots or beige cell types within WAT depots in Tet1−/− mice could account for the lean phenotype. However, were this the cause of such a significant discrepancy in fat deposition, I would have expected to see higher energy expenditure in the Tet1−/− animals on indirect gas calorimetry. The use of cold exposure with assessment of shivering, thermographic camera analysis, measurement of BAT depot weights and analysis of beige specific marker transcript levels in WAT depots (Ucp1, Hoxc9, Shox2, Tbx1, Tmem26, Fgf21) would give a definitive answer as to whether this is likely.

There is the tantalising possibility this may be a liver specific effect due to a putative role for Tet1 in the regulation of lipid synthesis and transport mediators. A key driver in the metabolic syndrome is gluconeogenesis from liver tissue secondary to hepatic insulin resistance exacerbated by the accumulation of lipid species within the liver. If indeed the loss of Tet1 within liver tissue prevents the activation of lipid synthesis and transport genes under obesogenic conditions, this would reduce hepatic insulin resistance and therefore gluconeogenesis. Thus less glucose would be available for peripheral adipose tissue uptake and fat depots would be reduced. Against this theory is the finding of equal glucose levels between genotypes. However, this can be confounded due to the stress of glucose tolerance testing and as such glycated haemoglobin measurement may be more informative of chronic glycaemic control. Alternatively, Tet1 may play a role in pancreatic beta cell development or function which would account for the observed hyperglycaemia in the context of lower insulin levels. If this theory is correct, it is unclear how Tet1 may be specifically targeted to these
lipid synthesis and transport gene sets. For this to occur an intermediary protein would be required to recruit Tet1 in a locus specific fashion. Candidate proteins for this function would be Ogt and Pparγ which are both expressed in liver tissue, have metabolic regulatory function, chromatin binding properties and are able to directly interact with Tet1. Detailed analysis of liver ChIP data sets for these proteins intersected with the 5hmC analyses in this thesis may provide further evidence for this.

My data therefore presents an intriguing biological question, how does Tet1 deficiency result in a lean phenotype? A significant limitation of the constitutive knock out model is difficulty in determining the predominant tissue affected. NAFLD is the result of a complex interplay between the liver, muscle and adipose tissue and thus Tet1 loss in any or all of these tissues may significant. The observed reduction in hepatic 5hmC would suggest that despite low expression in adult liver, Tet1 is important for normal liver function. However, the effect of Tet deficiency on adipose and muscle tissue metabolic function is unknown and deficiency in either or both of these tissues could be contributory to the phenotype. A further limitation of a constitutive knock out is differentiating between developmental and somatic effects. Histologically, Tet1−/− mouse liver tissue appears normal but the presence of a hepatic developmental abnormality that is only exposed by a metabolic challenge in adulthood cannot be excluded. Published gene ontology analysis of dysregulated transcripts in Tet1−/− ESCs report enriched pathways of WNT signalling, regulation of metabolism and NOTCH signalling. Thus developmental effects are possible. The use of inducible tissue specific knock out animals or Tet1 over expression in specific tissues of knock out animals may allow identification of the key tissue to focus further study. Once that tissue is identified, paired 5hmC and transcriptional analysis may demonstrate that Tet1 is required for the activation of key genes in lipid metabolism.

**Conclusion**

Taken together these data suggest that despite relatively low expression levels, Tet1 plays an important role in the maintenance of the global hydroxymethylome in adult liver tissue. Further, I provide evidence that Tet1 may play an important role in lipid metabolism and energy homeostasis that is yet to be defined. An essential question is whether or not the Tet1−/− effect is primarily hepatic or extra hepatic in origin and this may lead to an exciting new avenue in obesity research.
CHAPTER 7: THESIS DISCUSSION

General Discussion:

As a central mediator in nutritional state, rapid changes occur in the hepatic transcriptome in response to portal and systemic signals of nutrient availability. A detailed understanding of the signal transduction process by which hepatic lipids are synthesised and transported is essential if novel therapeutic targets in NAFLD are to be found. The current dearth of pharmacological interventions for NAFLD highlights the complexity of this process and need for further research. It should also be noted that NAFLD is an integral part of a multisystem pathology and associates strongly with Type 2 diabetes, cardiovascular death and common cancers. Thus effective treatment may have knock-on broad ranging positive effects. The work contained in this thesis explores for the first time in detail the potential role of Tet methyl dioxygenase enzymes, involved in the generation of 5hmC DNA profiles, in this field. Observational studies of perturbations in cytosine nucleosides in human NAFLD combined with molecular observations of a Tet interdependency with energy metabolites provide a strong rationale for this approach.

7.1 What this thesis adds

The association of genic 5hmC levels and transcriptional activity is well documented. Further, two elegant studies have already demonstrated the dynamic nature of 5hmC regulation in adult mouse liver tissue in which genic 5hmC enrichment can be induced by toxicological insult in a cumulative dose dependant manner and with partial reversibility. This thesis builds on current work by replicating these observations in the most globally prevalent liver pathology. Further these data suggest that intact Tet function is essential for normal energy homeostasis.

Firstly, these data support the proposition that 5hmC generation, presumably through Tet activity, is highly dynamic in differentiated liver tissue and rapidly responds to external cellular stimuli in a locus specific manner. This is an essential property for transcriptional regulators particularly in the co-ordination of energy metabolism in which nutrient availability fluctuates frequently. Thus in principle it supports the notion that Tets may play a role in this process. Previous studies using phenobarbital have examined genes grouped by the time point at which they are induced and resolved during a phenobarbital dosing regimen. This study utilises similar methodology but instead groups genes by biological function to interrogate the role of Tet activity in NAFLD. Chapters 4 and 5 demonstrate a preference for...
5hmC generation only within genes of functional relevance for lipid synthesis and transport with gene groups of alternative function that could be considered reactive to rather than causal of hepatic lipid accumulation. These data are significantly strengthened by replication of these findings cross species and within a uniform hepatocyte like cell population abrogating the in vivo confounder of heterogeneous cell population.

The degree to which 5hmC levels are functionally important in the maintenance of transcriptional activity or a bystander reaction remains unclear. The consistent finding that 5hmC profiles are highly conserved, tissue specific and precisely regulated favours a functional role. Further, the detrimental effects of Tet methyl dioxygenase enzyme deletion in ESCs and mice suggests an essential role in accurate cell differentiation. Although studies have examined the effect of metabolic liver disease on global 5hmC levels and Tet expression, the requirement of intact Tet function for normal cellular response to over nutrition is unknown. The use of a Tet1 deficient mouse was employed to address this question. To our knowledge we demonstrate for the first time that interruption of normal Tet function conveys a strikingly protected metabolic phenotype with marked resistance to diet induced obesity, insulin resistance and NAFLD.
Figure 7.1 Schematic representation of experimental sequence in this thesis and avenues for future research. HFD = high fat diet, ESCheps = Embryonic stem cell derived hepatocytes. LPO = Lactate/pyruvate/octanoate intervention (2.2).
7.2 Limitations and opportunities

Despite a rapid increase in our understanding of contributory processes in NAFLD a highly unsatisfactory translational gap remains. Much of our knowledge comes from the use of murine and cell culture models that attempt to recapitulate phenotypic similarities to observations in humans and subsequent analysis of putative pathways and mechanisms. Low cost, isogenic subjects, genetic manipulation and rapid turnover make these models an attractive option. However, basic causal mechanisms in these models may bear little or no resemblance to the disease process in the human population. Mouse lipid biology may differ considerably from human, and cell culture models are unable to reflect the important complex interplay between the liver, adipose and muscle tissues known to be important in this disease process. Therefore, the findings of this work must be interpreted in the context of the considerable technical and biological limitations implicit in the field. In chapter 5 I utilised embryonic stem cell derived hepatocytes and identified this model as a novel method for examining hepatic 5hmC regulation in health and disease. Progression of the stem cell biology field along with the generation of 3D organoids and co-culture perfusion systems have the potential to revolutionise research in this field and may overcome many of the limitations of traditional murine and cancer cell models.

Technical and informatic processes are also limiting here. In our lab, the use of DNA immunoprecipitation followed by semiconductor sequencing provides a highly reproducible method of profiling the hydroxymethylome and correlates well with quantitative PCR analysis and glucosyl-sensitive restriction enzyme digestion PCR. However, while this technique will be highly accurate in localising regions of 5hmC enrichment, the ability to accurately detect a 10-20% difference at each highly enriched locus between cohorts is as yet unclear. In this study I have attempted to compensate for this by reporting a mean read count per window over an entire annotated region. The pipelines for performing these processes are non-uniform between research groups, a problem that could be solved in part by publication of syntax scripts. In expression analysis, ChIP and the processing of Infinium methylation data, the use of quartile normalisation, application of Bayesian statistics and multiple test corrections are routine due to the presence of Bioconductor packages such as Methylumi, Lumi, Limma, Deseq2 and MACS2. An opportunity exists to generate similar accessible tools for affinity based DNA immunoprecipitation to maximise the potential of these techniques and allow meaningful comparison between studies. Techniques for single based resolution analysis of 5hmC are progressing such as Tet assisted bisulphide sequencing (TAB-seq) and oxidative bisulfite sequencing (oxBS-Seq). Sequential reports suggest these
techniques are highly reproducible and show enormous potential, however at present they remain prohibitively expensive.

7.3 Future work

Despite the remarkable lean phenotype of Tet1 deficient mice, pharmacological inhibition of Tet1 is unlikely to be a viable therapeutic target. The ubiquitous epigenomic effects of Tet proteins and findings of haematological malignancy in Tet1−/− rodents are prohibitive. However, continued elucidation of the genes specifically effected by Tet1 modulation during high fat diet intervention may uncover novel pathways for interrogation. Intersecting hmeDIP-seq and Tet1 ChIP data sets in the context of NAFLD may be informative to this end.

Future experiments must determine the dominant tissue affected by Tet1 deficiency in generating the observed phenotype. For liver tissue, this can be done in a variety of ways.

1) Generation of a Tet1loxP knock out animal and breed with a Cyp1a1cre or Albcre with a drug inducible promoter for liver specific deletion of Tet1 in adulthood prior to HFD intervention.

2) Isolation of hepatocytes from Tet1−/− and Tet1+/− deficient animals and challenge in vitro with energy substrates and examine effect of lipid accumulation and induction of lipid synthesis genes.

3) Small interfering RNA knock down of ESCheps late in the differentiation process and challenge with LPO intervention.

In addition, a logical next step for this work is to examine in detail Tet expression and activity in human subjects and correlate with NAFLD stage to determine if the findings of chapters 4 and 5 are replicated in the human population. Characterised cohorts like the NASH-CRN database, the Dallas heart study cohort and the Edinburgh Type 2 Diabetes Study are accumulating well characterised populations with NAFLD of all stages that may facilitate this process. However, this will be challenging. The majority of tissue isolated from human NAFLD biopsies is formalin fixed paraffin blocks unsuitable for expression analysis with which to target 5hmC analysis. Further the heterogeneous nature of the human population with respect to age, drug exposure, genotype and dietary composition will likely require a high number of biological replicates to achieve adequate statistical power as previously reported. Thus far previous expression data sets have not highlighted Tets as up or downregulated in NAFLD subtypes likely due to low levels of Tet expression and the use of the relatively
insensitive micro array platform. The exploration of expression quantitative trait loci in large genotyped populations with NAFLD and correlation with outcome is an alternative approach.

7.4 Exciting themes in current NAFLD Research – Pathogenesis, biomarkers and interventions

The gut microbiome

Mice raised in a sterile environment with no gut flora have lower body fat mass, higher insulin sensitivity and lower hepatic fats than conventionally housed animals. Remarkably, this protective effect is lost within 14 days by colonisation of germ free mice with the faecal microbiome of conventionally housed mice despite a reduction in food intake. The ability to transfer metabolic phenotype in this way suggests an important and complex relationship between the composition of colonised microbiota, host metabolism and immunity and has opened up an exciting new field in obesity research. The bacterial population of the gut is important for the digestion and fermentation of food stuffs and thus may play an important role in the composition of nutrient availability and absorption. Differences in the bacterial composition of gut flora in obese and lean humans has been demonstrated with the recurrent finding of a reduction in the abundance in Bacteroidetes bacteria with a relative increase in Firmicutes, consistent with animal models. Multiple mechanisms have been suggested for this phenomenon including an increase in gut permeability termed ‘metabolic endotoxaemia’ in generating a chronic host immune response. High fat diet in rodents impairs normal gut barrier function and reduces expression of gut endothelial tight junction molecules. This may allow transition of inflammatory mediators across the epithelial wall altering the relationship between bacterial flora and the host immune system. The complex interplay between commensal flora and metabolic disease is a rapidly evolving area of translational research with real opportunities for intervention.

Non-invasive Biomarkers of progression to NASH

One significant hurdle to the implementation of NAFLD pharmacotherapy is the ability to target patients with NASH or significant fibrosis in a patient acceptable and non-invasive manner. Currently the gold standard for histological severity remains liver biopsy which is an invasive procedure and not without complication. Further, there remains considerable intra observer variability in biopsy interpretation. The possibility that 5hmC profiling as a marker of cell phenotype could be used to stratify disease severity is intriguing and worthy of further
research. The robust nature of DNA and objective output may be a useful adjunct to histological examination.

Thus a number of non-invasive predictive models have been designed and validated against histological data sets. CK18 fragments are released into serum following apoptosis of injured hepatocytes and can be measured by ELISA. Serum CK18 is significantly elevated in NASH versus simple steatosis, a finding validated in multiple data sets. Meta-analysis of these studies suggests a sensitivity and specificity for predicting NASH of 0.78 and 0.87 respectively. Thus CK-18 alone is not sufficient for predicting NASH but may be a useful adjunct to a predictive model. Other future potential biomarkers include, lipid peroxidation products, microRNAs and extracellular micro-particles. The exploration of 5hmC profiling of circulating cell free DNA as a marker of hepatic apoptosis is also an exciting prospect that has yet to be examined.

**Restoration of energy balance**

Finally, it cannot be denied that the rapid advance of NAFLD incidence in the human population is predominantly due to positive energy balance and sedentary living. The use of bariatric surgery and ultra-low calorie diets are highly effective but unsustainable on a global scale. As such the solution to this epidemic ultimately lies in public health education and intervention. However, instigating behavioural change on a population scale is remarkably difficult as evidenced by smoking habits. Despite a sequence of public health interventions over the past 60 years, 23% of 16-24 year olds in the UK are smokers. Notable success has been demonstrated targeting the younger population in Finland through a combination of urban planning, nutritionist input and community nurses interacting with the education board. Alternatively, the use of medications to induce satiety and reduce appetite may be employed, however recent attempts to modify appetite pharmacologically using cannabinoid receptor inhibition have been associated with unacceptable mental health side effects limiting use.

7.5 Conclusion

These initial experiments highlighting the importance of Tet activity in lipid metabolism have the potential to open an exciting new avenue of translational research not only through understanding fundamental biological mechanism but also NAFLD pathophysiology. Although Tet1 inhibition is an unlikely therapeutic target, exploration of annotated sites of
Tet1 activity in a hypothesis free approach may identify downstream regulators amenable to intervention. It is hoped that research of this nature can limit the disease burden of NAFLD until sustainable global solutions can be found.
13. Neuschwander-Tetri, B. A. *et al*. Clinical, laboratory and histological associations in


56. Speliotes, E. K. *et al.* Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits.


69. Bhala, N. et al. The natural history of nonalcoholic fatty liver disease with advanced


71. No Title. http://www.odt.nhs.uk/


89. Higuchi, N. et al. Liver X receptor in cooperation with SREBP-1c is a major lipid synthesis regulator in nonalcoholic fatty liver disease. Hepatol. Res. 38, 1122–9 (2008).
120, 756–767 (2010).


111. Smith, K. Liver disease: Kupffer cells regulate the progression of ALD and NAFLD.


127. Vilar-Gomez, E. et al. Weight Loss via Lifestyle Modification Significantly Reduces


140. Gallou-Kabani, C. *et al.* C57BL/6J and A/J mice fed a high-fat diet delineate


188. Ben-Shushan, E., Pikarsky, E., Klar, A. & Bergman, Y. Extinction of Oct-3/4 gene expression in embryonal carcinoma x fibroblast somatic cell hybrids is accompanied...


202. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in
173


233. Williams, K. et al. TET1 and hydroxymethylcytosine in transcription and DNA


248. <oncotarget-02-627.pdf>.

175


doi:10.1371/journal.pgen.1002154


279. Jin, S.-G., Kadam, S. & Pfeifer, G. P. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-


348. Rocquain, J. et al. Combined mutations of ASXL1, CBL, FLT3, IDH1, IDH2, JAK2,
KRAS, NPM1, NRAS, RUNX1, TET2 and WT1 genes in myelodysplastic syndromes and acute myeloid leukemias. *BMC Cancer* **10**, 401 (2010).


364. Chen, M. L. *et al.* Quantification of 5-methylcytosine and 5-hydroxymethylcytosine


419. Thomas, K. H., Martin, R. M., Potokar, J., Pirmohamed, M. & Gunnell, D. Reporting of drug induced depression and fatal and non-fatal suicidal behaviour in the UK from
Presentation and Publication of data from this thesis

Published Manuscripts


Abstracts


## APPENDICES

Appendix 1: Gene list and fold changes for 17 week high fat diet fed mouse liver tissue.

<table>
<thead>
<tr>
<th>Probe ID</th>
<th>Gene Symbol</th>
<th>Gene Name</th>
<th>logFC</th>
<th>Adjusted P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>3180750</td>
<td>Dbp</td>
<td>D site albumin promoter binding protein</td>
<td>1.54</td>
<td>0.0216</td>
</tr>
<tr>
<td>6560497</td>
<td>Nnmt</td>
<td>nicotinamide N-methyltransferase</td>
<td>1.39</td>
<td>0.0030</td>
</tr>
<tr>
<td>670653</td>
<td>Cyp17a1</td>
<td>cytochrome P450, family 17, subfamily a, polypeptide 1</td>
<td>1.38</td>
<td>0.0160</td>
</tr>
<tr>
<td>6420253</td>
<td>Msmo1</td>
<td>methylsterol monoxygenase 1</td>
<td>1.37</td>
<td>0.0015</td>
</tr>
<tr>
<td>6450056</td>
<td>Sqle</td>
<td>squalene epoxidase</td>
<td>1.37</td>
<td>0.0009</td>
</tr>
<tr>
<td>3360608</td>
<td>Cfd</td>
<td>complement factor D (adipsin)</td>
<td>1.30</td>
<td>0.0385</td>
</tr>
<tr>
<td>2070519</td>
<td>Nnmt</td>
<td>nicotinamide N-methyltransferase</td>
<td>1.30</td>
<td>0.0037</td>
</tr>
<tr>
<td>7610673</td>
<td>Apoa4</td>
<td>apolipoprotein A-IV</td>
<td>1.14</td>
<td>0.0065</td>
</tr>
<tr>
<td>5560600</td>
<td>Apoa4</td>
<td>apolipoprotein A-IV</td>
<td>1.08</td>
<td>0.0146</td>
</tr>
<tr>
<td>2350368</td>
<td>Mfsd2a</td>
<td>major facilitator superfamily domain containing 2A</td>
<td>1.08</td>
<td>0.0074</td>
</tr>
<tr>
<td>7210039</td>
<td>Gpc1</td>
<td>glypican 1</td>
<td>1.05</td>
<td>0.0032</td>
</tr>
<tr>
<td>3460133</td>
<td>Apoa4</td>
<td>apolipoprotein A-IV</td>
<td>1.04</td>
<td>0.0100</td>
</tr>
<tr>
<td>3370551</td>
<td>Nnmt</td>
<td>nicotinamide N-methyltransferase</td>
<td>1.00</td>
<td>0.0096</td>
</tr>
<tr>
<td>2650154</td>
<td>Aldh1b1</td>
<td>aldehyde dehydrogenase 1 family, member B1</td>
<td>0.98</td>
<td>0.0266</td>
</tr>
<tr>
<td>6940521</td>
<td>NA</td>
<td>NA</td>
<td>0.95</td>
<td>0.0077</td>
</tr>
<tr>
<td>6350333</td>
<td>Cyp2a5</td>
<td>cytochrome P450, family 2, subfamily a, polypeptide 5</td>
<td>0.86</td>
<td>0.0205</td>
</tr>
<tr>
<td>Gene Accession</td>
<td>Gene Name</td>
<td>Description</td>
<td>Log2 Fold Change</td>
<td>p-value</td>
</tr>
<tr>
<td>---------------</td>
<td>------------</td>
<td>------------------------------------------------------------------------------</td>
<td>------------------</td>
<td>---------</td>
</tr>
<tr>
<td>2260082</td>
<td><em>Chrna4</em></td>
<td>cholinergic receptor, nicotinic, alpha polypeptide 4</td>
<td>0.84</td>
<td>0.0449</td>
</tr>
<tr>
<td>150066</td>
<td><em>1810055G02Rik</em></td>
<td>RIKEN cDNA 1810055G02 gene</td>
<td>0.84</td>
<td>0.0260</td>
</tr>
<tr>
<td>5890553</td>
<td><em>Lss</em></td>
<td>lanosterol synthase</td>
<td>0.83</td>
<td>0.0030</td>
</tr>
<tr>
<td>6940711</td>
<td><em>Clstn3</em></td>
<td>calsyntenin 3</td>
<td>0.83</td>
<td>0.0446</td>
</tr>
<tr>
<td>7160193</td>
<td><em>Aqp4</em></td>
<td>aquaporin 4</td>
<td>0.82</td>
<td>0.0045</td>
</tr>
<tr>
<td>6560382</td>
<td><em>Dhcr24</em></td>
<td>24-dehydrocholesterol reductase</td>
<td>0.82</td>
<td>0.0061</td>
</tr>
<tr>
<td>7040243</td>
<td><em>Nrep</em></td>
<td>neuronal regeneration related protein</td>
<td>0.79</td>
<td>0.0272</td>
</tr>
<tr>
<td>1500180</td>
<td><em>Cyp2a5</em></td>
<td>cytochrome P450, family 2, subfamily a, polypeptide 5</td>
<td>0.79</td>
<td>0.0258</td>
</tr>
<tr>
<td>3520382</td>
<td><em>Upp2</em></td>
<td>uridine phosphorylase 2</td>
<td>0.77</td>
<td>0.0446</td>
</tr>
<tr>
<td>3140768</td>
<td>NA</td>
<td>NA</td>
<td>0.75</td>
<td>0.0207</td>
</tr>
<tr>
<td>840114</td>
<td><em>Cyp2a5</em></td>
<td>cytochrome P450, family 2, subfamily a, polypeptide 5</td>
<td>0.75</td>
<td>0.0229</td>
</tr>
<tr>
<td>1780594</td>
<td><em>Gnat1</em></td>
<td>guanine nucleotide binding protein, alpha transducing 1</td>
<td>0.75</td>
<td>0.0447</td>
</tr>
<tr>
<td>1070097</td>
<td><em>Hsd17b7</em></td>
<td>hydroxysteroid (17-beta) dehydrogenase 7</td>
<td>0.75</td>
<td>0.0142</td>
</tr>
<tr>
<td>4250228</td>
<td><em>Dhcr24</em></td>
<td>24-dehydrocholesterol reductase</td>
<td>0.73</td>
<td>0.0149</td>
</tr>
<tr>
<td>6220026</td>
<td><em>Zfp36</em></td>
<td>zinc finger protein 36</td>
<td>0.73</td>
<td>0.0116</td>
</tr>
<tr>
<td>2350324</td>
<td><em>Gstt3</em></td>
<td>glutathione S-transferase, theta 3</td>
<td>0.72</td>
<td>0.0075</td>
</tr>
<tr>
<td>6520022</td>
<td><em>Slc47a1</em></td>
<td>solute carrier family 47, member 1</td>
<td>0.71</td>
<td>0.0103</td>
</tr>
<tr>
<td>2100097</td>
<td><em>Mvd</em></td>
<td>mevalonate (diphospho) decarboxylase</td>
<td>0.70</td>
<td>0.0065</td>
</tr>
<tr>
<td>Gene ID</td>
<td>Gene Name</td>
<td>Description</td>
<td>Score</td>
<td>p-value</td>
</tr>
<tr>
<td>---------</td>
<td>--------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-------</td>
<td>---------</td>
</tr>
<tr>
<td>4280112</td>
<td>Dhcr7</td>
<td>7-dehydrocholesterol reductase</td>
<td>0.68</td>
<td>0.0229</td>
</tr>
<tr>
<td>1980639</td>
<td>Fads2</td>
<td>fatty acid desaturase 2</td>
<td>0.67</td>
<td>0.0045</td>
</tr>
<tr>
<td>5290671</td>
<td>Fdps</td>
<td>farnesyl diphosphate synthetase</td>
<td>0.66</td>
<td>0.0500</td>
</tr>
<tr>
<td>4070402</td>
<td>Gm4956</td>
<td>predicted gene 4956</td>
<td>0.63</td>
<td>0.0456</td>
</tr>
<tr>
<td>6370681</td>
<td>Dhcr24</td>
<td>24-dehydrocholesterol reductase</td>
<td>0.63</td>
<td>0.0166</td>
</tr>
<tr>
<td>3610082</td>
<td>Cxcl1</td>
<td>chemokine (C-X-C motif) ligand 1</td>
<td>0.60</td>
<td>0.0444</td>
</tr>
<tr>
<td>4260014</td>
<td>Them7</td>
<td>thioesterase superfamily member 7</td>
<td>0.60</td>
<td>0.0301</td>
</tr>
<tr>
<td>5260367</td>
<td>Cyp1a2</td>
<td>cytochrome P450, family 1, subfamily a, polypeptide 2</td>
<td>0.60</td>
<td>0.0456</td>
</tr>
<tr>
<td>290168</td>
<td>Sqle</td>
<td>squalene epoxidase</td>
<td>0.59</td>
<td>0.0036</td>
</tr>
<tr>
<td>5270082</td>
<td>Tef</td>
<td>thyrotroph embryonic factor</td>
<td>0.58</td>
<td>0.0227</td>
</tr>
<tr>
<td>1980470</td>
<td>Ero1lb</td>
<td>ERO1-like beta (S. cerevisiae)</td>
<td>-0.58</td>
<td>0.0065</td>
</tr>
<tr>
<td>6250193</td>
<td>Txnip</td>
<td>thioredoxin interacting protein</td>
<td>-0.58</td>
<td>0.0037</td>
</tr>
<tr>
<td>6220070</td>
<td>Elovl3</td>
<td>elongation of very long chain fatty acids (FEN1/Elo2, SUR4/Elo3, yeast)-like 3</td>
<td>-0.61</td>
<td>0.0149</td>
</tr>
<tr>
<td>70341</td>
<td>Txnip</td>
<td>thioredoxin interacting protein</td>
<td>-0.71</td>
<td>0.0036</td>
</tr>
<tr>
<td>7570292</td>
<td>Spon2</td>
<td>spondin 2, extracellular matrix protein</td>
<td>-0.76</td>
<td>0.0113</td>
</tr>
<tr>
<td>6290743</td>
<td>Fgf21</td>
<td>fibroblast growth factor 21</td>
<td>-0.76</td>
<td>0.0045</td>
</tr>
<tr>
<td>430037</td>
<td>Osigin1</td>
<td>oxidative stress induced growth inhibitor 1</td>
<td>-0.93</td>
<td>0.0420</td>
</tr>
</tbody>
</table>
Appendix 2: Gene list and fold changes for human ESC derived hepatocytes treated with LPO energy substrate cocktail versus untreated cells.

<table>
<thead>
<tr>
<th>Probe</th>
<th>Gene Symbol</th>
<th>Gene Name</th>
<th>Log FC</th>
<th>Adjusted P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ILMN_2412336</td>
<td>AKR1C2</td>
<td>aldo-keto reductase family 1, member C2</td>
<td>2.1812</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1687757</td>
<td>AKR1C4</td>
<td>aldo-keto reductase family 1, member C4</td>
<td>1.7285</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1801077</td>
<td>PLIN2</td>
<td>perilipin 2</td>
<td>1.6008</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1675706</td>
<td>APOA4</td>
<td>apolipoprotein A-IV</td>
<td>1.5015</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_2174437</td>
<td>CIDEC</td>
<td>cell death-inducing DFFA-like effector c</td>
<td>1.3658</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_2138765</td>
<td>PLIN2</td>
<td>perilipin 2</td>
<td>1.3521</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_2278335</td>
<td>AKR1B15</td>
<td>aldo-keto reductase family 1, member B15</td>
<td>1.1536</td>
<td>0.0015</td>
</tr>
<tr>
<td>ILMN_1672148</td>
<td>AKR1B10</td>
<td>aldo-keto reductase family 1, member B10 (aldose reductase)</td>
<td>1.1393</td>
<td>0.0004</td>
</tr>
<tr>
<td>ILMN_2219681</td>
<td>RBP2</td>
<td>retinol binding protein 2, cellular</td>
<td>1.1042</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1815203</td>
<td>HMGCS2</td>
<td>3-hydroxy-3-methylglutaryl-CoA synthase 2 (mitochondrial)</td>
<td>1.0833</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_2387385</td>
<td>IGFBP1</td>
<td>insulin-like growth factor binding protein 1</td>
<td>1.0332</td>
<td>0.0011</td>
</tr>
<tr>
<td>ILMN_2124802</td>
<td>MT1H</td>
<td>metallothionein 1H</td>
<td>1.0191</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1684306</td>
<td>S100A4</td>
<td>S100 calcium binding protein A4</td>
<td>0.9749</td>
<td>0.0491</td>
</tr>
<tr>
<td>ILMN_2188862</td>
<td>GDF15</td>
<td>growth differentiation factor 15</td>
<td>0.9451</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_1750974</td>
<td>S100A9</td>
<td>S100 calcium binding protein A9</td>
<td>0.9441</td>
<td>0.0223</td>
</tr>
<tr>
<td>ILMN_2349393</td>
<td>MKD</td>
<td>midkine (neurite growth-promoting factor 2)</td>
<td>0.9256</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1754055</td>
<td>APOA5</td>
<td>apolipoprotein A-V</td>
<td>0.8811</td>
<td>0.0002</td>
</tr>
<tr>
<td>ILMN_1736178</td>
<td>AEBP1</td>
<td>AE binding protein 1</td>
<td>0.8807</td>
<td>0.0000</td>
</tr>
<tr>
<td>Accession</td>
<td>Gene Symbol</td>
<td>Description</td>
<td>Score</td>
<td>P-Value</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>-------------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>ILMN_1807291</td>
<td>CYP1A1</td>
<td>cytochrome P450, family 1, subfamily A, polypeptide 1</td>
<td>0.8524</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_2108735</td>
<td>EEF1A2</td>
<td>eukaryotic translation elongation factor 1 alpha 2</td>
<td>0.8498</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_1804822</td>
<td>SRXN1</td>
<td>sulfiredoxin 1</td>
<td>0.8287</td>
<td>0.0004</td>
</tr>
<tr>
<td>ILMN_1747067</td>
<td>NPAS1</td>
<td>neuronal PAS domain protein 1</td>
<td>0.8221</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1735816</td>
<td>CYP4A11</td>
<td>cytochrome P450, family 4, subfamily A, polypeptide 11</td>
<td>0.8135</td>
<td>0.0084</td>
</tr>
<tr>
<td>ILMN_1713124</td>
<td>AKR1C3</td>
<td>aldo-keto reductase family 1, member C3</td>
<td>0.8127</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1684982</td>
<td>PDK4</td>
<td>pyruvate dehydrogenase kinase, isozyme 4</td>
<td>0.7880</td>
<td>0.0007</td>
</tr>
<tr>
<td>ILMN_2212999</td>
<td>KIF5C</td>
<td>kinesin family member 5C</td>
<td>0.7845</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_1767470</td>
<td>SCPEP1</td>
<td>serine carboxypeptidase 1</td>
<td>0.7749</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1666733</td>
<td>CXCL8</td>
<td>chemokine (C-X-C motif) ligand 8</td>
<td>0.7725</td>
<td>0.0433</td>
</tr>
<tr>
<td>ILMN_3307693</td>
<td>WFDC2</td>
<td>WAP four-disulfide core domain 2</td>
<td>0.7722</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1715401</td>
<td>MT1G</td>
<td>metallothionein 1G</td>
<td>0.7612</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_3249142</td>
<td>ZG16</td>
<td>zymogen granule protein 16</td>
<td>0.7494</td>
<td>0.0045</td>
</tr>
<tr>
<td>ILMN_2046073</td>
<td>LCT</td>
<td>lactase</td>
<td>0.7393</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_1684308</td>
<td>DEFB103A</td>
<td>defensin, beta 103A</td>
<td>0.7339</td>
<td>0.0421</td>
</tr>
<tr>
<td>ILMN_1717056</td>
<td>TXNRD1</td>
<td>thioredoxin reductase 1</td>
<td>0.7315</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_2352009</td>
<td>ACADVL</td>
<td>acyl-CoA dehydrogenase, very long chain</td>
<td>0.7271</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1829555</td>
<td>LCE6A</td>
<td>late cornified envelope 6A</td>
<td>0.7192</td>
<td>0.0092</td>
</tr>
<tr>
<td>ILMN_1744817</td>
<td>UGT1A1</td>
<td>UDP glucuronosyltransferase 1 family, polypeptide A1</td>
<td>0.7063</td>
<td>0.0005</td>
</tr>
<tr>
<td>ILMN_2121774</td>
<td>ZG16</td>
<td>zymogen granule protein 16</td>
<td>0.6997</td>
<td>0.0049</td>
</tr>
<tr>
<td>Gene ID</td>
<td>Gene Symbol</td>
<td>Description</td>
<td>Log2 Fold Change</td>
<td>p-Value</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td>------------------------------------------------------------------------------</td>
<td>------------------</td>
<td>---------</td>
</tr>
<tr>
<td>ILMN_1728445</td>
<td><em>IGFBP1</em></td>
<td>insulin-like growth factor binding protein 1</td>
<td>0.6860</td>
<td>0.0161</td>
</tr>
<tr>
<td>ILMN_2388484</td>
<td><em>MAP2</em></td>
<td>microtubule-associated protein 2</td>
<td>0.6791</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_1757387</td>
<td><em>UCHL1</em></td>
<td>ubiquitin carboxyl-terminal esterase L1 (ubiquitin thiolesterase)</td>
<td>0.6728</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1757406</td>
<td><em>HIST1H1C</em></td>
<td>histone cluster 1, H1c</td>
<td>0.6594</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_1704531</td>
<td><em>PTGRI</em></td>
<td>prostaglandin reductase 1</td>
<td>0.6563</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_1758731</td>
<td><em>CYP2J2</em></td>
<td>cytochrome P450, family 2, subfamily J, polypeptide 2</td>
<td>0.6558</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1737298</td>
<td><em>MAT2A</em></td>
<td>methionine adenosyltransferase II, alpha</td>
<td>0.6490</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1789733</td>
<td><em>CLIP3</em></td>
<td>CAP-GLY domain containing linker protein 3</td>
<td>0.6388</td>
<td>0.0003</td>
</tr>
<tr>
<td>ILMN_2324421</td>
<td><em>TXNRD1</em></td>
<td>thioredoxin reductase 1</td>
<td>0.6381</td>
<td>0.0006</td>
</tr>
<tr>
<td>ILMN_1801216</td>
<td><em>S100P</em></td>
<td>S100 calcium binding protein P</td>
<td>0.6369</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1779448</td>
<td><em>EFHD1</em></td>
<td>EF-hand domain family, member D1</td>
<td>0.6355</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1731948</td>
<td><em>PCK1</em></td>
<td>phosphoenolpyruvate carboxykinase 1 (soluble)</td>
<td>0.6347</td>
<td>0.0069</td>
</tr>
<tr>
<td>ILMN_2065773</td>
<td><em>SCG5</em></td>
<td>secretogranin V</td>
<td>0.6296</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_1707339</td>
<td><em>BTG3</em></td>
<td>BTG family, member 3</td>
<td>0.6265</td>
<td>0.0002</td>
</tr>
<tr>
<td>ILMN_1654262</td>
<td><em>ZMAT3</em></td>
<td>zinc finger, matrin-type 3</td>
<td>0.6218</td>
<td>0.0014</td>
</tr>
<tr>
<td>ILMN_1653200</td>
<td><em>SLC22A17</em></td>
<td>solute carrier family 22, member 17</td>
<td>0.6176</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_1813175</td>
<td><em>ADGRL1</em></td>
<td>adhesion G protein-coupled receptor L1</td>
<td>0.6154</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1794190</td>
<td><em>CCPG1</em></td>
<td>cell cycle progression 1</td>
<td>0.6087</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1700081</td>
<td><em>FST</em></td>
<td>follistatin</td>
<td>0.6082</td>
<td>0.0050</td>
</tr>
<tr>
<td>ILMN_1786388</td>
<td><em>RNF113A</em></td>
<td>ring finger protein 113A</td>
<td>0.6081</td>
<td>0.0012</td>
</tr>
<tr>
<td>Gene ID</td>
<td>Symbol</td>
<td>Description</td>
<td>Log2 Fold Change</td>
<td>P Value</td>
</tr>
<tr>
<td>------------</td>
<td>--------</td>
<td>----------------------------------------------------------</td>
<td>-----------------</td>
<td>----------</td>
</tr>
<tr>
<td>ILMN_1760414</td>
<td>AADAC</td>
<td>arylacetamide deacetylase</td>
<td>0.6064</td>
<td>0.0005</td>
</tr>
<tr>
<td>ILMN_1757467</td>
<td>HIF0</td>
<td>H1 histone family, member 0</td>
<td>0.5994</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1657744</td>
<td>FAM219B</td>
<td>family with sequence similarity 219, member B</td>
<td>0.5931</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_2401344</td>
<td>PPP2R2C</td>
<td>protein phosphatase 2, regulatory subunit B, gamma</td>
<td>0.5879</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_1670899</td>
<td>FBN2</td>
<td>fibrillin 2</td>
<td>-0.5807</td>
<td>0.0004</td>
</tr>
<tr>
<td>ILMN_1713807</td>
<td>MAN1C1</td>
<td>mannosidase, alpha, class 1C, member 1</td>
<td>-0.5830</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_1698213</td>
<td>RBM3</td>
<td>RNA binding motif (RNP1, RRM) protein 3</td>
<td>-0.5834</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1776905</td>
<td>TMEM236</td>
<td>transmembrane protein 236</td>
<td>-0.5874</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1701613</td>
<td>RARRES3</td>
<td>retinoic acid receptor responder (tazarotene induced) 3</td>
<td>-0.5900</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_1695658</td>
<td>KIF20A</td>
<td>kinesin family member 20A</td>
<td>-0.5951</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_2179083</td>
<td>LOXL4</td>
<td>lysyl oxidase-like 4</td>
<td>-0.5980</td>
<td>0.0130</td>
</tr>
<tr>
<td>ILMN_1728049</td>
<td>S100A16</td>
<td>S100 calcium binding protein A16</td>
<td>-0.5983</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1800573</td>
<td>RPS21</td>
<td>ribosomal protein S21</td>
<td>-0.6019</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_2218104</td>
<td>PAH</td>
<td>phenylalanine hydroxylase</td>
<td>-0.6032</td>
<td>0.0002</td>
</tr>
<tr>
<td>ILMN_1763837</td>
<td>ANPEP</td>
<td>alanyl (membrane) aminopeptidase</td>
<td>-0.6060</td>
<td>0.0021</td>
</tr>
<tr>
<td>ILMN_2213297</td>
<td>C11orf54</td>
<td>chromosome 11 open reading frame 54</td>
<td>-0.6078</td>
<td>0.0025</td>
</tr>
<tr>
<td>ILMN_1734773</td>
<td>PRSSI</td>
<td>protease, serine, 1 (trypsin 1)</td>
<td>-0.6138</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1753584</td>
<td>KRT8</td>
<td>keratin 8, type II</td>
<td>-0.6199</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_1784364</td>
<td>STARD5</td>
<td>StAR-related lipid transfer (START) domain containing 5</td>
<td>-0.6215</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1728262</td>
<td>SAA2</td>
<td>serum amyloid A2</td>
<td>-0.6296</td>
<td>0.0006</td>
</tr>
<tr>
<td>Locus</td>
<td>Ensembl</td>
<td>Description</td>
<td>Log2 Fold Change</td>
<td>P-Value</td>
</tr>
<tr>
<td>------------</td>
<td>-----------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------------------</td>
<td>---------</td>
</tr>
<tr>
<td>ILMN_1661595</td>
<td>C1orf53</td>
<td>chromosome 1 open reading frame 53</td>
<td>-0.6312</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_2125869</td>
<td>ACTA1</td>
<td>actin, alpha 1, skeletal muscle</td>
<td>-0.6375</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1690866</td>
<td>KLHL41</td>
<td>kelch-like family member 41</td>
<td>-0.6382</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_1679262</td>
<td>DPYSL3</td>
<td>dihydrophosphorymidinase-like 3</td>
<td>-0.6412</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_3251409</td>
<td>PDE6A</td>
<td>phosphodiesterase 6A, cGMP-specific, rod, alpha</td>
<td>-0.6480</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1777325</td>
<td>STAT1</td>
<td>signal transducer and activator of transcription 1, 91kDa</td>
<td>-0.6488</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1702489</td>
<td>TRIM63</td>
<td>tripartite motif containing 63, E3 ubiquitin protein ligase</td>
<td>-0.6510</td>
<td>0.0002</td>
</tr>
<tr>
<td>ILMN_1760087</td>
<td>SLC26A3</td>
<td>solute carrier family 26 (anion exchanger), member 3</td>
<td>-0.6528</td>
<td>0.0025</td>
</tr>
<tr>
<td>ILMN_1737517</td>
<td>RPL29</td>
<td>ribosomal protein L29</td>
<td>-0.6533</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_1791726</td>
<td>TUBB3</td>
<td>tubulin, beta 3 class III</td>
<td>-0.6546</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1788122</td>
<td>GSTA5</td>
<td>glutathione S-transferase alpha 5</td>
<td>-0.6562</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_1716859</td>
<td>TDO2</td>
<td>tryptophan 2,3-dioxygenase</td>
<td>-0.6566</td>
<td>0.0007</td>
</tr>
<tr>
<td>ILMN_1698246</td>
<td>MTMR11</td>
<td>myotubularin related protein 11</td>
<td>-0.6582</td>
<td>0.0020</td>
</tr>
<tr>
<td>ILMN_1719543</td>
<td>MAF</td>
<td>v-maf avian musculoaponeurotic fibrosarcoma oncogene homolog</td>
<td>-0.6651</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1754576</td>
<td>KRT6C</td>
<td>keratin 6C, type II</td>
<td>-0.6671</td>
<td>0.0186</td>
</tr>
<tr>
<td>ILMN_1720710</td>
<td>HSPB3</td>
<td>heat shock 27kDa protein 3</td>
<td>-0.6698</td>
<td>0.0004</td>
</tr>
<tr>
<td>ILMN_2358074</td>
<td>MAD1L1</td>
<td>MAD1 mitotic arrest deficient-like 1 (yeast)</td>
<td>-0.6732</td>
<td>0.0005</td>
</tr>
<tr>
<td>ILMN_1755720</td>
<td>SLC2A2</td>
<td>solute carrier family 2 (facilitated glucose transporter), member 2</td>
<td>-0.6745</td>
<td>0.0019</td>
</tr>
<tr>
<td>ILMN_2133205</td>
<td>GPX2</td>
<td>glutathione peroxidase 2</td>
<td>-0.6750</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1696302</td>
<td>FABP5</td>
<td>fatty acid binding protein 5 (psoriasis-associated)</td>
<td>-0.6807</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_2219002</td>
<td>KRT6A</td>
<td>keratin 6A, type II</td>
<td>-0.6820</td>
<td>0.0062</td>
</tr>
<tr>
<td>ILMN_1789648</td>
<td>SCGN</td>
<td>secretagogin, EF-hand calcium binding protein</td>
<td>-0.6949</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_1707975</td>
<td>SERPIND1</td>
<td>serpin peptidase inhibitor, clade D (heparin cofactor), member 1</td>
<td>-0.6952</td>
<td>0.0015</td>
</tr>
<tr>
<td>ILMN_1729251</td>
<td>MYH4</td>
<td>myosin, heavy chain 4, skeletal muscle</td>
<td>-0.7016</td>
<td>0.0006</td>
</tr>
<tr>
<td>ILMN_2389054</td>
<td>CLDN18</td>
<td>claudin 18</td>
<td>-0.7294</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1671337</td>
<td>SLC2A5</td>
<td>solute carrier family 2 (facilitated glucose/fructose transporter), member 5</td>
<td>-0.7301</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_1808677</td>
<td>UGT2B17</td>
<td>UDP glucuronosyltransferase 2 family, polypeptide B17</td>
<td>-0.7335</td>
<td>0.0005</td>
</tr>
<tr>
<td>ILMN_1771051</td>
<td>RPL29</td>
<td>ribosomal protein L29</td>
<td>-0.7340</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_1769547</td>
<td>DIO1</td>
<td>deiodinase, iodothyronine, type I</td>
<td>-0.7405</td>
<td>0.0008</td>
</tr>
<tr>
<td>ILMN_2414786</td>
<td>DIO1</td>
<td>deiodinase, iodothyronine, type I</td>
<td>-0.7417</td>
<td>0.0008</td>
</tr>
<tr>
<td>ILMN_1754247</td>
<td>SLC3A1</td>
<td>solute carrier family 3 (amino acid transporter heavy chain), member 1</td>
<td>-0.7464</td>
<td>0.0022</td>
</tr>
<tr>
<td>ILMN_1662214</td>
<td>PAH</td>
<td>phenylalanine hydroxylase</td>
<td>-0.7481</td>
<td>0.0003</td>
</tr>
<tr>
<td>ILMN_1709847</td>
<td>KCNJ13</td>
<td>potassium channel, inwardly rectifying subfamily J, member 13</td>
<td>-0.7489</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_1691364</td>
<td>STAT1</td>
<td>signal transducer and activator of transcription 1, 91kDa</td>
<td>-0.7504</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1667018</td>
<td>ACE2</td>
<td>angiotensin I converting enzyme 2</td>
<td>-0.7509</td>
<td>0.0006</td>
</tr>
<tr>
<td>ILMN_1905548</td>
<td>MBNL3</td>
<td>muscleblind-like splicing regulator 3</td>
<td>-0.7587</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1803429</td>
<td>CD44</td>
<td>CD44 molecule (Indian blood group)</td>
<td>-0.7602</td>
<td>0.0035</td>
</tr>
<tr>
<td>ILMN_2400935</td>
<td>TAGLN</td>
<td>transgelin</td>
<td>-0.7664</td>
<td>0.0003</td>
</tr>
<tr>
<td>ILMN_1766675</td>
<td>CDH6</td>
<td>cadherin 6, type 2, K-cadherin (fetal kidney)</td>
<td>-0.7721</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_1694588</td>
<td>C4BPB</td>
<td>complement component 4 binding protein, beta</td>
<td>-0.7769</td>
<td>0.0000</td>
</tr>
<tr>
<td>Gene ID</td>
<td>Gene Symbol</td>
<td>Description</td>
<td>p-value</td>
<td>FDR</td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>ILMN_1663390</td>
<td>CDC20</td>
<td>cell division cycle 20</td>
<td>-0.7866</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_2207504</td>
<td>LEP</td>
<td>leptin</td>
<td>-0.7930</td>
<td>0.0014</td>
</tr>
<tr>
<td>ILMN_2374449</td>
<td>SPP1</td>
<td>secreted phosphoprotein 1</td>
<td>-0.8197</td>
<td>0.0041</td>
</tr>
<tr>
<td>ILMN_1690105</td>
<td>STAT1</td>
<td>signal transducer and activator of transcription 1, 91kDa</td>
<td>-0.8235</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1823750</td>
<td>JAKMIP3</td>
<td>Janus kinase and microtubule interacting protein 3</td>
<td>-0.8251</td>
<td>0.0002</td>
</tr>
<tr>
<td>ILMN_1734176</td>
<td>CGA</td>
<td>glycoprotein hormones, alpha polypeptide</td>
<td>-0.8286</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1778668</td>
<td>TAGLN</td>
<td>transgelin</td>
<td>-0.8408</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_2239408</td>
<td>RNASE4</td>
<td>ribonuclease, RNase A family, 4</td>
<td>-0.8409</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1737041</td>
<td>HABP2</td>
<td>hyaluronan binding protein 2</td>
<td>-0.8426</td>
<td>0.0002</td>
</tr>
<tr>
<td>ILMN_2146761</td>
<td>FABP5</td>
<td>fatty acid binding protein 5 (psoriasis-associated)</td>
<td>-0.8459</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1776112</td>
<td>SLC10A1</td>
<td>solute carrier family 10 (sodium/bile acid cotransporter), member 1</td>
<td>-0.8588</td>
<td>0.0113</td>
</tr>
<tr>
<td>ILMN_1716925</td>
<td>FSIP1</td>
<td>fibrous sheath interacting protein 1</td>
<td>-0.8645</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_1651354</td>
<td>SPP1</td>
<td>secreted phosphoprotein 1</td>
<td>-0.8648</td>
<td>0.0073</td>
</tr>
<tr>
<td>ILMN_2162819</td>
<td>UGT2B11</td>
<td>UDP glucuronosyltransferase 2 family, polypeptide B11</td>
<td>-0.8651</td>
<td>0.0002</td>
</tr>
<tr>
<td>ILMN_1771544</td>
<td>FMO9P</td>
<td>flavin containing monoxygenase 9 pseudogene</td>
<td>-0.8997</td>
<td>0.0092</td>
</tr>
<tr>
<td>ILMN_1764690</td>
<td>NTS</td>
<td>neurotensin</td>
<td>-0.9169</td>
<td>0.0017</td>
</tr>
<tr>
<td>ILMN_1808114</td>
<td>LYVE1</td>
<td>lymphatic vessel endothelial hyaluronan receptor 1</td>
<td>-0.9200</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1810233</td>
<td>UGT2B11</td>
<td>UDP glucuronosyltransferase 2 family, polypeptide B11</td>
<td>-0.9391</td>
<td>0.0008</td>
</tr>
<tr>
<td>ILMN_1808494</td>
<td>ITIH2</td>
<td>inter-alpha-trypsin inhibitor heavy chain 2</td>
<td>-0.9481</td>
<td>0.0041</td>
</tr>
<tr>
<td>ILMN_1766955</td>
<td>VCAM1</td>
<td>vascular cell adhesion molecule 1</td>
<td>-0.9890</td>
<td>0.0000</td>
</tr>
<tr>
<td>Gene ID</td>
<td>Gene Symbol</td>
<td>Description</td>
<td>Log2 Fold Change</td>
<td>P-value</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>-------------------------------------------------------------</td>
<td>------------------</td>
<td>----------</td>
</tr>
<tr>
<td>ILMN_1750234</td>
<td>PRSS2</td>
<td>protease, serine, 2 (trypsin 2)</td>
<td>-0.9950</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1685699</td>
<td>PRSS3</td>
<td>protease, serine, 3</td>
<td>-1.0235</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_2307903</td>
<td>VCAM1</td>
<td>vascular cell adhesion molecule 1</td>
<td>-1.0786</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_1781859</td>
<td>UGT2B28</td>
<td>UDP glucuronosyltransferase 2 family, polypeptide B28</td>
<td>-1.0927</td>
<td>0.0001</td>
</tr>
<tr>
<td>ILMN_2297626</td>
<td>PEG10</td>
<td>paternally expressed 10</td>
<td>-1.1448</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1742444</td>
<td>UGT2B10</td>
<td>UDP glucuronosyltransferase 2 family, polypeptide B10</td>
<td>-1.2029</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1777797</td>
<td>AFM</td>
<td>afamin</td>
<td>-1.2589</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1685043</td>
<td>CYP3A7</td>
<td>cytochrome P450, family 3, subfamily A, polypeptide 7</td>
<td>-1.3091</td>
<td>0.0002</td>
</tr>
<tr>
<td>ILMN_1678841</td>
<td>UBD</td>
<td>ubiquitin D</td>
<td>-1.3929</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1696284</td>
<td>CLDN18</td>
<td>claudin 18</td>
<td>-1.4834</td>
<td>0.0000</td>
</tr>
<tr>
<td>ILMN_1780575</td>
<td>CRP</td>
<td>C-reactive protein, pentraxin-related</td>
<td>-1.7992</td>
<td>0.0018</td>
</tr>
</tbody>
</table>