Investigating the nexus between chromatin and metabolism

.



Silvia Raineri
St. Catherine's College
Trinity Term 2020
Doctor of Philosophy, University of Oxford

Investigating the nexus between chromatin and metabolism

A thesis submitted to the University of Oxford for the degree of Doctor of Philosophy in Biochemistry

Silvia Raineri, St Catherine's College, Trinity Term 2020

Word Count: 49239

Abstract

Metabolism and epigenetics are tightly connected within the cell, as many chromatin-modifying enzymes use metabolites as co-factors or substrates, and the energetic state of the cell can drive the activation of several transcriptional programmes. This thesis examines the nexus between chromatin and metabolism, both at the systemic and cellular level, in two different projects. First, the impact on a panel of weight-loss drugs on the gut microbiome of obese rats is investigated through shotgun metagenomic sequencing. Results highlight a concomitant change in weight and gut microbiome composition following treatment with Sibutramine, which acts through inhibition of appetite.

A second project focuses on alterations induced by a gain-of-function mutation in the metabolic enzyme Isocitrate Dehydrogenase I (IDH1), which spurs production of the oncometabolite 2-hydroxyglutarate (2HG), in the context of glioma. Using an isogenic human glioma cell line (U87), either wild-type or bearing the R132H IDH1 mutation, an IDH1 inhibitor (GSK864) and a combination of multi-omics techniques, we characterize the metabolic and epigenetic alterations induced. Results suggest an uncoupling between metabolism, transcriptome and the epigenome, as treatment with an IDH1 inhibitor, while successfully rescuing the metabolic defect and promoting, upon a longer treatment, changes in gene expression, did not correct the epigenetic alterations.

Experimental delays due to the Covid19 pandemic

To assess DNA methylation levels in our cellular model, we set up a collaboration with Prof. Kriaucionis laboratory to perform TET-assisted pyridine borane sequencing (TAPS) on WT and R132H IDH1 U87 cells following a week-long treatment with GSK864. At the time of writing, due to the Covid19 epidemic this experiment has been seriously delayed and is currently on queue to be sequenced. Results from this dataset will be key in confirming whether the loss of CTCF detected corresponds to DNA sequences that have been hypermethylated in R132H cells.

Acknowledgements

First and foremost, I would like to thank my supervisor, Prof. Jane Mellor, for the constant support, guidance, and for the precious scientific discussions. Working in the Mellor lab has been a great formative experience, which I thoroughly enjoyed, and I will always be grateful for the opportunity she has given me.

This thesis would not have been possible without the help of our collaborators: Dr. Nicholas Ilott for his help in running the CGAT pipeline on our shotgun metagenomics sequencing dataset; Prof. Skirmantas Kriaucionis and his lab, for collaborating with us on the DNA methylation sequencing; Dr. Oscar Yanes and all the members of his lab, particularly Joan Mirò Blanch, for welcoming me during my secondment and for their collaboration with the metabolomics experiments.

Parts of this thesis also include a published Perspective article written by me and Prof. Jane Mellor, from whom I have received permission to use this material (Raineri & Mellor, 2018).

A special thanks to Drs Nicholas Ilott, Charlotte George, Thomas Brown, Struan Murray and Salva Casanì for the bioinformatics advice provided over the years. As a beginner, I have learned a lot about this new field through our discussions, and I have always appreciated the time and patience you have offered.

Last, but not least, I would also like to thank past and current members of the Mellor lab, particularly Anna Lamstaes, Ülkü Uzun, Meredith Wouters, Harry Fischl and Philip Lorenz, for creating a welcoming and friendly environment, which has made facing the ups and occasional downs of a PhD a lot easier.

Publications

Raineri, S., and Mellor, J. (2018). IDH1: Linking Metabolism and Epigenetics. Front. Genet. *9*, 493.

List of abbreviations

2-hydroxy-glutarate	2HG
Capturing chromosome conformation	3C
3D-structure illumination microscopy	3D-SIM
Acute myeloid leukemia	AML
Area under curve	AUC
Body mass index	BMI
Chromatin Immunoprecipitation followed by sequencing	ChIP-Seq
Central nervous system	CNS
CCCTC-binding factor	CTCF
DNA damage response	DDR
Food and drug administration	FDA
Fumarate hydratase	FH
Faecal microbiota transplantation	FMT
Gas chromatography followed by mass spectrometry	GC/MS
Generalized effect size	GES
Gastrointestinal stromal tumours	GIST
Gene ontology	GO
Gene set enrichment analysis	gsea
Proton-magnetic resonance spectroscopy	H-MRS
Acetyltransferases	HATs
High fat diet	HFD
High throughput conformation capture	HiC
Isocitrate dehydrogenase I	IDH1
Integrative genome browser	IGV
Interleukin 8	IL8
Immunoprecipitation	IP
Jumonji	JMJ
Lysine acetyltransferases	KATs
Lysine deacetylases	KDACs
Kyoto encyclopedia of genes and genomes	KEGG
Liquid chromatography followed by mass spectrometry	LC/MS
Log2 fold -change	Log2FC
Lipopolysaccharide	LPS
Nuclear lamina associated domains	NADs
Normalized enrichment score	NES
Oral glucose tolerance test	OGTT
Adjusted P-Value	P.adj
Principal component analysis	PCA
Programmed cell-death protein 1	PD1
PDGF Receptor alpha	PDGFRA

Pentose phosphate pathway	PPP
Quantitative real-time PCR	qPCR
3' end mRNA sequencing	RNASeq
S-adenosyl-homocysteine	SAH
S-adenosyl-methyonine	SAM
Short-chain fatty acids	SCFAs
Succinate dehydrogenase	SDH
Topologically associated domains	TADs
TET-assisted pyridine borane sequencing	TAPS
Tricarboxylic acid cycle	TCA
The cancer genome atlas	TCGA
Ten-eleven translocation	TET
Toll-like Receptor	TLR
Toll-like Receptor 5	TLR5
Trimmed mean of M values	TMM
Transcription start site	TSS

Table of Contents

Abstract 2
Publications 5
Table of Contents 8
Table of Figures 12
1. Introduction 15
1.1 Epigenetic modifications on chromatin and DNA fine tune gene expression 15
1.2 AcetylCoA is the primary substrate for histone acetylation 18
1.3 DNA and Histone methylation is promoted by one-carbon metabolism 20
1.4 Coordination between epigenetics and metabolism 23
1.5 The metabolism-epigenetics axis in disease development 25
1.6 Aims 26
Metagenomic analysis of microbiomes in drug-treated rats 28
2. Metagenomic analysis of microbiomes in drug-treated rats 29
2.1 Introduction 29
2.1.1 Role of the gut microbiome in human metabolism 29
2.1.2 Microbiome alterations in disease states 30
2.2 Materials and Methods 34
2.2.1 Faecal collection protocol 34
2.2.2 In vivo experimental design 35
2.2.3 Oral glucose tolerance test (OGTT) 36
2.2.4 Faecal DNA extraction 36
2.2.5 Shotgun metagenomics sequencing 37
2.2.6 Bioinformatics Analysis 38
2.2.7 Preprocessing files 38
2.3 Results 41
2.3.1 Differential effects on weight loss after long term treatment with a panel of drugs 41
2.3.2 Treatment with Sibutramine increases Insulin sensitivity 43
2.3.3 Sibutramine affects food intake 46
2.3.4 Impact of weight loss drugs on the microbiome 50
2.3.5 Sibutramine impacts the microbiome 52
2.3.6 Analysis of differential species confirms an association between Sibutramine treatment and gut microbiome composition 53
2.3.7 Functional analysis of rats microbiome reveals significant decrease of "Flagellar Assembly" related-genes 56
2.4 Discussion 61
IDH1: Linking chromatin and metabolism 65

3. Materials and Methods 66
3.1 RNA Sequencing 66
3.1.1 48-hours treatment with GSK864 of U87 WT and R132H IDH1 cells 66
3.1.2 One week treatment with GSK864 of U87 WT and R132H IDH1 cells 67
3.1.3 RNA extraction and library preparation 67
3.1.4 Bioinformatic Analysis 68
3.2 Reverse transcription-polymerase chain reaction (RT-PCR) and real time PCR 69
3.2.1 Reverse Transcription 69
3.2.2 PCR 72
3.2.3 Real time PCR 72
3.3 Sample preparation for TAPS sequencing 72
3.4 Chromatin Immunoprecipitation followed by sequencing (ChIPSeq) 73
3.4.1 Immunoprecipitation 73
3.4.2 Quality Check: qPCR 74
3.4.3 Library Preparation and Sequencing 74
3.4.4 Bioinformatics Analysis 75
3.5 Metabolic Assays: Glucose and Glutamine Uptake 77
3.6 Metabolic Flux of Glucose and Glutamine 77
3.6.1 Sample preparation for Mass Spectrometry 79
3.7 Immunofluorescence on H3K9me3, H4K5ac and H3K27ac levels in WT or R132H IDH1 cells 83
3.8 Bioinformatics analyses 85
4. Metabolic characterization of U87 WT and R132H cells 86
4.1 Introduction[1] 86
4.2 Discrepancies between literature studies on the metabolic alterations in IDH1 mutant cells 88
4.3 Metabolic Flux Optimization: Glucose and Glutamine Uptake 93
4.4 GSK864 reduces levels of Glutamine-derived 2HG in R132H cells 94
4.5 Methionine cycle-related metabolites levels remain unchanged in IDH1 mutant cells 97
4.6 Increased levels of TCA cycle-related metabolites in R132H cells 100
4.7 Glutamine fuels both TCA cycle and reductive carboxylation 103
4.8 Discussion 107
5. Characterization of gene expression in R132H U87 cells using RNASeq 110
5.1 IDH1 mutation in Glioma patients 110
5.2 Assessment of gene expression differences following GSK864 treatment in U87 WT or R132H cells 114
5.3 RNA Sequencing of U87 cells reveals little effect of GSK864 treatment on R132H cells gene expression 114

5.4 Gene expression differences between R132H and WT cells 118
5.5 1 week treatment with GSK864 has greater impact on gene expression 124
5.6 Gsea analysis of 1 week dataset hints at increased Fatty AcylCoA Biosynthesis ..130
5.7 Discussion 137
6. Characterization of epigenetic alterations in R132H cells with ChIPSeq 142
6.1 Aberrant histone methylation in IDH1 mutated cells impairs different cellular processes 142
6.2 Characterization of histone H3K9 methylation in R132H cells with ChIPSeq 144
6.3 Assessment of CTCF binding in R132H IDH1 mutated cells 148
6.4 No effect of GSK864 on CTCF or H3K9me3 levels in R132H cells 156
6.5 Discussion 158
6.6 Note on the assessment of genome-wide DNA methylation levels in R132H cells with TAPS 161
7. Assessment of histone H3K27ac in IDH1 R132H cells using ChIP-Seq 163
7.1 Potential uncoupling between gene expression and H3K27ac around promoters in R132H cells 166
7.2 EnhancerAtlas annotation confirms uncoupling between gene expression and enhancer acetylation 179
7.3 Discussion 182
8. Deconstructing the models proposing loss of insulation of chromatin domains in R132H cells[1] 185
8.1 Adding a new piece to the puzzle: Does the IDH1 mutation influence the formation of chromatin domains? 185
8.2 From local to a genome-wide mechanism: challenges ahead 192
8.3 Deconstructing the model based on Flavahan (2016) 193
8.4 Assessing the overlap of CTCF binding sites with lost boundaries from Flavahan et al. 2016 197
8.5 PDGFRA expression is not a suitable readout for loss of insulation between TADs in R132H IDH1 U87 cells 203
8.6 1 week dataset confirms DNAJC12 upregulation in R132H cells and little to no change in PDGFRA levels 207
8.7 No change of CTCF binding around PDGFRA insulator 211
8.8 No detection of a CTCF peak around Flavahan's annotated DNAJC12 boundary 218
8.9 No CTCF binding site at PTPRN Flavahan-predicted boundary 222
8.10 No significant change in CTCF binding detected at the SOX2 locus 226
8.11 Future directions 230
8.12 Discussion 231
9. Effect of IDH1 R132H mutation on levels of H3K9me3 on Chromosome X 236
9.1 Heterochromatin role in genome organization within the nucleus 236

9.2 Loss of H3K9me3 on chromosomes 18 and X 238
9.3 Loss of H3K9me3 around Transposable Elements 245
9.4 Investigating morphological changes in the nucleus and chromatin marks distribution of R132H cells using 3D-SIM 249
9.5 Discussion 250
10. Discussion 252
10.1 Summary 252
10.2 Future directions 254
10.3 Concluding remarks 257
11. Appendix 258
Bibliography 263

Table of Figures

Figure 1.1. Three distinct pools contribute to AcetylCoA cellular concentrations. *20*
Figure 1.2. Folate and Methionine cycles participate to SAM production *22*
Figure 2.1. Weight of rats over the course of the treatment *41*
Figure 2.2. Sibutramine and RDCOB, alone or in combination, improve Insulin secretion in obese rats. *45*
Figure 2.3 Sibutramine reduces food intake in obese rats. *47*
Figure 2.4 Tukey's post hoc test on Sibutramine reveals significant timepoint association at Day 2 and 3. ..... *49*
Figure 2.5 Diversity analyses on all treatment groups reveal no clear alteration of alpha diversity, but few differences in beta diversity. *52*
Figure 2.6 Sibutramine treatment effect on the microbiome. *53*
Figure 2.7 Differentially enriched species between rats before (Day -3) and after (Day 42) Sibutramine treatment. *54*
Figure 2.8. Differentially abundant genes between Sibutramine-treated rats at the start and end of the treatment. *57*
Figure 2.9 Gene Set Enrichment Analysis highlights decrease of genes related to flagellar assembly. *58*
Figure 2.10 Species composition of three representative genes from the "Flagellar Assembly" pathway. *60*
Figure 3.1. Timeline of one-week treatment with GSK864 experiments. *67*
Figure 3.2. Diagram of metabolic flux samples collected *78*
Figure 3.3 Timeline of Metabolic flux experiment. *79*
Figure 4.1.Current model of the impact of IDH1 mutation on chromatin remodelling. *87*
Figure 4.2. Diagram of main metabolic pathways potentially affected by IDH1 mutation. *90*
Figure 4.3 Glucose and Glutamine uptake unchanged in WT vs R132H IDH1 cells. *93*
Figure 4.4 Treatment with GSK864 reduces 2HG levels in R132H cells. *95*
Figure 4.5 Glutamine labelling patterns of 2HG. *96*
Figure 4.6 Intermediates of the Methionine cycle show no change between WT or R132H cells. *98*
Figure 4.7 SAM labelling patterns reveal no differences between WT or R132H cells. *100*
Figure 4.8 R132H IDH1 induces increase in metabolites other than 2HG. *101*
Figure 4.9 Labelling patterns of Acetyl-CoA suggest increased Glutamine-labelling of acetyl group in R132H cells. *102*
Figure 4.10 Glutamine labelling patterns of WT and R132H cells reveal Glutamine dependency and activation of reductive carboxylation in R132H cells. *105*
Figure 4.11 Glucose labelling patterns remain largely unchanged between WT and R132H cells. *106*
Figure 4.12 R132H IDH1 cells show increased 2HG levels and Glutamine flux, together with broad metabolic changes. *107*
Figure 5.1. Schematic classification of gliomas. *110*
Figure 5.2. Principal component analysis of RNASeq data distribution before and after normalization. *115*
Figure 5.3. Expression levels of 19 genes change after treatment with GSK864 in R132H cells. *118*
Figure 5.4. Gene set enrichment analysis suggests increase in immune response, folate metabolism and decreased Insulin receptor mediated signaling in R132H cells. *119*
Figure 5.5 Gene expression levels of core genes from immune response- or Insulin signaling-related pathways. *120*

Figure 5.6 Gene set enrichment analysis suggests increased Folate metabolism in R132H cells. *121*
Figure 5.7 Methylation related genes show distinct pattern in IDH1 mutant cells. *122*
Figure 5.8 PCA showing data distribution before or after normalization in the 1 week dataset. *124*
Figure 5.9. Common genes between Genotype (R132H Ctrl vs WT Ctrl) and R132H treatment (R132H GSK vs R132H Ctrl) *126*
Figure 5.10 Top 10 Gene Ontology categories enriched in genes changing upon GSK864 treatment in R132H cells. *128*
Figure 5.11. Example GO categories enriched in R132H cells treated with GSK864. ..... *130*
Figure 5.12 Top 20 gene sets enriched or decreased in one week dataset. *131*
Figure 5.13 Expression levels of core genes from the "Fatty Acyl-CoA biosynthesis" set. ... *132*
Figure 5.14 Expression levels of core genes from the "SUMOylation of DNA damage response and repair" set. *134*
Figure 5.15 Methylation-related differentially expressed genes in the one-week dataset. *136*
Figure 6.1 mRNA expression levels of histone demethylases in R132H control cells. *145*
Figure 6.2 Genome-wide increase of H3K9me3 in R132H cells. *146*
Figure 6.3 Distribution of H3K9me3 differential peaks highlights subset of loci with decreased methylation. *147*
Figure 6.4 Genome-wide loss of CTCF in R132H cells. *149*
Figure 6.5. Annotation of CTCF loci showing a significant decrease in R132H cells. *150*
Figure 6.6 No clear link between loss of CTCF and nearby gene expression. *152*
Figure 6.7 IGV tracks of genes with loss of CTCF binding but opposite expression levels. . *153*
Figure 6.8 Absence of direct link between lost CTCF binding and gene expression is confirmed in one week dataset analysis. *154*
Figure 6.9 IGV tracks of selected genes from GREAT analysis on one week dataset. *156*
Figure 6.10 . Effect of GSK864 treatment on CTCF or H3K9me3 levels. *158*
Figure 7.1 Rearrangement of H3K27ac in R132H cells. *166*
Figure 7.2 Analysis of H3K27ac levels around promoters and corresponding gene expression levels. *167*
Figure 7.3 IGV tracks showing the ENO2 and ALDH1L1 loci. *169*
Figure 7.4 Annotation with EnhancerAtlas confirms uncoupling between enhancer acetylation and target gene expression. *182*
Figure 8.1. Changes in insulation of TADs as a consequence of IDH1 mutation. *187*
Figure 8.2. SOX2 loses contact with its enhancer following introduction of three genetic lesions (three-hit model): a gain of function IDH1 mutation combined with loss of ATRX and p53. ..... *191*
Figure 8.3. CTCF binding around Flavahan-annotated boundaries. *198*
Figure 8.4. Distribution of CTCF around the F11R locus. *199*
Figure 8.5 CTCF distribution around the GSTM locus. *201*
Figure 8.6. Volcano plot of differentially expressed genes shows commonalities with the Flavahan dataset. *203*
Figure 8.7. Validation of PDGFRA and DNAJC12 expression. *207*
Figure 8.8. Gene expression levels of Flavahan related genes in 1 week dataset. *211*
Figure 8.9. Description of PDGFRA locus. *212*
Figure 8.10. TADs distribution around the PDGFRA locus (hg19). *214*
Figure 8.11. CTCF ChIPSeq shows no statistically significant loss of CTCF around PDGFRA Insulator. *216*
Figure 8.12. Distribution of H3K27ac and H3K9me3 around the PDGFRA locus. *217*
Figure 8.13. DNAJC12 locus description. *219*
Figure 8.14. Lack of CTCF peak around annotated boundary for DNAJC12 locus. *221*

Figure 8.15. Distribution of TADs around the PTPRN locus (hg19)......*223*
Figure 8.16 No clear CTCF binding site at PTPRN predicted boundary......*225*
Figure 8.17. SOX2 locus shows cell type-dependent TAD boundaries......*227*
Figure 8.18. CTCF binding remains unchanged between WT or R132H U87 cells at the SOX2 locus.....*229*
Figure 9.1. Schematic representation of nuclear and chromosome organization. *237*
Figure 9.2 Some chromosomes show loss of H3K9me3 in R132H cells. *240*
Figure 9.3 GO enrichment analysis on H3K9me3 differential peaks in the one week dataset. *241*
Figure 9.4 H3K9me3 distribution on Chromosome X in WT or R132H IDH1 U87 cells......*242*
Figure 9.5 Distribution of H3K9me3 peaks lost across three autosomal chromosomes. *244*
Figure 9.6 Annotation of decreased H3K9me3 peaks along chromosome X......*245*
Figure 9.7 Optimization of H3K9me3, H3K27ac and H4K5ac immunofluorescence in WT or R132H U87 cells......*249*
Figure 11.1 Sibutramine induces weight loss in obese rats. .....***Error! Bookmark not defined.***
Figure 11.2 Analysis of H3K27ac changes across enhancers in the 48 hours reveals potential uncoupling with gene expression......***Error! Bookmark not defined.***
Figure 11.3 IGV tracks of SYNPO and MAP2. .....***Error! Bookmark not defined.***
Figure 11.4 . Analysis of H3K27ac changes across enhancers in the one week dataset..***Error! Bookmark not defined.***
Figure 11.5 . IGV tracks of the NR0B1 and SPINK1 loci. .....***Error! Bookmark not defined.***
Figure 11.6 . Gene expression changes fall preferably between, rather than within, H3K9me3 peaks lost. .....***Error! Bookmark not defined.***
Figure 11.7 Gene expression changes around BGW gene cluster...... ***Error! Bookmark not defined.***
Figure 11.8. Tukey's post-hoc test following ANOVA statistical test on Timepoint reveals significance mostly due to fasting on Day 36......*260*
Figure 11.9. Classification of higher order chromatin structures. ..... ***Error! Bookmark not defined.***

1. Introduction

1.1 Epigenetic modifications on chromatin and DNA fine tune gene expression

Chromatin indicates the complex of DNA and proteins, called histones, which bind to it in order to compact it into the nucleus (Kouzarides, 2007). Its smallest unit is the nucleosome, composed of four core histones, H3, H4, H2A and H2B, joint in an octamer around which 147 bp of DNA are wrapped (Kouzarides, 2007). Histones present a globular domain and a tail, and both can be modified on certain amino acids with a wide range of chemical groups, although lysine and arginine methylation, lysine acetylation, serine or threonine phosphorylation and lysine ubiquitination are so far the best characterized (Berger, 2007). This thesis will focus primarily on lysine methylation and acetylation.

Histone methylation is the reversible addition of a mono-, di-, or trimethyl group on specific lysine or arginine residues (B. Li, Carey, & Workman, 2007). This modification is promoted by methyltransferases, and its removal is mediated by demethylases (Black, Van Rechem, & Whetstine, 2012; Gibney & Nolan, 2010). Depending on the residue, histone methylation can be associated with different transcriptional states: residues such as H3K9me3 and H3K27me3 can be associated to transcriptional repression, while others are linked to active gene expression, such as H3K4me1 or me3, present at active gene promoters, and H3K36me3, located in transcriptionally active gene bodies (Berger, 2007). On the contrary, upon acetylation, an acetyl group is added by acetyltransferases to the tail of selected histones, mainly H3 and H4, and removed by histone deacetylases (Haberland, Montgomery, & Olson, 2009; Utley et al., 1998). This modification is thought to be more dynamic, and correlates with transcriptional activation (T. H. Kim et al., 2005; Rivera & Ren, 2013).

Both modifications can impact on gene expression in two ways, either by promoting the recruitment of non-histone proteins on a specific site, or by tightening or loosening histones'

grip on DNA, thus favouring or hindering gene expression (Shogren-Knaak et al., 2006; Wysocka et al., 2006).

On a higher order architecture level, the presence of certain histone modifications separates the genome into two functionally distinct sections: heterochromatin and euchromatin. Heterochromatin is a transcriptionally silenced compartment, characterized by low levels of histone acetylation and high histone methylation at the H3K9, H3K27 and H4K20 residues (Kouzarides, 2007). In particular, H3K9me3 is thought to mark constitutive heterochromatin, tightly packed nucleosomes around DNA sequences that must be kept perpetually silenced, such as telomeres, centromeres or transposable elements (Mikkelsen et al., 2007; Ninova, Tóth, & Aravin, 2019; C. Wang et al., 2018). Indeed, this latter class is composed of endogenous retroviruses (ERVs), transposons and long- or short interspersed nuclear elements (LINEs or SINEs), whose activation might disrupt genomic integrity by retrotransposition in different areas of the genome (J. He et al., 2019; Payer & Burns, 2019). H3K27me3 lines facultative heterochromatin in a cell-type specific manner, silencing developmental genes whose expression is no longer needed in terminally differentiated cells (E. Bernstein et al., 2006). Euchromatin, on the other hand, is a dynamic compartment that upon reception of stimuli, can favour gene expression: here, transcriptionally active sequences usually correlate with both acetylated and methylated histone residues, such as H3K27ac, H3K4me1, H3K36me3 and H3K79me2 (Berger, 2007; B. Li et al., 2007). Each of these marks shows a distinct distribution pattern and correlates with a specific transcriptional status, summarised in Table 1.

Table 1. Positioning of best characterized histone modifications along genes and relative correlation with the transcriptional state. *Summary of the main histone lysine acetylations and methylations. These marks have been extensively studied and correlate to different transcriptional states, from activation (see H3K27ac, H3K4me3, H3K36me3), to repression (H3K9me3), and a so-called bivalent state, commonly found in stem cells, which primes promoters with both H3K4me3 and H3K27me3, easing their transcriptional activation following specific signals* (B. E. Bernstein et al., 2006).

Histone Modification	Localization	Correlation with Transcription	References

H3K27ac	Promoter	Activation	(Creyghton et al., 2010a; T. H. Kim et al., 2005)
H3K27ac	Enhancer	Activation	(Creyghton et al., 2010a)
H3K4me3	Promoter	Activation	(B. E. Bernstein et al., 2005; T. H. Kim et al., 2005)
H3K36me3	Gene body	Activation	(Barski et al., 2007)
H3K9me	Highly correlated to DNA methylation and associated with transcriptional repression (H3K9me3 and me2). Barski et al find H3K9me1 enriched at gene promoters and suggest its correlation with transcriptional activation.		Messner 2008, Bernstein et al 2006a, Barski et al, 2007 (Barski et al., 2007; B. E. Bernstein et al., 2006)
H3K4me1 + H3K27ac	Active Enhancer	Activation	(Heintzman et al., 2009; Rada-Iglesias et al., 2011)
H3K4me1 + H3K27me3	Promoter	Poised/Bivalent	(B. E. Bernstein et al., 2006)

Another critical epigenetic modification is DNA methylation, upon which a methyl group is added to Cytosine bases, forming 5-methylcytosine (5mC) (Jones, 2012). Methylation on DNA is carried out by three DNA methyltransferases: DNMT1, and the *de novo* methyltransferases DNMT3A and 3B (Okano, Bell, Haber, & Li, 1999). The modification is reversible and its removal can be either through passive dilution following multiple rounds of cell division, or active, thanks to DNA demethylases such as the Ten Eleven Translocation methylcytosine dioxygenases (TET) (Kagiwada, Kurimoto, Hirota, Yamaji, & Saitou, 2013; Wu & Zhang, 2017). CpG Islands are short stretches (~ 1 kb) of CpG nucleotides found in more than half the genes in the genome (Luo, Hajkova, & Ecker, 2018). When localised at gene promoters, most CpG Islands usually remain unmethylated in a normal cell, with the exception of the inactivated X chromosome, imprinted genes, and a subset of tissue- or germline-specific genes, whose DNA methylation around the TSS promotes their long-term silencing (Farthing et al., 2008;

Jones, 2012; Lock, Takagi, & Martin, 1987). This situation is often overturned in cancer development, so much so that aberrant DNA methylation is considered an hallmark of cancer (Bartlett et al., 2013; Jones & Baylin, 2002). In this context, DNA methylation occurs around the promoter of onco-suppressor genes, further contributing to cancer progression (Jones & Baylin, 2002). In contrast, genome-wide methylome profiling has revealed the presence of distinct, tissue- and cell type-specific methylation patterns in non-CpG Island sequences, called differentially methylated regions (DMRs) (M. D. Schultz et al., 2015; Ziller et al., 2013). Such regions overlap with tissue-specific enhancers, and are characterized by loss of 5mC during development, followed by a gain of DNA methylation as cell differentiation progresses (Y. He et al., 2020).

Thus, DNA methylation, as well as post-translational modification on histones, add another layer of regulation, and their concerted action results in the fine tuning of gene expression. Research in recent years has highlighted the connection between chromatin and metabolism. On one hand, metabolism provides the co-factors for many post-translational modifications that regulate the tightness with which histones are wrapped around DNA, favouring or inhibiting transcription (Dai, Ramesh, & Locasale, 2020). On the other, such mechanisms further regulate the expression of genes coding for enzymes involved in biochemical pathways, amongst others (Kinnaird, Zhao, Wellen, & Michelakis, 2016). Of these modifications, the effect of histone acetylation and methylation on transcription have been extensively studied, and are representatives of the relationship between chromatin and metabolism.

1.2 AcetylCoA is the primary substrate for histone acetylation

Histone acetylation is a chromatin modification generally associated with transcriptional activation (Creyghton et al., 2010b; Shogren-Knaak et al., 2006). Upon acetylation on specific lysine residues on histone tails, the positively charged lysine is neutralized by the addiction of an acetyl- moiety, loosening histones' grip on the DNA and relaxing the chromatin, allowing for the binding of the transcriptional machinery (Shogren-Knaak et al., 2006). Indeed, H3K27

acetylation (H3K27ac) is generally found at gene promoters and marks active enhancers (Creyghton et al., 2010a). This reaction is mediated by lysine acetyltransferases (KATs) and is reversed by the activity of lysine deacetylases (KDACs) (Haberland et al., 2009; Utley et al., 1998). The primary source for histone acetylation is AcetylCoA, a key metabolite involved in multiple biosynthetic pathways (Takahashi, McCaffery, Irizarry, & Boeke, 2006). It can be found in the mitochondria, as it is a tricarboxylic cycle (TCA) intermediate, as well as in the cytosol and the nucleus. In the cytosol, Citrate, Acetate and Malonyl-CoA can all be converted into Acetyl-CoA by ATP-Citrate Lyase (ACLY), AcetylCoA Synthetase (AceCS1) or MalonylCoA decarboxylase (MCD), respectively (Luong, Hannah, Brown, & Goldstein, 2000; Sacksteder, Morrell, Wanders, Matalon, & Gould, 1999; Wellen et al., 2009b). Both Citrate and Acetate can freely diffuse into the nucleus, where local ACLY and AceCS1 transform them into nuclear AcetylCoA (Ariyannur et al., 2010). More recent studies confirmed the presence of nuclear Pyruvate, which can further contribute to nuclear AcetylCoA concentration following conversion by pyruvate dehydrogenase (Sutendra et al., 2014b). Interestingly, lysine deacetylases are inhibited by metabolites like the short chain fatty acid butyrate or beta-hydroxybutyrate, generated by fatty acid oxidation, ketogenesis or indirectly produced by the microbiota (Shimazu et al., 2013).

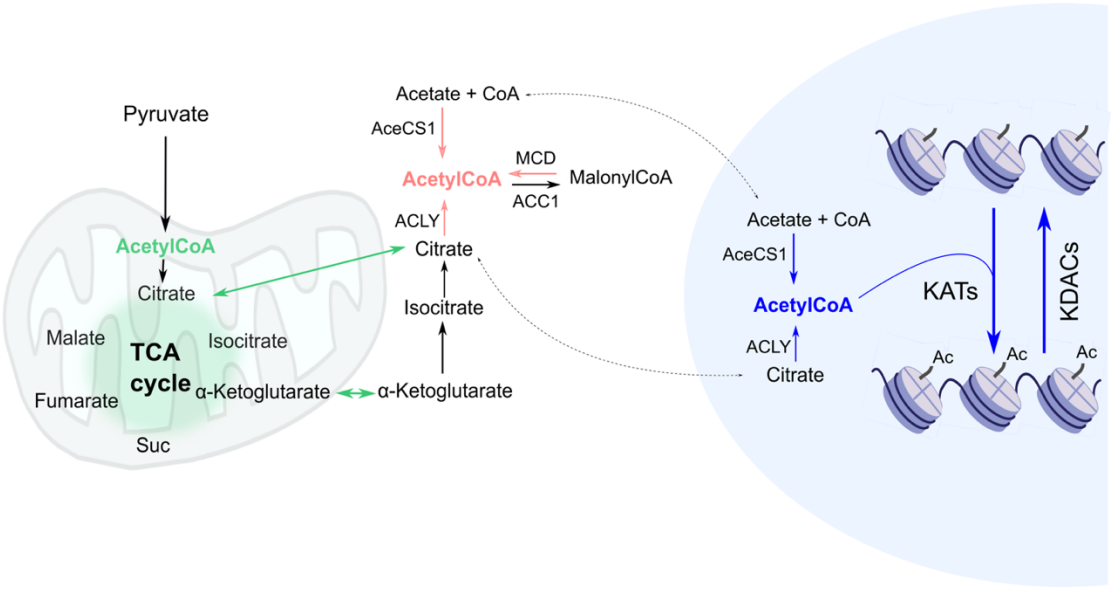


Figure 1.1. Three distinct pools contribute to AcetylCoA cellular concentrations*. Diagram of the mitochondrial (green), cytosolic (pink) and nuclear (blue) AcetylCoA pools and the metabolic pathways involved in its synthesis. While mitochondrial AcetylCoA is mainly used to fuel the TCA cycle, cytosolic precursors Citrate and Acetate can freely diffuse into the nucleus to locally produce AcetylCoA, which will be used by Lysine acetyltransferases (KATs) to acetylate lysine residues on histones.*

Thus, levels of AcetylCoA, and more in general, the energy state of the cell directly impact on chromatin conformation. The reverse is also true: deacetylase SIRT6 is a co-repressor of HIF1a expression, and helps maintaining correct blood glucose concentration and mitochondrial respiration (Zhong et al., 2010). Furthermore, it has been shown to recruit CLOCK on the promoter of genes coding for enzymes involved in fatty acid metabolism in the liver (Masri et al., 2014). Nuclear acetyltransferase KAT2A acetylates Peroxisome proliferator-activated receptor gamma coactivator 1-alpha (PGC1-a), with the downstream effect of reducing gluconeogenesis (Lerin et al., 2006).

Thus, histone acetylation provides an example of chromatin modification largely dependent on metabolic pathways.

1.3 DNA and Histone methylation is promoted by one-carbon metabolism

Methylation is the acquisition of a methyl (-CH_3) group on Cytosine bases on the DNA, or Lysine or Arginine residues in histones (Barski et al., 2007). On DNA, methylation of CpG residues is generally a mechanism of gene expression inhibition performed by DNA

methyltransferases and enriched around repetitive elements, heterochromatin and inactive gene promoters, while CpG islands in active promoters are characteristically unmethylated (Schübeler, 2015). On histones, methyltransferases can add up to three methyl moieties, but its impact on gene expression varies depending on the target lysine or arginine (B. Li et al., 2007). Only lysine methylation is discussed here. H3K4me3 and H3K36me3 are concomitant to, but not necessarily causing, gene expression, marking active promoters or actively transcribed regions, respectively (Howe, Fischl, Murray, & Mellor, 2017; Kolasinska-Zwierz et al., 2009). Instead, methylation of lysine 27 or 9 on histone 3 is a marker of heterochromatin, with H3K27me3 more commonly found in facultative heterochromatin, around genes necessarily active during development or differentiation, then silenced, and H3K9me3 enriched in constitutive heterochromatin, silencing genomic regions such as repetitive DNA elements to maintain genomic integrity (Hawkins et al., 2010; B. Wen, Wu, Shinkai, Irizarry, & Feinberg, 2009; Zhu et al., 2013).

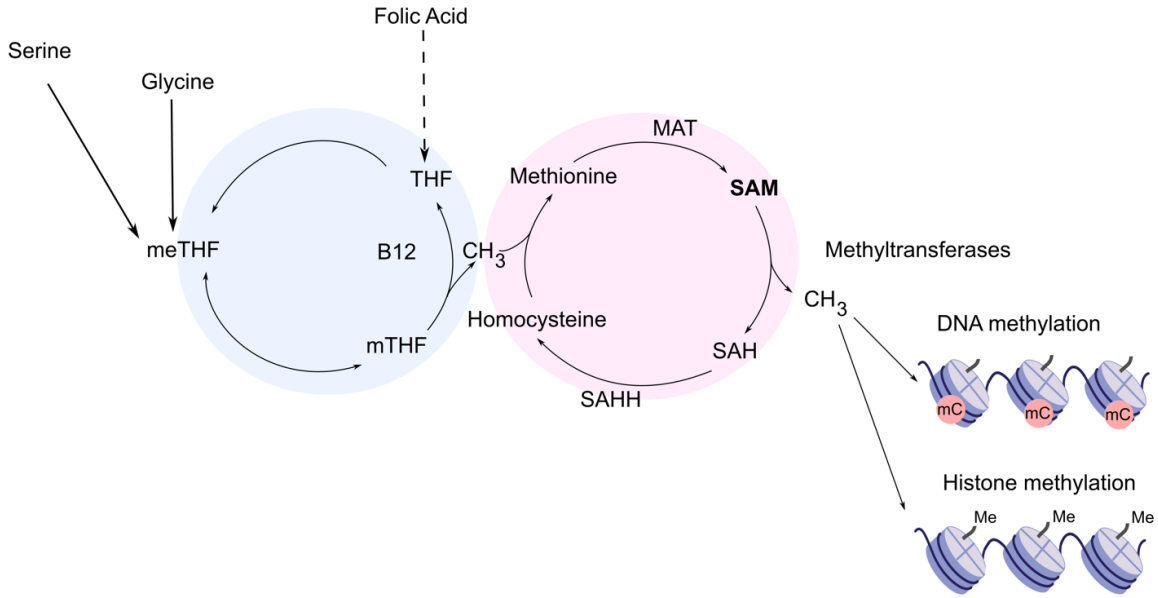


Figure 1.2. Folate and Methionine cycles participate to SAM production *Diagram representing the main steps of folate cycle, fueled by Serine, Glycine and the diet-derived Folic acid (blue circle), and its connection to the Methionine cycle (pink circle). Briefly, the addiction of a methyl group from 5 methyl-tetrahydrofolate (mTHF) to Homocysteine forms Methionine, which is further converted into SAM by Methionine adenosyl-transferase (MAT). Upon methylation, SAM is used as a substrate by methyltransferases, and loses its methyl group, becoming SAH. This intermediate is then reconverted into Homocysteine by S-adenosyl homocysteine hydrolase (SAHH).*

The main methyl donor in the cell is S-Adenosyl-Methionine (SAM) (Sanderson, Gao, Dai, & Locasale, 2019). Upon demethylation to S-Adenosyl-Homocysteine (SAH), SAM loses a methyl group which is then transferred by methyltransferases upon their target. SAH is then reconverted into Homocysteine, refuelling the Methionine cycle (Sanderson et al., 2019). Thus, the ratio between SAM and SAH regulates the availability of methyl groups in the cell, and consequently the rates of DNA or protein methylation, linking once again the metabolic to the epigenetic state of the cell (Shyh-Chang et al., 2013). SAM production is closely connected to systemic metabolism and nutrition, as it is generated from the Folate and Methionine cycles, which are in turn fuelled by levels of Folic Acid, acquired through the diet or produced by commensal bacteria, as well as Glycine and Serine (Maddocks, Labuschagne, Adams, & Vousden, 2016; Shyh-Chang et al., 2013).

Thus, methylation levels are linked to cellular metabolism and, due to its dependence on diet-derived intermediates such as folate and methionine, systemic metabolism.

1.4 Coordination between epigenetics and metabolism

The coordination between metabolic inputs and epigenetic modifications is summarised by three different models.

The first model describes the production of metabolites that act as co-factors for, activate or inhibit the enzymes inducing chromatin modifications. Some histone demethylases (Jumonji domain-containing histone-lysine demethylases, Jmj-KDMs) and Ten Eleven Translocation methylcytosine dioxygenases DNA demethylases (TET) belong to the α-ketoglutarate dependent dioxygenases protein family (X. Li, Egervari, Wang, Berger, & Lu, 2018). As such, their activity depends on the availability of alpha-ketoglutarate, but increased concentrations of succinate, fumarate and of the oncometabolite 2 hydroxyglutarate (2HG) have been shown to inhibit their function through competitive binding, resulting in an overall increase of both DNA and histone methylation, which contributes to cancer development (Lu et al., 2012; Turcan et al., 2018; Xiao et al., 2012). Furthermore, SAM and AcetylCoA are used for generation of methyl and acetyl groups by methyl or acetyltransferases, respectively, whose activity is inhibited by higher concentrations of their side products, SAH and CoA, adding a layer of metabolic regulation to epigenetic modifications in the cell (J. V. Lee et al., 2014; Locasale, 2013). Depending on their affinity with the cofactor (K_m) or the inhibitor (K_d), chromatin modifying enzymes activity will be more or less immediately affected by the changes in concentration of their metabolic cofactors, leading to differences in the sensitivities of epigenetic modifications in response to metabolic changes (Dai et al., 2020). Depletion of SAM has been shown to cause loss of histone di- and tri- methylation, in favour of monomethylation (Haws et al., 2020). Members of the large family of α-ketoglutarate-dependent lysine demethylases (KDMs) have been shown to have different K_d for the oncometabolite 2HG, leading to differences in the extent of their inhibition (Black et al., 2012). Since each KDM has a preferential target in terms of histone modification, and is more or less inhibited by 2HG

based on its K_d, there is a combinatorial effect that results in a generalised raise in histone methylation, with deleterious consequences (Thinnes et al., 2014).

A second example of coordination between metabolism and chromatin modification is nutrient sensing. As described in the previous sections, methylation relies on the availability SAM. Its production is based on the essential amino acid methionine and diet-derived folate cycles (Sanderson et al., 2019). Indeed, SAM and SAH plasma levels seem to correlate with diet, potentially linking methylation levels with nutrition (Poirier, Wise, Delongchamp, & Sinha, 2001; Van Driel et al., 2009). This might also have an impact on tumor development, as a methyl donor deficient diet was reported to have a protective effect on colon cancer formation in predisposed animals (Kadaveru, Protiva, Greenspan, Kim, & Rosenberg, 2012). AcetylCoA, the main source for histone acetylation, is also produced through two major metabolic pathways: glycolysis and fatty acid oxidation (Kinnaird et al., 2016). Its levels are dynamic and its higher concentration is associated with cellular growth and proliferation (Cai, Sutter, Li, & Tu, 2011). Thus, cells might sense AcetylCoA levels to couple energy availability with proliferation, by using AcetylCoA to acetylate histones, favouring the activation of genes whose products further promote cell growth. Indeed, studies in yeast and mammals have shown that deletion of the cytosolic or nuclear enzyme AcetylCoA-synthesizing enzyme *ACS2* (*S.cerevisiae*) or the mammalian ACLY result in a drop of histone acetylation levels (Takahashi et al., 2006; Wellen et al., 2009a).

Finally, there is growing evidence of the presence of metabolic enzymes in the nucleus, producing the chromatin-modifying metabolites needed by epigenetic writers to locally impact the epigenome. Enzymes involved in production of AcetylCoA, such as pyruvate dehydrogenase (PDH), AcetylCoA synthetase 2 (ACSS2) and ACLY, have been shown to localise in the nucleus, where their activity is linked to histone acetylation (Bulusu et al., 2017; Sutendra et al., 2014a; Wellen et al., 2009a). Furthermore, the folate pathway enzyme MTHFD1

has recently been found in the nucleus, where it maintains a pool of folate metabolism intermediates and influences gene expression (Sdelci et al., 2019). Thus, a set of nuclear metabolic enzymes actively interacts with chromatin modifiers.

While these mechanisms are generally accepted, the exact molecular players involved in the coordination between metabolic substrate production and subsequent chromatin modifications are still unclear, making research in the field even more relevant, considering the potential implications that any alterations to this equilibrium can have in disease development.

1.5 The metabolism-epigenetics axis in disease development

Given the tight coordination between metabolism and epigenetics, it is easy to assume that alterations in the former will impact the latter, and vice-versa.

Such events can occur in the context of cancer, where the connection between chromatin and metabolism contributes to disease progression in different ways. Mutations in genes coding for metabolic enzymes can alter the epigenome, as exemplified by Isocitrate Dehydrogenase I (IDH1), Succinate Dehydrogenase (SDH) and Fumarate Hydratase (FH) (Lu et al., 2012; Turcan et al., 2018; Xiao et al., 2012). All three genes can be mutated in different types of cancer, and the resulting higher levels of 2HG, succinate or fumarate inhibit α-ketoglutarate-dependent dioxygenases, including DNA and histone demethylases, with a resulting increase in DNA and histone methylation (Lu et al., 2012; Turcan et al., 2012; Xiao et al., 2012). Rewiring of cancer cell metabolism can impact on the epigenome without the presence of genetic mutations: reduced oxygen levels caused by tumour hypoxia leads to a reduction in the activity of the of TET proteins, leading to DNA hypermethylation, in a wide range of human and murine cell lines (Thienpont et al., 2016). Finally, aberrant activation of oncogenes, such as c-MYC or KRAS has been shown to alter cell metabolism by inducing glutamine dependence (Gaglio et al., 2011; P. Gao et al., 2009). This metabolic rewiring can be promoted directly, through mRNA and protein upregulation of glutamine transporters, as well as Glutaminase

(GLS), or as a downstream effect of other oncogene-activated signalling cascades, such as mTOR (Csibi et al., 2014; Son et al., 2013).

Much less is known about the role of the metabolism-epigenetics axis in the context of obesity. Emerging evidence suggests an involvement of gut microbiota-derived metabolites on the host epigenome and phenotype (Sampson et al., 2016; Thaiss et al., 2016; Turnbaugh, Bäckhed, Fulton, & Gordon, 2008; Turnbaugh et al., 2006). The most characterized example is provided by short-chain fatty acids (SCFAs; e.g. butyrate, crotonate; propionate etc), which can influence the host protein acetylation levels, either directly, as butyrate has been shown to inhibit lysine deacetylases, and indirectly, by being processed into units of AcetylCoA, thus providing the acetyl group for histone acetylases (Krautkramer et al., 2016). It has been proposed that energy harvest through SCFAs production promoted by gut microbes can participate to obesity development, although more research is needed to clarify whether this effect is due, at least in part, by a raise in acetylation on the host cells (Turnbaugh et al., 2008, 2006).

Research in the field of cancer metabolism and epigenetics has highlighted multiple ways in which alterations to this axis can impact on cancer progression, but its involvement in the development of other diseases is still unclear. Thus, a better understanding of the interplay between these two major players in cellular physiology and disease can lead to the design of better, specific and ultimately more successful therapies.

1.6 Aims

In this thesis, the nexus between chromatin and metabolism is examined, to evaluate its role in disease development at the cellular and systemic level, within two separate projects. First, the effects of a panel of weight-loss drugs on the composition of the gut microbiome of obese rats maintained on a high fat diet is examined using shotgun metagenomics sequencing. To support this, data on the physiological state of rats will help dissect whether changes to the microbiome impact on obesity and whether these are potentially linked to changes to the host metabolism.

The second project exploits a genetically induced change in metabolism as a result of a gain of function mutation in isocitrate dehydrogenase (IDH1 R132H) that creates an oncometabolite D-2-hydroxyglutarate. Here, using glioma-derived cells (U87) with or without the mutation, together with an inhibitor (GSK864) of mutated IDH1 enzyme, the potential impact of D-2-HG on the metabolism and epigenetics of cell will be investigated with the aim of understanding more about glioma development and progression. Due to the profound differences between the two projects, more detailed aims and introductory information are provided at the start of each results chapter.

Metagenomic analysis of microbiomes in drug-treated rats

2. Metagenomic analysis of microbiomes in drug-treated rats

2.1 Introduction

2.1.1 Role of the gut microbiome in human metabolism

Nutrients introduced with our diet are broken down into complex carbohydrates within the stomach. The human genome, however, does not include genes that code for all the enzymes that would allow the breakdown of complex carbohydrates into smaller molecules. These are further digested in the gut thanks to its colonizing bacteria, and about 3.3 x 10^6 non-redundant genes from over 1200 different species make up for the gut microbiome (Postler & Ghosh, 2017). Recent advances in the field have highlighted its important contribution to human metabolism: indeed, it is estimated that up to 10% of blood-circulating metabolites are products of reactions carried out by the microbiota (Wikoff et al., 2009).

An important class of metabolites produced are short-chain fatty acids (SCFAs), carbohydrates with as little as two (acetate), three (propionate) or four (butyrate) carbons. SCFAs can act as signalling molecules, sending messages to the Autonomous Nervous System, stimulating the release of peptide YY and the synthesis of Leptin, the satiety hormone (Samuel et al., 2008; Xiong et al., 2004). They have been shown to have a role in the regulation of the Immune System, influencing the chemotactic response of neutrophils, promoting T cells differentiation into regulatory T cells, downregulating inflammatory cytokines in dendritic cells and apoptosis in CD4+ T cells (Park et al., 2015; Postler & Ghosh, 2017). Interestingly, within human cells SCFAs also impact chromatin remodelling, either directly due to their inhibition of lysine deacetylases (KDACs) or indirectly, by providing Acetyl- groups to lysine acetylase transferases (KATs) (Donohoe et al., 2012), that target chromatin proteins including histones.

Furthermore, the microbiota participates in bile acid metabolism, converting unabsorbed bile acids into secondary bile acids, which can activate FXR signalling, regulating bile acid

production, glucose metabolism and hepatic autophagy, or TGR5, a GPCR receptor expressed in a wide range of tissues (Hylemon et al., 2009; Inagaki et al., 2006; J. M. Lee et al., 2014).

Gut-colonizing bacteria also contribute to the production of other micronutrients and metabolites, such as polyamines, indoles and vitamins, mainly K and B, with B12 being produced exclusively by the gut microbiota (Degnan, Taga, & Goodman, 2014; Rooks & Garrett, 2016; Wikoff et al., 2009).

Given its extensive role in human metabolism, it is not difficult to imagine how disruption in microbiome composition can participate in a disease state.

2.1.2 Microbiome alterations in disease states

Alterations to the microbiome have been linked to a wide variety of diseases, both metabolic, such as obesity and cancer, and psychiatric. Interestingly, the gut-brain axis has been proposed to contribute to several Central Nervous System (CNS) conditions, such as depression, anxiety and Parkinson's (Cryan & Dinan, 2012). While it is not always immediately understood whether disruption of the host microbiome is simply associated with a disease, or is playing a causative role, research in the field is trying to highlight common patterns of microbiome alterations found in individuals affected by a specific disease.

A striking example is obesity, a pandemic affecting over 600 million adults worldwide (WHO, n.d.). By definition, obese patients have a body mass index (BMI, calculated as the weight in kilograms divided by the square of the height of the patient in meters) over 30 (Longo, Heymsfield, & Wadden, 2017). At the basis of a such a high BMI there is a disbalance between caloric intake and energy expenditure, which is reflected by an accumulation of triglycerides in adipose tissue and an increase in fat mass distribution all over the body. As a consequence, patients often present other conditions, such as increased insulin secretion and insulin resistance, which often lead to the development of type 2 diabetes, high blood pressure and dyslipidemia, linked to the occurrence of stroke events and heart failure (Bornfeldt & Tabas, 2011; Hall et al.,

2010; Heymsfield & Wadden, 2017). Obesity is also difficult to treat, as most interventions, pharmaceutical, surgical or behavioural, prove to be effective only in the short term, with the vast majority of patients regaining weight over time (Table 2, Fothergill et al., 2016).

Table 2. Current pharmacotherapies for obesity. *Summary of the most common pharmacological therapies for the treatment of obesity. With the exception of Sibutramine, which has been withdrawn from the market due to its cardiological and, to a lesser extent, psychiatric adverse effects, all drugs are currently being used in the clinic, often together with a lifestyle change promoting diet and exercise* (Longo et al., 2017; Srivastava & Apovian, 2017a).

Drug	**Physiological mechanism**	**Chemical features**	**Administration**	**Adverse effects**
Phentermine	Suppression of appetite	Sympathomimetic Amine	15 or 37.5 mg orally once daily; 8 mg orally twice daily. Often used in combination with Topiramate. ER formulation available	Insomnia, dry mouth, constipation, dizziness
Orlistat	Decreased absorption of fat	Pancreatic and gastric lipase inhibitor	120 mg three times daily	Flatulence, bloating, diarrhea
Lorcaserine	Modulation of food intake and appetite	Serotonin receptor agonist (5-HT2c)	10 mg or 20 mg (ER formulation) orally once daily	Headache, dizziness, dry mouth, constipation
Liraglutide	Modulation of glucose homeostasis, food intake and satiety	GLP1 receptor agonist with incretin properties	Start with 0.6 mg subcutaneously once daily for a week, then weekly titration up to 1.2 mg, 2.4 mg, 3 mg	Nausea, vomiting, constipation, hypoglicemia, headache, dizziness, diarrhea, fatigue, abdominal pain
Bupropion		Dopamine and Norepinephrine uptake inhibitor	Upwards titration over 4	(Often used as combination treatment)

Naltrexone	Suppression of appetite	Opioid antagonist	weeks up to two tablets/day	Nausea, constipation, headache, vomiting, dizziness, insomnia, dry mouth, diarrhea
Sibutramine	Promotion of satiety	Monoamine reuptake inhibitor	Withdrawn	Cardiovascular events (stroke, heart attack), mood swings, seizures

Although faecal microbiota transplantation (FMT) experiments provide evidence of a link between obesity and an altered microbiome, only limited studies assessed whether drug treatments for obesity influence the specific composition of the microbiome. Weight loss induced by Orlistat, a reversible gastric and pancreatic lipase inhibitor, was shown to modify the host microbiome of high fat diet-fed obese mice by reducing α diversity and increasing the abundance of Rhodococcus, Pseudomonas, Acetivibrio and Roseburia genera, correlated with an increase in several pathways linked to lipid metabolism (Ke et al., 2020). However, in vitro studies on human gut model systems fed with Orlistat, olive oil, or both, found no notable effects on microbiome population dynamics, but highlighted a great variability in the lipid degradation capabilities of the gut microbiomes from different donors (Hoyles L., 2009). Further investigation led to the isolation of 378 strains with lipase activity, belonging to Firmicutes, Bacteroidetes and Actinobacteria (Hoyles L., 2009).

Although in the context of morphine dependence, treatment with the opioid receptor antagonist Naltrexone has also been found to induce changes in the microbiome, rescuing the morphine-induced gut dysbiosis and its effects on bile acid metabolism by promoting the abundance of species that contributed to increasing levels of deoxycholic acid and decreasing phosphatidylethanolamines (F. Wang et al., 2018).

Interestingly, FMT from obese mice or humans to germ-free mice increased fat deposition and induced the development of obesity-associated metabolic defects compared to the transplantation of faecal microbiome from lean individuals (Ridaura et al., 2013; Turnbaugh et al., 2008). Furthermore, the development of sequencing technologies such as 16S and shotgun metagenomics have allowed for descriptive analyses of the differences between the microbiota of obese individuals compared to lean counterparts. Despite a certain degree of discordance between the findings reported, potentially due to the extreme variability of the microbiome and differences in study design, a few common patterns have emerged. Obese patients show a less diverse microbiome, as both their α diversity, the number of different microbial species found within a subject, and their β diversity, the degree of species diversity between individuals, is decreased compared to healthy controls (Turnbaugh et al., 2009).

Another potential diagnostic feature is a shift of the ratio between Bacteroidetes and Firmicutes towards a higher presence of bacteria belonging to the latter phylum in obese individuals. However, while this has been widely described in both obese human and mice, other studies report little or no change (Ley et al., 2005; Nardelli et al., 2020; Schwiertz et al., 2010; Zhang et al., 2009). This increased proportion of Firmicutes is thought to play a key role in weight gain, as microbes from this phylum are more efficient at digesting dietary polysaccharides and crosstalk, promoting the increased storage of triglyceride within host adipocytes (Bäckhed et al., 2004; Turnbaugh et al., 2006).

Finally, an increased presence of members of the Proteobacteria phylum, such as *Desulfovibrionaceae* and *Enterobacteriaceae*, seems to be contributing to the obesity-induced inflammation state in patients. These Gram-negative bacteria add to the host microbiome a panel of genes associated to lipopolysaccharide (LPS) and peptidoglycan biosynthesis. A release of LPS into the bloodstream would elicit activation of Toll-like Receptor 4 (TLR4), thus igniting an inflammatory response (De La Serre et al., 2010; R. Liu et al., 2017).

Despite the significant progress made in the microbiome field in the past ten years, more research is needed to fully understand the role of the microbiome in obesity. A lack of guidelines for study design or bioinformatics analysis, combined with a high individual variability and our little knowledge of the microbiology behind it makes results interpretation difficult and mostly descriptive. Moreover, while many studies have focused on describing the changes in the microbiome of obese patients after surgical procedures, such as bariatric surgery or Roux-en-Y gastric bypass, when it comes to a pharmacological treatment, little is known about its impact on the microbiome (R. Liu et al., 2017; Tremaroli et al., 2015; Zhang et al., 2009).

The aims of this section are to use shotgun metagenomics to sequence the microbiome of two cohorts of obese female rats before and after a 42-day treatment with a panel of drugs associated with weight loss, immunosuppression, amelioration of alcohol, opiate or nicotine dependence, and treatment of major depressive disorders, to assess any changes induced in the microbiome. Choosing shotgun metagenomics, a technique that allows the sequencing of all the genes in the faecal sample, as opposed to amplifying only the 16s rRNAs, allowed us to identify not only the species composition for each sample, but the individual genes as well, adding information regarding the metabolic pathways potentially active in each rat. This resulted in a large metagenomics dataset that will be publicly available as a resource for the field, to be consulted as more information about the gut microbiome and the functional role of each species becomes accessible.

2.2 Materials and Methods

2.2.1 Faecal collection protocol

Faecal samples from 82 female Wistar rats were received from Chronos Therapeutics and stored at -80°C until DNA extraction was performed.

2.2.2 In vivo experimental design

The in vivo experiment was carried out by RenaSci (Nottingham, United Kingdom) on behalf of Chronos Therapeutics. Briefly, one hundred and six female Wistar rats (weight range 250-300 g) were ordered from Charles River (Margate, Kent). In order to induce obesity, animals had free access to powdered high fat diet (VRF1 plus 20% lard), ground chocolate, ground peanuts and tap water at all times. Animals were housed in pairs for 17 weeks for the induction of obesity (Part 1). Eighty seven animals were then selected on the basis of body weight and condition and housed singly in polypropylene cages with wire grid floors to enable the food intake of each rat to be recorded. After a two-week accustomization phase, animals were divided into eight treatment groups and treated as described in the table below for forty two days. The panel of weight-loss drugs used includes RDCOB001(from here onwards referred to as RDCOB), Bupropion, Naltrexone and Sibutramine, either alone or in combination. While RDCOB is a compound Chronos Therapeutics intended to repurpose as a weight-loss drug, Bupropion and Naltrexone have been approved by the FDA and are currently used in the clinic as a treatment for obesity (Heymsfield & Wadden, 2017). Sibutramine is an oral anorexiant and while still available in some countries, in many others it has been withdrawn from the market following its association with adverse effects (Srivastava & Apovian, 2017a).

Group	Treatment 1 (4 ml/kg po)	Treatment 2 (2 ml/kg ip)	n
A	Vehicle (castor oil/dehydrated alcohol po)	Vehicle (Saline ip)	12
B	RDCOB001 (0.6 mg/kg po)	Vehicle (Saline ip)	10
C	Vehicle (castor oil/dehydrated alcohol po)	Bupropion (20 mg/kg ip)	10
D	Vehicle (castor oil/dehydrated alcohol po)	Naltrexone (1 mg/kg ip)	10
E	RDCOB001 (0.6 mg/kg po)	Bupropion (20 mg/kg ip)	10
F	RDCOB001 (0.6 mg/kg po)	Naltrexone (1 mg/kg ip)	10
G	Vehicle (castor oil/dehydrated alcohol po)	Bupropion (20 mg/kg ip) plus Naltrexone (1 mg/kg ip)	10
H	Sibutramine (5 mg/kg po)	Vehicle (Saline ip)	10

Table 3. Treatment groups of the in vivo experiment. *A panel of weight-loss drugs was tested to evaluate the extent of the induced weight-loss and whether it was accompanied by any changes in the animals' microbiome.*

On the mornings of Days -2, 15 and 42 immediately prior to dosing a fresh cage pad was placed on the tray beneath the wire grid floor of the cage. The following morning (i.e. 24 hours later) approximately 15 faecal pellets were removed and placed into a sterile 50 ml tube and stored at -20°C or -80°C prior to shipment to our laboratory.

2.2.3 Oral glucose tolerance test (OGTT)

An oral glucose tolerance test (OGTT) was performed on Day 37 for each cohorts of rats (N= 5-6, total of 41 animals per OGTT) by Renasci on behalf of Chronos Therapeutics. On Day 36, rats were deprived of food overnight before the OGTT but allowed free access to water. The following day (Day 37), the 41 fasted animals were subject to the OGTT. Animals were cannulated and a first blood sample was taken (Baseline 1 timepoint (B1)). Animals were then dosed with vehicle or drug as appropriate (60 min before the glucose challenge). Four minutes before the glucose challenge, an additional blood sample was taken (Baseline 2 timepoint (B2).

Thus, baseline blood samples were taken both before vehicle/drug treatment and before the glucose challenge at 0 min. Animals were then dosed orally with D-glucose (2 g/kg) and further blood samples were taken 10, 20, 30, 45, 60 and 120 min later. Between blood-sampling animals were returned to the home cage with free access to water but not food. At all time-points, blood samples (~90 µl) were collected and plasma separated by centrifugation (2,400 g for 5 min at 4°C) to produce a single aliquot of plasma (~45 µl) which was frozen (-80°C) and subsequently assayed for glucose and insulin as described below. After the final tail vein bleed, the cannulae were removed and feeding jars were returned to the animals. All cages were then moved back to the holding room (reverse-phase lighting). On the day of the OGTT, the other cohort of rats was dosed as normal.

2.2.4 Faecal DNA extraction

Faecal DNA was extracted from samples using the QIAmp PowerFecal DNA kit (Qiagen, Hilden, Germany), following the manufacturer's instructions. DNA was quantified on a Nanodrop (Thermofisher, Massachusetts, USA), then diluted to a final concentration of 10 ng/µl. To get 200 base-pairs (bp) fragments, samples were sonicated (Bioruptor, Diagenode, Belgium) with the following settings: 70 mins, 30" on/30" off, Low followed by 20 mins, 30" on/30" off, Medium. DNA was purified using a QIAQuick PCR Clean up and Purification kit (Qiagen) and quantified using Qubit (Thermofisher). All samples were randomized prior to DNA extraction.

2.2.5 Shotgun metagenomics sequencing

Library preparation was performed with 150 ng of starting material using NEBNext Fast DNA Library Prep Set for Ion Torrent (New England Biolabs, Massachusetts, USA) and following the manufacturer's instructions. A barcode adapter (Ion Xpress Barcode Adapters 1-16, Thermofisher) was ligated to each library to allow for multiplexing (see Appendix, Table 25). DNA fragments of around 250-300 bp were size-selected on an E-Gel™ SizeSelect™ II Agarose Gels 2% (Thermofisher) as an intermediate step of the library preparation. Final

concentration of each single-end library was assessed on a High Sensitivity DNA Chip for Bioanalyser (Agilent, California, USA). Shotgun metagenomics sequencing was performed on an Ion Proton sequencer (Ion Torrent, Thermofisher). Eight to ten DNA libraries with different barcodes were pooled together at a final concentration of 250 pM. Pooled libraries were loaded on an Ion PI v3 sequencing chip using the Ion Chef System (Thermofisher). A single sequencing run was performed for each pooled library. Samples were randomized across libraries.

2.2.6 Bioinformatics Analysis

2.6.1 Preprocessing files

Sequencing fastq files were loaded on Galaxy (Afgan et al., 2018) (https://usegalaxy.org) for quality control. FASTQC (Andrews, n.d.) was used to assess sequencing quality and evaluate the presence of adapters or other contaminant sequences. Sequences were trimmed using Trimmomatic (Bolger, Lohse, & Usadel, 2014), using the SLIDINGWINDOW and MINLEN features to discard sequences below 30 bp long or whose average quality across a 4-nucleotide window drops below a defined threshold. Newly trimmed sequences were then re-run on FASTQC to assess final quality. Host-derived sequences were removed from files through alignment with Bowtie2 (Langmead & Salzberg, 2012a) against the rat canonical genome (Rn4). Unaligned sequences were written in a separate fasta file, which has been used as input for further analyses.

2.6.2 Taxonomic profiling (Oxford Centre for Microbiome Studies Pipeline)

Reads were processed using a cgat-core based pipeline, in collaboration with Dr. Nicholas Ilott from the Oxford Centre for Microbiome Studies. The pipeline (https://github.com/microbialman/MetaSequencing) assembles reads into contigs, annotates those contigs with taxonomic and functional information and produces counts tables for each category of annotation. Within this framework, reads were assembled into contigs using

megahit with default parameters (v1.1.3, Li et al., 2015), which produced 44,789 (± 29,626) contigs of greater than 200bp in length. Open reading frames were identified in contigs using prodigal (v2.6.3, Hyatt et al., 2010) with default parameters. We annotated open reading frames with functional information using eggnog mapper (emapper v2.0.0, Huerta-Cepas et al., 2017). We used the Kyoto Encyclopedia of Genes and Genomes (KEGG, Kanehisa and Goto, 2000) in downstream analyses. We annotated contigs with taxonomic information using kraken2 (Wood, Lu, & Langmead, 2019) against the kraken2-microbial-fatfree database available from https://lomanlab.github.io/mockcommunity/mc_databases.html that contains "representative" or "complete" genomes across bacteria, archaea, fungi, protozoa, virus and UniVec_Core (i.e. potential vector contaminants in genome assemblies) domains.

Reads from each sample were aligned to the corresponding assembled contigs using bowtie2 (v2.3.4.1 with default parameters, Langmead and Salzberg, 2012b) and counts tables were produced using eggNOG mapper- and kraken2-derived features (KEGG orthologous groups and taxonomy) using featureCounts (v1.6.0 with default parameters, (Liao, Smyth, & Shi, 2014)). Counts across samples were combined based on feature identifiers with a sample given a count of zero if the feature was not present. The resulting tables contained 6391 unique KEGG ko features, 54 phyla, 122 classes, 254 orders, 559 families, 1621 genera and 4850 species.

2.6.3 Microbiome composition analysis

Taxonomic profiling resulted in a table collecting microbiome composition for each sample at the species level. Diversity analyses were carried out using R/Bioconductor package Phyloseq, following a check of its suitability for shotgun metagenomic data (Supplementary Information S1, McMurdie and Holmes, 2013). Data preparation for import into a Phyloseq object was carried out following a dedicated tutorial (https://joey711.github.io/phyloseq/import-data.html), using the species counts table as OTU table, the mapping file as sample data and splitting the rownames in the Kingdom|Phylum|Class|Order|Family|Genus|Species format into a taxonomy

table. α diversity was calculated using Simpson's diversity index through the estimate_richness function within the Phyloseq package, while β diversity was assessed on relative abundances using Bray-Curtis dissimilarity index to obtain a distance matrix (vegan, R package Oksanen et al., 2016), then visualized with principal component analysis (PCA, base R). Proportion of Bacteroidetes to Firmicutes was obtained by extracting counts from species belonging to either phylum, then calculating the ratio of Bacteroidetes over Firmicutes relative abundance in each sample. Downstream analysis was prioritized based on the extent of drug treatment on weight loss and microbiome diversity. Differentially abundant genes or enriched species between Sibutramine-treated rats at Day -3 compared to Day 42 were assessed using DESeq2 (Love, Huber, & Anders, 2011), with the following linear model: ~ rat id + Day. Heatmaps representing differentially abundant genes or enriched species were produced with the Heatmap3 Bioconductor package (Zhao, Yin, Guo, Sheng, & Shyr, 2020). To perform Gene Set Enrichment Analysis, genes have been ranked by log2 fold change multiplied by the -log10(p.adjusted), then gsea was performed with the fgsea R/Bioconductor package (Sergushichev, 2016).

Analysis of species contribution to differentially abundant genes of interest were performed using a genes by species table obtained after taxonomic profiling, then absolute raw counts were transformed into relative counts. The top 10 contributing species were chosen for further representation, while the remainder was collapsed into the "other" category. KEGGREST Bioconductor package was used to decode KEGG-based KO gene identifiers into each gene name (Tenenbaum, 2019). Statistical analysis was conducted using a combination of the stats, ggpubr and tidyverse R packages (Kassambara, n.d.; Wickham et al., 2019). All results tables have been loaded on a dedicated github page at the URL: https://github.com/SilviaRaineri/Thesis

2.3 Results

2.3.1 Differential effects on weight loss after long term treatment with a panel of drugs

Two cohorts of obese female Wistar rats were treated with a panel of weight-loss drugs. Thus, we first assessed the extent of weight loss induced by each drug. To evaluate whether rats from the same treatment group had successfully lost weight during the experiment, a repeated measures ANOVA was calculated according to the formula: Weight ~ Timepoint + Rat number + Cohort (Table 4). Results showed a significant change in weight in four treatment groups (Ctrl, Bupropion, Naltrexone, Bupropion and Naltrexone and Sibutramine), in spite of the great inter-group variability (Figure 2 and Table 4, Rat number GES). To better understand whether the rats had lost, rather than gained, weight by the end of the treatment, a paired t-test comparing weights at Day -3 and weights at Day 42 of rats within the same treatment group was performed (Figure 2.1).

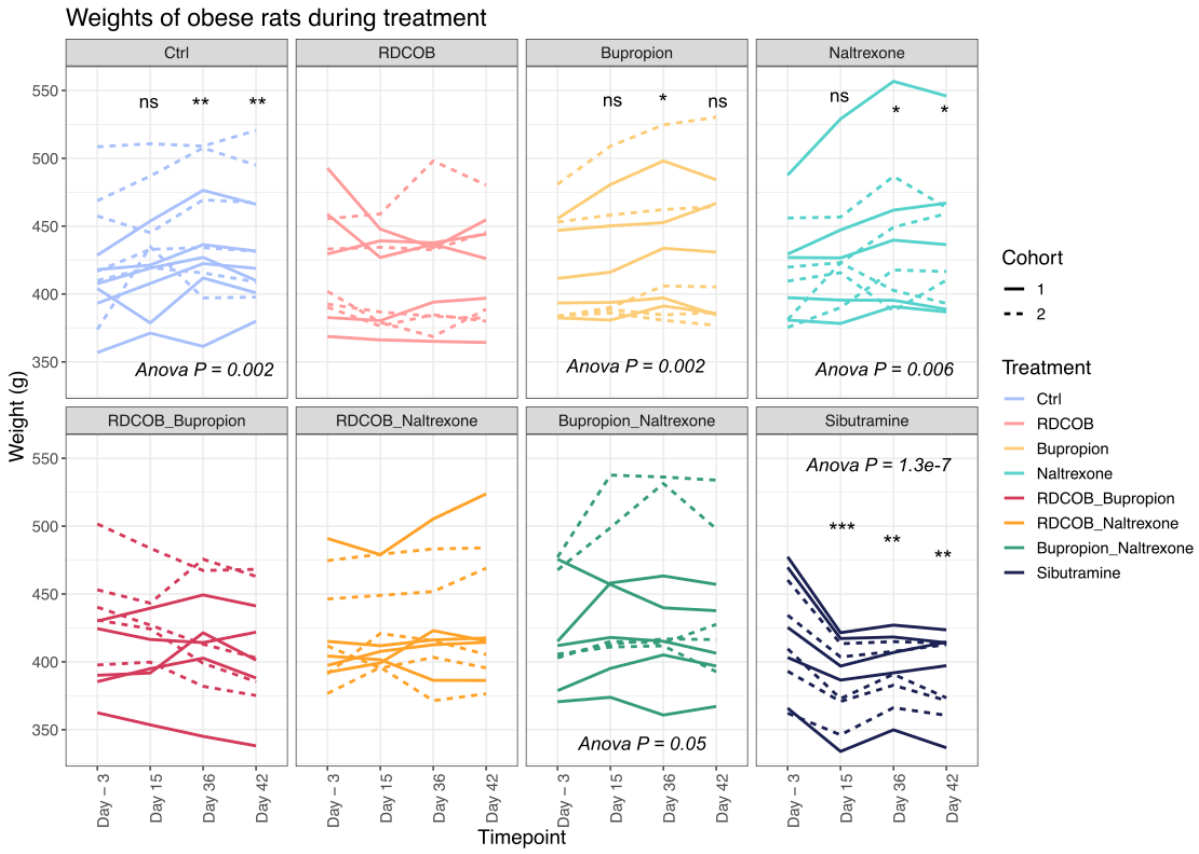


Figure 2.1. Weight of rats over the course of the treatment *Each panel shows the weights of rats treated with the respective weight loss drug at specific time points over the course of the study, starting three days before the first administration of the drug, then at Day 15, 36 and 42, the final day of the treatment. Statistical significance has been assessed first by repeated measures Anova, using the formula: weight*

*~ Timepoint + Rat number + Cohort, then significant timepoints were calculated with a series of paired t-tests, comparing Day – 3 weights with those at Day 15, 36 or 42. Paired T-test analysis within rats from the same treatment group showed only Sibutramine treatment significantly reduced weight by the end of the treatment (Day -3 vs Day 42,Benjamini-Hochberg corrected p.values: Ctrl = 0.005 (significantly higher), RDCOB = 0.52, Bupropion = 0.06, Naltrexone = 0.04 (significantly higher), RDCOB and Bupropion = 0.14, RDCOB and Naltrexone = 0.17 , Bupropion and Naltrexone = 0.22, Sibutramine = 0.005). Adjusted P-values are further detailed in Table 4. Statistical significance levels: ns P > 0.05; * P < 0.05; ** P < 0.01; *** P < 0.00 (Control: N=12; Sibutramine: N=10 rats per group).*

Only Sibutramine treatment induced a statistically significant weight loss by Day 42, causing the biggest drop in weight within the first 15 days and then maintaining the lower weight throughout the duration of the study. To a lesser extent, a similar trend is seen in RDCOB in combination with Bupropion, but rats belonging to these groups show more intra-

group variability and many regain some weight by the end of the treatment (4 rats out of 10, Table 4). Furthermore, rats belonging to the Control and Naltrexone-treated groups show a significantly higher weight by the end of the treatment, as demonstrated by the group average weight reported in Table 4.

Table 4. Average weight before (Day -3) or at the end of the experiment (Day 42). *Top: Table describing the generalized effect size (GES) and Benjamini-Hochberg-corrected P.values from the repeated measures ANOVA, calculated with the anova_test function within the rstatix package using the formula: weight ~ timepoint + rat number + cohort, within rats from the same treatment group. Results are reported for the Timepoint and Rat number terms of the formula, as weights showed a high within-group variation* (Kassambara, 2020). *Bottom: table describing the starting or final average weight for each treatment group. A third column shows P.values of the paired T-Test (weight at Day -3 vs weight at Day 42), following correction with Benjamini-Hochberg method. The corresponding signed t-statistic is reported in the last column. Paired t-test was calculated with the t.test function from baseR, using Day-3 weights as the first term (x) and Day 42 weights as second term (y). Thus, a positive sign of the t-statistic should be interpreted as the average weight at Day -3 being higher than the average weight at Day 42 (= weight loss, Sibutramine), while a negative one would reflect a lower average weight at Day-3 compared to Day 42 (= weight gain, Control and Naltrexone), as confirmed by the average weights reported in the table. Notably, Ctrl and Naltrexone groups show a significantly increased weight by the end of the treatment, while Sibutramine-treated rats significantly lost weight.*

	ANOVA: weight ~ timepoint + rat number + cohort			
	Timepoint		*Rat Number*	
Treatment Group	*GES*	*Adjusted P.value*	*GES*	*Adjusted P.value*
Ctrl	0.362	0.002	0.93	9.60E-15
RDCOB	0.38	0.3	0.92	2.09E-12
Bupropion	0.414	0.002	0.973	6.00E-18
Naltrexone	0.373	0.006	0.928	5.92E-13
RDCOB + Bupropion	0.169	0.189	0.901	2.60E-11

RDCOB + Naltrexone	0.12	0.318	0.946	2.08E-14
Bupropion + Naltrexone	0.252	0.05	0.918	2.40E-12
Sibutramine	0.731	0.00000013	0.93	5.60E-13
Treatment Group	*Average Weight (Day -3)*	*Average Weight (Day 42)*	**Paired T-test**	
			Adjusted P.value	*T-statistic*
Ctrl	420.3	435.69	0.005	-4.31 (higher)
RDCOB	420.6	416.27	0.52	0.65
Bupropion	417.45	431.56	0.06	-2.5
Naltrexone	416.38	436.76	0.04	-2.9 (higher)
RDCOB + Bupropion	421.61	408.54	0.147	1.88
RDCOB + Naltrexone	420.05	428.79	0.178	-1.64
Bupropion + Naltrexone	421.03	433.39	0.226	-1.39
Sibutramine	420.12	391.83	0.005	4.06 (lower)

Interestingly, the two drugs currently used in the clinic to treat obesity, Bupropion and Naltrexone, did not induce a significant weight loss, either alone or in combination. Thus, our panel of treatments has induced different levels of weight-loss in obese rats, each group showing high variability.

While the complete analysis has been carried out for each treatment group, given the consistent decrease in weight following Sibutramine treatment, moving forward we will mainly focus our attention on the comparison between the Control and Sibutramine groups.

2.3.2 Treatment with Sibutramine increases Insulin sensitivity

Obese patients often show other metabolic complications, such as insulin resistance, which can lead to the development of type 2 diabetes (Bornfeldt & Tabas, 2011). While no data was collected at the start of the treatment regarding the level of insulin secretion in these rats, their response to sugar was tested through an oral glucose tolerance test (OGTT) towards the end of the treatment (Day 36). During OGTT, blood samples were taken before (-60, 0) and after (10,

20, 30, 45, 60 and 120 minutes) a small dose of Glucose was given to the rats, to evaluate plasma concentration of both Glucose and Insulin over time (see section 2.2.3).

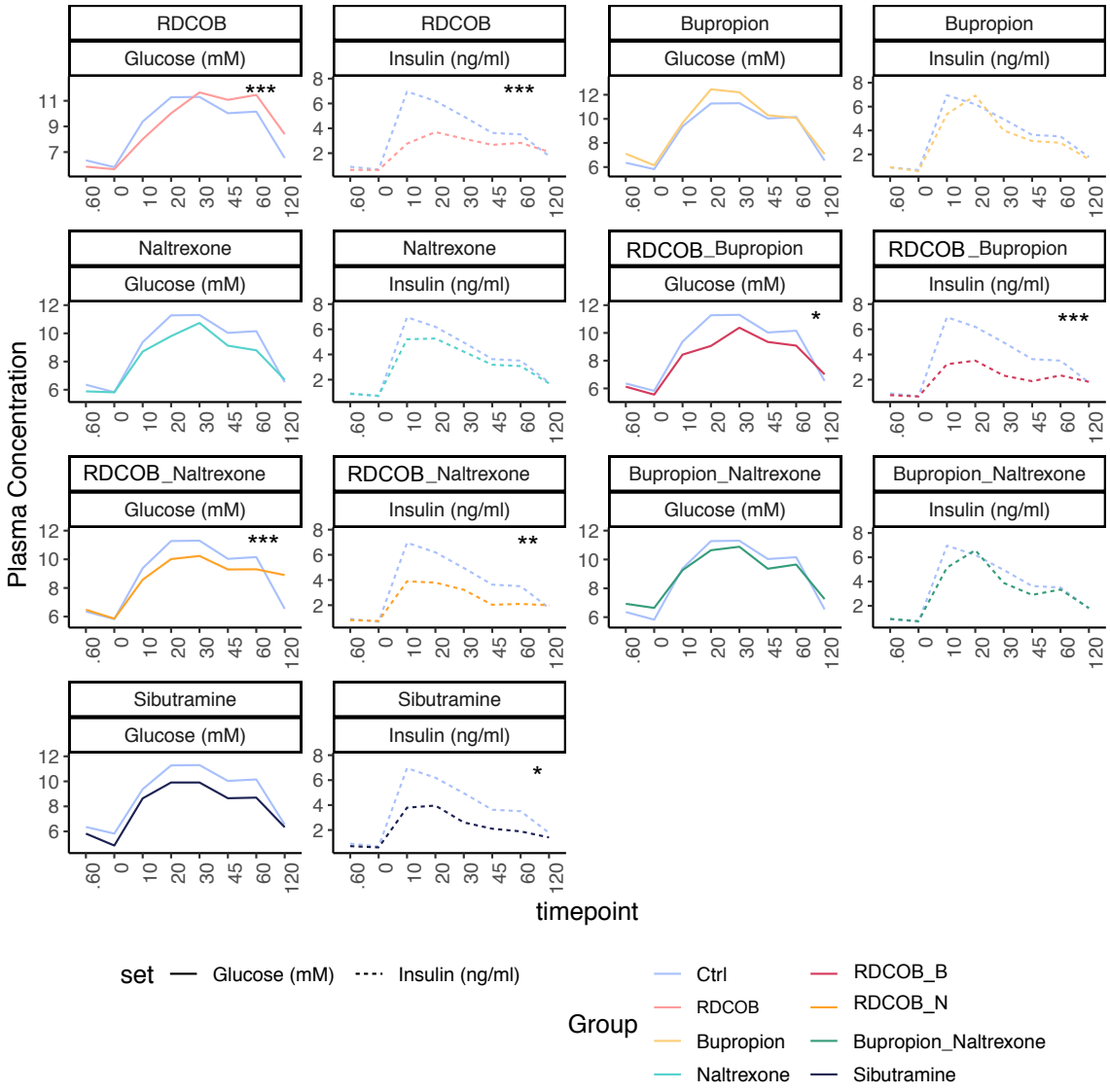


Figure 2.2. Sibutramine and RDCOB, alone or in combination, improve Insulin secretion in obese rats. *Line plot of the treatment group average values for plasmatic Glucose and Insulin throughout the OGTT. Each treatment is plotted against a violet line, representing Glucose and Insulin levels in the controls. Statistical significance was assessed by ANOVA with the formula: value ~ Group + timepoint + timepoint * Group + sample, (function anova_test within the rstatix R package, parameters: wid = sample, within = Day, between = treatment, dv = value) (Kassambara, 2020). P.values were corrected using the Benjamini-Hochberg method (Adjusted p.values: RDCOB, Glucose = 9.69e-04, Insulin = 1.98e-04; RDCOB_Buproprion, Glucose = 0.022, Insulin =4.17e-04; RDCOB_Naltrexone, Glucose = 4.46e-04, Insulin = 7.41e-03; Sibutramine, Insulin = 0.0256). Statistically significant results have been printed on the corresponding panels. Statistical significance levels = ns P > 0.05; * P < 0.05; ** P < 0.01; *** P < 0.001 (Control: N=12; Treatment Groups: N=10 rats per group)*

OGTT results suggest that treatment with RDCOB, alone or in combination with either Bupropion or Naltrexone, as well as Sibutramine, lower the concentration of plasmatic Insulin when compared to the control group (Figure 2.2, pink (RDCOB), red (RDCOB in combination

with Bupropion), yellow (RDCOB combined with Naltrexone) or blue (Sibutramine) lines compared to violet (Control)). This would potentially suggest an amelioration of Insulin sensitivity, although performing a first OGTT prior to the start of the treatment would have provided a better insight into the level of insulin resistance, glucose absorbance and insulin secretion in these rats. Nevertheless, these results suggest that Sibutramine treatment not only induces weight loss but also improves metabolic complications associated with obesity.

2.3.3 Sibutramine affects food intake

Since Sibutramine is an oral anorexiant which is thought to act through inhibition of appetite, we next asked whether rats undergoing this treatment showed a decrease in food intake. Over the course of the experiment, rats were allowed to eat *ad libitum* from three different pots containing chocolate, peanuts or powdered High Fat Diet (HFD). Food intake was calculated as the difference in weight of the chocolate, peanuts and HFD pots in the rats' cages between a certain day and the day before, then multiplied by the food energy values 20.79 kJ/g (HFD), 23.44 kJ/g (chocolate) and 30.34 kJ/g (ground peanuts). The sum of these three values is the combined food intake for a particular day.

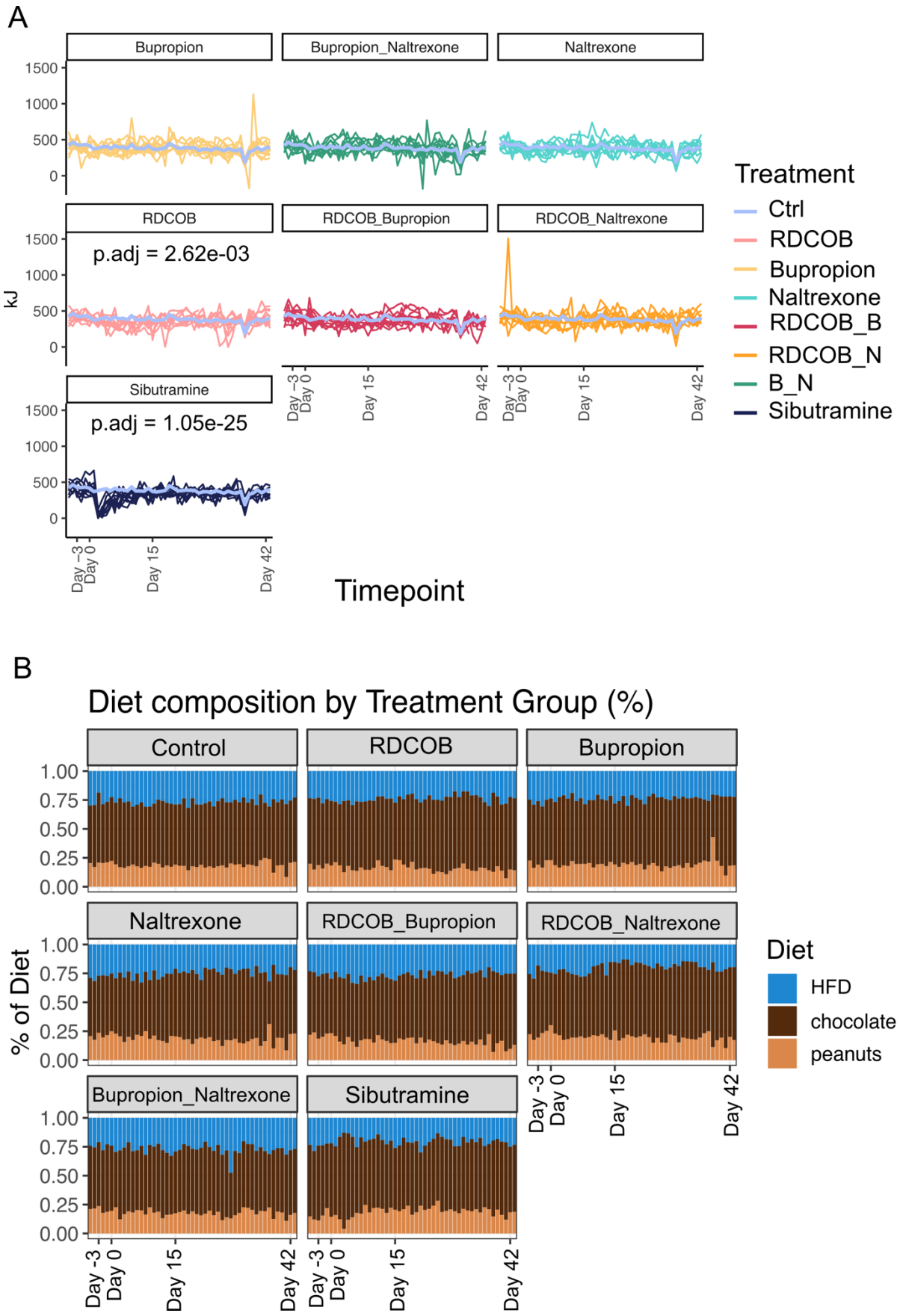


Figure 2.3 Sibutramine reduces food intake in obese rats. *A) Combined food intake (kJ) for each rat throughout the duration of the treatment. A violet line represents the average food intake for rats in the*

*control group. Statistical significance was assessed by ANOVA with the model: value ~ Day + Group + Day*Group + sample, (function anova_test within the rstatix R package, parameters: wid = sample, within = Day, between = treatment, dv = value) (Kassambara, 2020), and significant adjusted p-values for the day by group statistic have been reported on the appropriate panels. B) Bar plot representing the percentage of high fat diet (HFD), chocolate or peanuts consumed by each rat throughout the experiment.*

In line with its physiological mechanism, Sibutramine significantly lowers food intake, especially during the first two weeks of treatment (Figure 2.3, Sibutramine panel). While a similar result is shown in rats treated with RDCOB alone, albeit to a lesser extent, none of the other drugs tested significantly affect food intake.

Table 5. ANOVA results for combined food intake consumption. *Statistical significance for differences in food intake was computed with ANOVA using the model: value ~ Day + Treatment + Day*Treatment + sample, with the function anova_test within the rstatix R package (parameters: wid = sample, within = Day, between = treatment, dv = value) (Kassambara, 2020). Resulting p.values were corrected into adjusted p. values (p.adj) using the Benjamini-Hochberg correction. Adjusted P-values for every comparison in the model, together with generalized effect sizes (GES), representing the proportion of variance explained, are listed below. Notably, while all treatment groups show a significant change over time, investigation of the timepoints that contribute to this effect via Tukey's post-hoc test reveals that, aside from Sibutramine, this significance is driven by the fasting occurred on Day 36, the day before the OGTT test (see Appendix). Furthermore, Sibutramine and RDCOB have a significant association of Day by Treatment, but inspection of their GES reveals that while in Sibutramine this explains a good proportion of the variance observed (around 21%), in RDCOB this is much lower (7%).*

ANOVA	Day		Day by Treatment	
	P.adj	GES	P.adj	GES
RDCOB	3.88E-24	0.20	2.62e-03	0.07
Bupropion	1.5e-23	0.20	0.74	0.05
Naltrexone	2.25e-27	0.22	0.42	0.05
RDCOB Bupropion	2.59e-25	0.21	0.22	0.05
RDCOB Naltrexone	3.8e-19	0.18	0.42	0.05
Bupropion Naltrexone	5.27e-19	0.22	0.40	0.05

Sibutramine	9.95e-47	0.30	1.05E-25	0.21

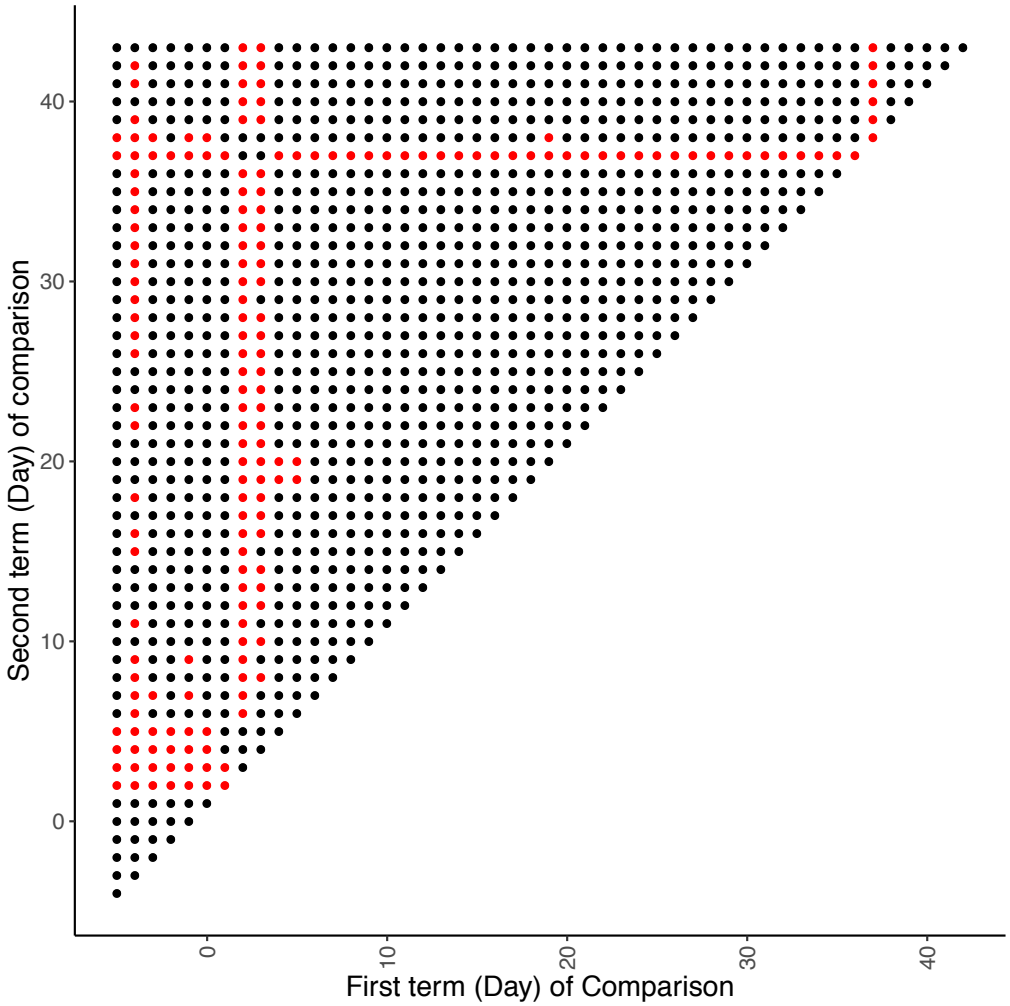


Figure 2.4 Tukey's post hoc test on Sibutramine reveals significant timepoint association during first week of treatment. *Visualization of p-values associated with the comparisons from Tukey's post-hoc test (P.value <= 0.05, red; P.value > 0.05, black). While the ANOVA test by Day resulted statistically significance in all treatment groups, mainly due to the values at Day 36, where rats were fasted in preparation for the OGTT test (see Figure 11.8 in Appendix), Sibutramine also shows significant differences during the first week of treatment (Day 2 and 3), as mirrored in Figure 2.3, confirming its effect on the reduction of food intake.*

Analysis of the food intake composition for each group revealed no particular dietary differences between Controls or Sibutramine-treated rats, as both groups showed a preference for chocolate, making up to 40-50% of their daily intake, followed by HFD (around 30-40%). Taken together, our results show that Sibutramine significantly reduces food intake.

2.3.4 Impact of weight loss drugs on the microbiome

To assess whether any of the weight loss drugs tested affected the microbiome, we initially performed diversity analyses on all treatment groups. α diversity measures the diversity of microbial species within an individual and it is generally lower in obese patients compared to their lean counterparts. Indeed, evaluation of the Simpson index, measuring the probability that two individuals randomly selected from a sample will belong to the same species, shows no real change in α diversity in any of the treatments tested, as the index is often around 1, meaning little change to diversity (Figure 2.5A).

β diversity assesses the diversity in species composition between samples. Results show that only Bupropion and Naltrexone treatment, as well as Sibutramine, clusters separately from controls (Figure 2.6B), suggesting a different species composition between these groups and, to a lesser extent, between rats treated with these drugs at the end of the treatment compared to their composition at the start (green or blue triangles compared to circles, Figure 2.5B).

Thus, while most drugs did not cause significant changes in α or β diversity, the combination treatment of Bupropion and Naltrexone, despite promoting little weight loss, had an impact on the microbiome composition of the rats. However, of all the drug tested, Sibutramine had the best performance by inducing weight loss and increasing Insulin sensitivity. For this reason, over the next paragraphs we will examine in more detail its impact on the microbiome.

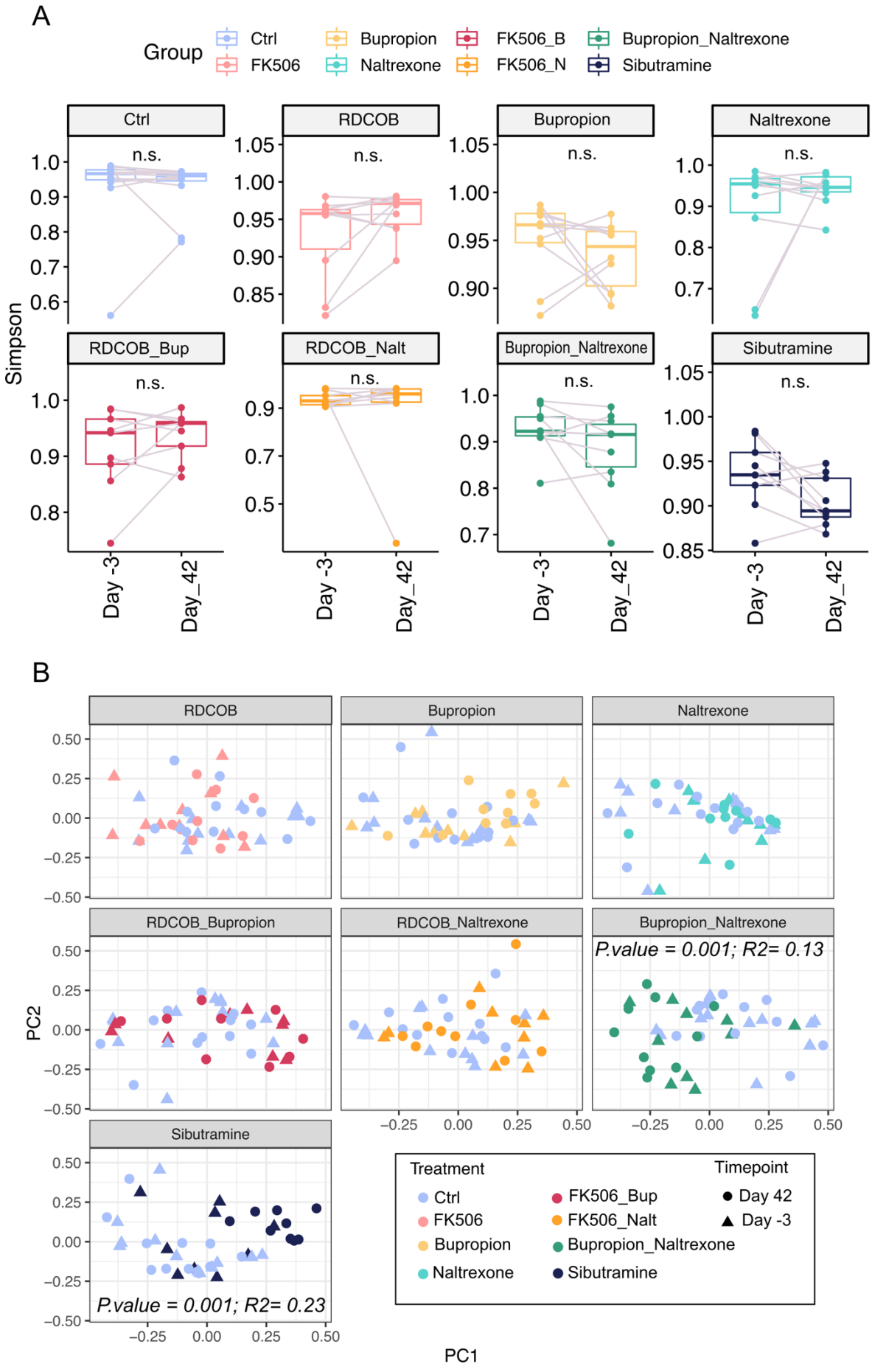

A
Group
Ctrl
FK506
Bupropion
Naltrexone
FK506_B
FK506_N
Bupropion_Naltrexone
Sibutramine
Simpson
Ctrl
RDCOB
Bupropion
Naltrexone
RDCOB_Bup
RDCOB_Nalt
Bupropion_Naltrexone
Sibutramine
n.s.
Day -3
Day_42
B
RDCOB
Bupropion
Naltrexone
RDCOB_Bupropion
RDCOB_Naltrexone
Bupropion_Naltrexone
Sibutramine
P.value = 0.001; R2= 0.13
P.value = 0.001; R2= 0.23
PC1
PC2
Treatment
Ctrl
FK506
Bupropion
Naltrexone
FK506_Bup
FK506_Nalt
Bupropion_Naltrexone
Sibutramine
Timepoint
Day 42
Day -3

Figure 2.5 Diversity analyses on all treatment groups reveal no clear alteration of alpha diversity, but few differences in beta diversity. *A) Boxplots of Simpson's index to estimate α diversity changes in each treatment group between the start (Day -3) and the end of the treatment (Day 42). Diversity analysis was carried out using Phyloseq's estimate richness function. Statistical significance was calculated by Student's paired t-test and resulting P.values were further corrected by the Benjamini-Hochberg method, but none of the comparisons were statistically significant. B) PCA plots showing β diversity amongst samples, calculated on relative abundances by Bray-Curtis distance between each sample, using the phyloseq R/Bioconductor package* (McMurdie & Holmes, 2013). *Statistical significance was assessed by adonis test (Treatment * Group + cohort + number). The combination of Bupropion and Naltrexone, as well as Sibutramine, were statistically significant. Bupropion and Naltrexone: P.value = 0.001, R^2= 0.13, Sibutramine: see section 2.3.5. Statistical significance levels: ns P > 0.05; * P < 0.05; ** P < 0.01; *** P < 0.00 (Control: N=12; Sibutramine: N=10 rats per group).*

2.3.5 Sibutramine impacts the microbiome

To evaluate the effect of Sibutramine treatment on the rats' microbiome, we first looked for changes in α and β diversity. While α diversity wasn't altered, β diversity showed significant differences (Figure 2.6A; R^2 = 0.23, p.value = 0.001).

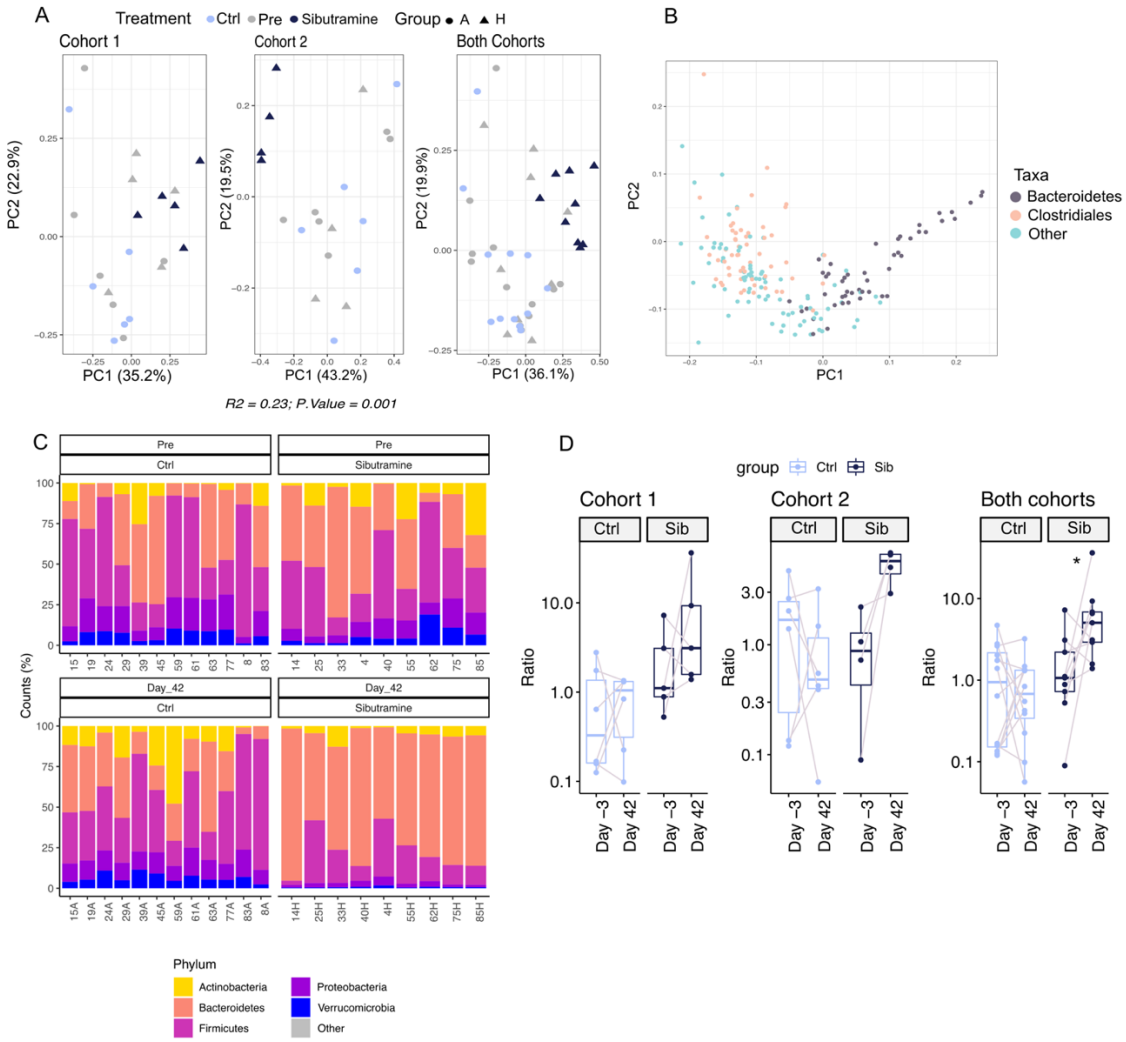


Figure 2.6 Sibutramine treatment effect on the microbiome. *A) PCA of β diversity amongst samples, calculated on relative abundances by Bray-Curtis distance between each sample, using the phyloseq R/Bioconductor package* (McMurdie & Holmes, 2013)*. Statistical significance was assessed by adonis test (Treatment * Group + cohort) and resulting variance explained (R^2) and p.value are printed at the bottom of the plot. Plot shapes describe the different treatment groups, with circles representing the Control group (A) and triangles the Sibutramine-treated samples (H). B) PCA of β diversity amongst taxa, calculated on relative abundances by Bray-Curtis distance between each sample, using the phyloseq R/Bioconductor package* (McMurdie & Holmes, 2013)*. C) Bar plot summarizing the percent composition of the top 5 phyla across control (left panel) or Sibutramine (right panel) samples, at the start (top) or the end (bottom) of the treatment. D) Boxplots of the comparison of Ratio Bacteroidetes/Firmicutes between Day -3 and Day 42 of Ctrl and Sibutramine-treated group. Statistical significance was assessed by T-test. Statistical significance levels: ns P > 0.05; * P < 0.05; ** P < 0.01; *** P < 0.00 (Control: N=12; Sibutramine: N=10 rats per group).*

In line with what is found in literature, our principal component analysis (PCA) shows a grouping of Control samples with the samples taken before the start of the treatment (Pre), while samples from Sibutramine treated rats are falling into a separate cluster (Figure 2.6A). Furthermore, ordination analysis of the taxa present in our dataset highlights a group of Bacteroidetes species separating along the first principal component (PC1), while a smaller cluster of Clostridiales spreads across PC2 (Figure 2.6B).

We next evaluated whether treatment with Sibutramine caused a shift in the ratio between Firmicutes and Bacteroidetes (Figure 2.6C and 2.6D). Several lines of evidence suggest that in obese individuals, microbes belonging to Firmicutes phylum are generally enriched compared to Bacteroidetes, thus giving a lower ratio of Bacteroidetes to Firmicutes (Ley et al., 2005; S. Louis, Tappu, Damms-Machado, Huson, & Bischoff, 2016; Turnbaugh et al., 2008). Upon weight loss, in both cohorts Sibutramine treated rats had an increased Bacteroidetes to Firmicutes ratio at day 42 vs. day -3 (Sibutramine mean at Day -3: 3.04 cohort 1, 1.07, cohort 2, 2.16, both cohorts; Sibutramine mean at Day 42: 12.23 cohort 1, 5.75, cohort 2, 9.35, both cohorts). Taken together, results show an influence of Sibutramine treatment on the gut microbiome.

2.3.6 Analysis of differential species confirms an association between Sibutramine treatment and gut microbiome composition

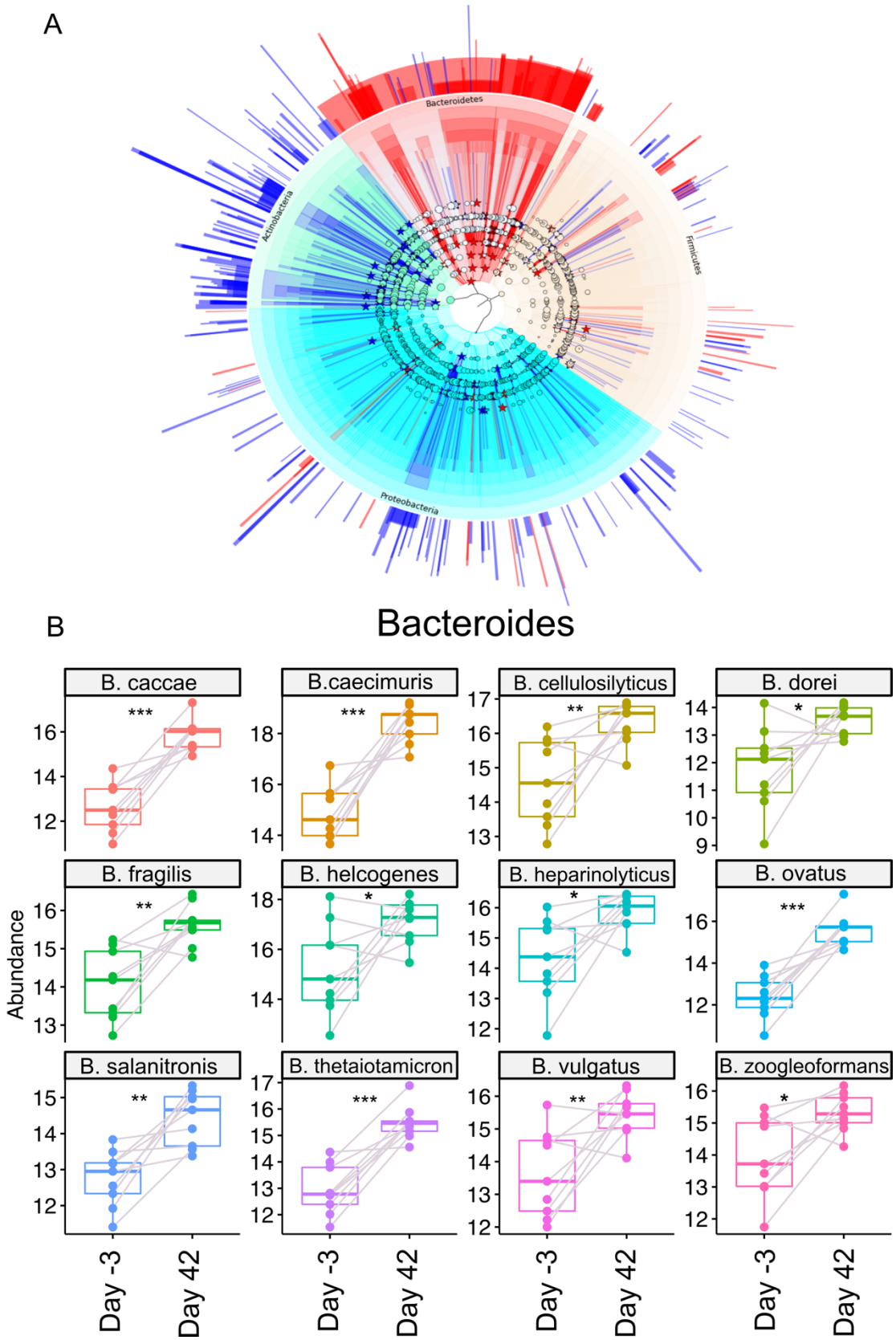


Figure 2.7 Differentially enriched species between rats before (Day -3) and after (Day 42) Sibutramine treatment. *A) Phylogenetic tree representation of differentially expressed taxa between samples at the start and end of the treatment. Differential enrichment at every classification level (Phylum, Class, Order, Family, Genus and Species) was calculated with R package DESeq2 using a linear model (~ Day + rat number). P values were adjusted by Benjamini-Hochberg correction and only species with P.adj <= 0.05 were considered significantly different, then tables with log2 and library size normalized counts (rlog) for every taxon were extracted* (Love, Huber, & Anders, 2014a; Zhao et

al., 2020). *These were fed into GraPhlAn for phylogenetic tree visualization. External annotation corresponds to Log2 fold-change values for every differentially expressed species. Species of interest were indicated in the circular tree by stars and colored according to their Log2 Fold change (Log2 fold-change < 0, blue; Log2 fold-change > 0, red). Concentric circles each represent a taxonomical level, starting from Phylum (Proteobacteria, light blue; Firmicutes, yellow; Bacteroidetes, red; Actinobacteria, green), then proceeding outwards to Class, Order, Family, Genus and Species. Notably, Bacteroidetes as a whole are significantly increased by the end of the treatment (see red star in the innermost circle). B) Boxplots showing levels of differentially enriched Bacteroides species, before and after Sibutramine treatment. Grey lines connect dots corresponding to the same rat. Counts used were log2 transformed and normalized by library size (Love et al., 2014a).*

We next focused on the differences between the gut microbiome of Sibutramine-treated rats before (Day – 3) and after Sibutramine treatment (Day 42). Analysis of differentially enriched species between the two groups highlighted statistically significant changes in 192 species (Figure 2.7). Amongst these, thirteen different species belonging to the Bacteroides genus, *B. thetaiotamicron, B. caecimuris, B. heparinolyticus, B. cellulosilyticus, B. fragilis, B. caccae, B. vulgatus, B. salanitronis, B. ovatus, B. helcogenes, B. zoogleoformans, B. dorei*, were significantly increased by Day 42. Interestingly, bacteria of this genus, and more generally of the Bacteroidetes phylum, have been found decreased in obese individuals in favour of Firmicutes species, thus their significant enrichment at the end of the treatment with Sibutramine might suggest a positive effect of the drug on the microbiome composition of the rats, further confirmed by the fact that the ratio of Bacteroidetes to Firmicutes shows a trend towards increasing by the end of the treatment (Figure 2.6D) (S. Louis et al., 2016; Turnbaugh et al., 2006). At the phylum level, Bacteroidetes are significantly increased by Day 42, while Firmicutes do not show overall significant changes, perhaps due to the fact that some of their species are increasing, while others are decreasing. Interestingly, most Actinobacteria species detected show a significant decrease by the end of the treatment (Figure 2.7A, green panel).

Moreover, we observe a redistribution of several probiotics species. Indeed, five members of Bifidobacteriaceae, such as *B. pseudolongum, B.animalis, B. catenulatum, B. dentium* and *B. scardovii* are significantly decreasing, while two Lactobacillales, *L.reuteri and L. ruminis*, are increased by the end of the treatment. Taken together, results show that Sibutramine treatment

significantly alters microbiome composition in obese rats, promoting potentially beneficial changes.

2.3.7 Functional analysis of rats microbiome reveals significant decrease of "Flagellar Assembly" related-genes

Single species changes might not explain all the differences between microbiome composition before or after Sibutramine treatment, as multiple species can code for the same gene. An advantage of shotgun metagenomics is the possibility of sequencing the whole microbial genome, rather than only its 16S rRNA, allowing the detection of genes that have the potential of being actively expressed within a sample. However, this method sequences the microbes' DNA, as opposed to their RNA, so it does not provide information regarding these genes' actual expression. Analysis of differentially abundant genes revealed statistically significant changes in 1047 genes, of which 536 are increased and 509 decreased in Sibutramine rats at the end of the treatment compared to the beginning (Figure 2.8).

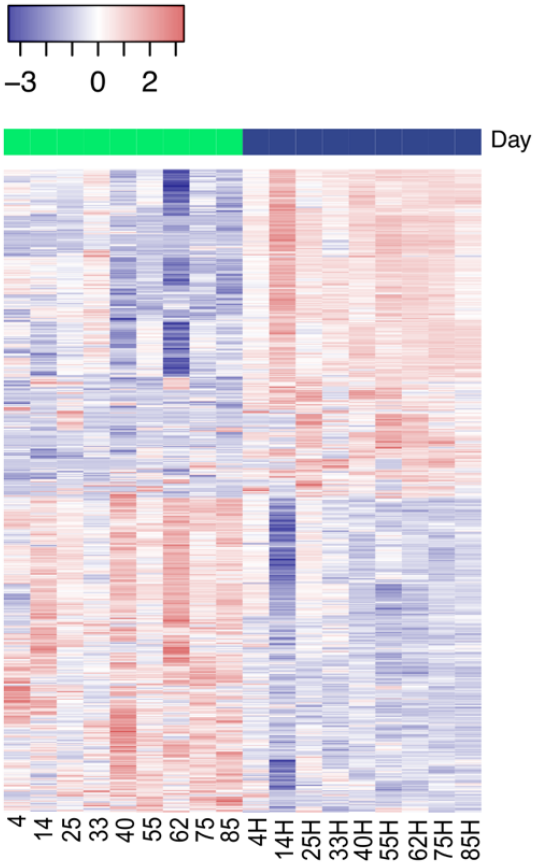


Figure 2.8. Differentially abundant genes between Sibutramine-treated rats at the start and end of the treatment. *A) Heatmap representing log2-transformed and normalized read counts for each of the 1047 differentially abundant genes. As with species, significant changes in gene levels were calculated using DESeq2 with the model ~ rat number + Day (Love et al., 2014a). P values were corrected with the Benjamini-Hochberg method, and only genes with adjusted P value <= 0. 05 were considered statistically significant. Top bar shows which columns correspond to Day -3 (green) or Day 42 (blue) samples.*

To functionally characterize these rats' microbiome, we performed gene set enrichment analysis (GSEA). This method aims at finding enrichment of entire gene sets rather than single genes, thus providing insight into the biological pathways contributing to a particular phenotype (Subramanian et al., 2005). Results show a significant enrichment of genes from the "Flagellar Assembly" set at the lower end of the ranked list, indicating their decrease in Sibutramine-treated samples (Normalized enrichment score (NES) = -1.71, p.adjusted = 0.03, Figure 2.9C). In bacteria, the flagellum is an hair-like structure used for locomotion, and has been shown to induce inflammation due to stimulation of interleukin 8 (IL8) release from intestinal epithelial cells (Neville et al., 2013). Moreover, different studies have shown an enrichment of this

pathway in obese patients, as well as in patients with metabolic syndrome (Hou et al., 2017; S. Louis et al., 2016). Further analysis of genes associated with this pathway highlighted a panel of 20 genes, all of which are differentially abundant and significantly decreased by the end of the treatment (Figure 2.9B).

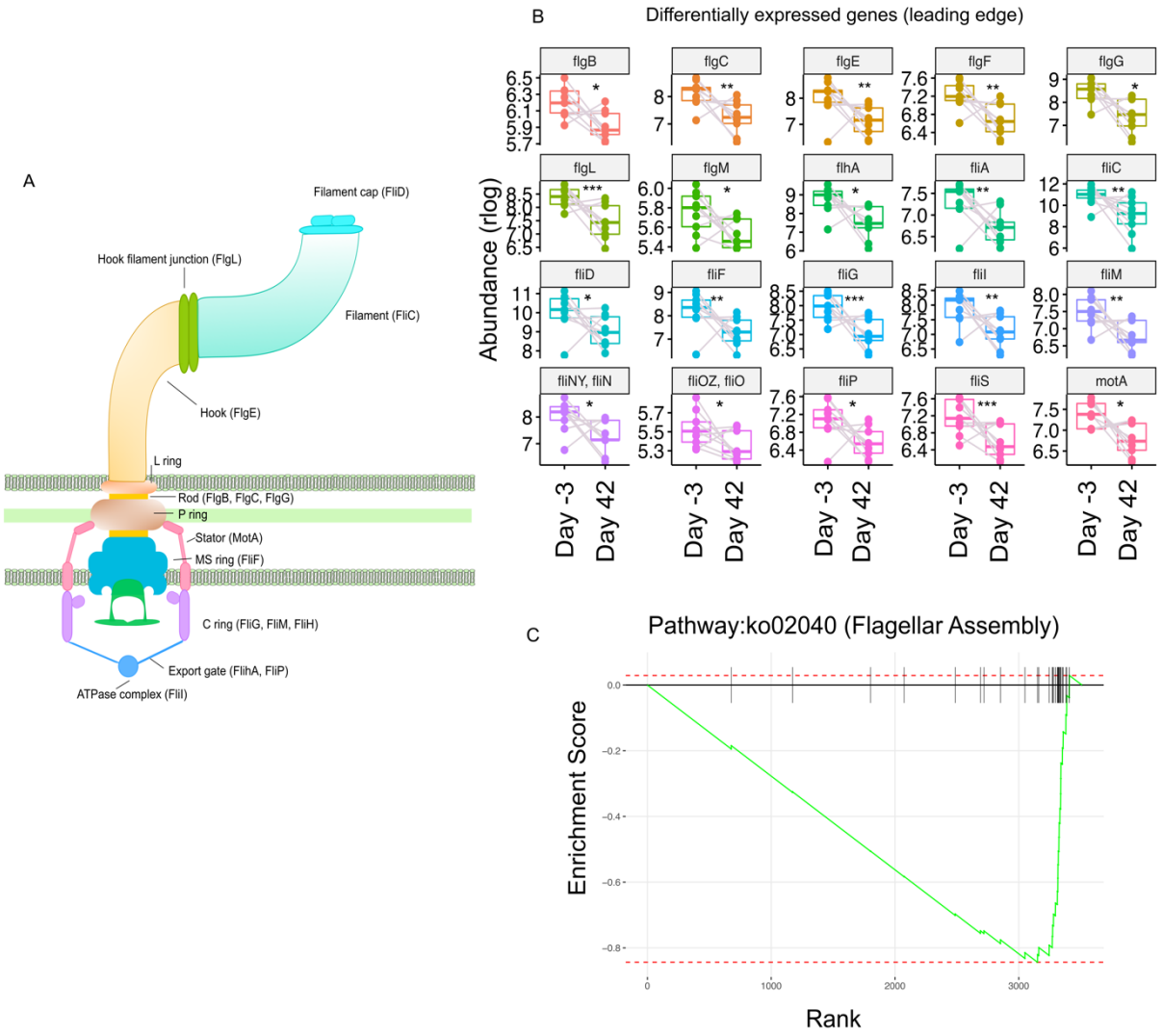


Figure 2.9 Gene Set Enrichment Analysis highlights decrease of genes related to flagellar assembly. *A) Schematic representation of the bacterial flagellum and its protein components. Genes detected in our analysis are indicated on the figure. Adapted from Minamino et al., 2018* (Minamino, Morimoto, Kawamoto, Terashima, & Imada, 2018)*. B) Boxplots of leading edge genes belonging to this pathway, all of which are differentially abundant in Sibutramine samples at Day 42 compared to Day -3. Boxplots were obtained by plotting the log2-normalized expression values for each gene. C) Enrichment plot for the pathway "Flagellar Assembly" (ko:02040). Briefly, genes have been ranked by log2 fold change multiplied by the -log10(p.adjusted), then gsea was performed with the fgsea R/Bioconductor package (Sergushichev, 2016). "Flagellar Assembly" was the only statistically significant result (normalized*

Enrichment Score: -1.7, p.adjusted = 0.03). Full gsea results table is available at https://github.com/SilviaRaineri/Thesis/tree/master/Microbiome.

Interestingly, the genes detected code for all the different components of the flagellum. FliD, fliC, flgL and flgE gene products make up for the extracellular portion of the flagellum, coding for the filament cap, the filament itself, the hook junction and the hook, respectively (Minamino et al., 2018). FlgB, flgC, flgF and flgG code for the rod, while motA and fliF are representative of the membrane-spanning basal body, coding for a part of the rotor and the MS ring, which provides the base for flagellar assembly (Bergeron, 2016; Takekawa et al., 2016). Finally, five genes, fliM, fliN, flhA and fliP, code for proteins making up the export apparatus on the cytosolic side (Barker, Meshcheryakova, Kostyukova, & Samatey, 2010) .

Further analysis of the species contribution to three representative genes, one for each major flagellum component, flgE, fliC and motA, found *Flavonifractor plautii* and *Roseburia hominis* as the two major sources of all the three genes considered (Figure 2.10). Interestingly, while *F.plautii* is differentially enriched at Day 42, *R. hominis* levels do not significantly change over time (*F. plautii*: Log2-fold change = 1.86, p.adjusted = 0.03; *R. hominis*: Log2 fold change = 0.93, p.adjusted = 0.49).

R. Hominis is a butyrate-producing bacterium from the Firmicutes phylum, and while its impact on flgE counts seems to be increasing by Day 42, in both motA and fliC it remains somewhat constant throughout the treatment (Figure 2.10B).

F.plautii is a flavonoid-metabolizing species from the Clostridiales family (Firmicutes) and its contribution to gene counts of flgE and motA increases in Sibutramine-treated rats by Day 42 (Figure 2.10B and D), while it is always present in the counts of fliC (Figure 2.10C). Interestingly, this species has been described as enriched in lean individuals compared to obese patients and has been linked to anti-inflammatory activities, potentially suggesting a positive change in the rats microbiome induced by Sibutramine treatment (Castaner et al., 2018; Mikami et al., 2020). This apparent discrepancy between levels of *F.plautii* increasing and flgE, motA

and fliC decreasing could be due to the high percentages of unassigned reads contributing to the gene counts (Appendix, Table 26).

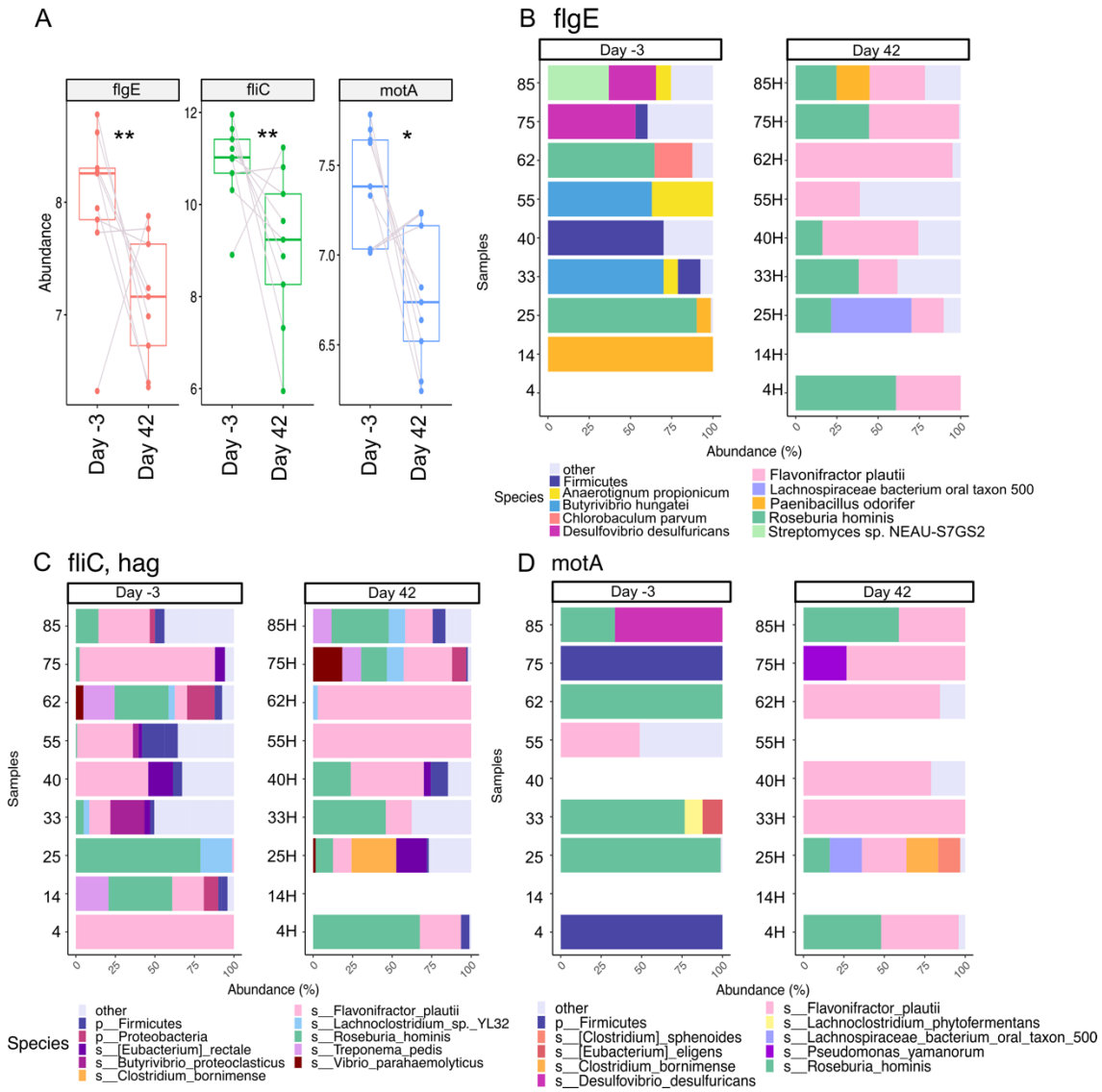


Figure 2.10 Species composition of three representative genes from the "Flagellar Assembly" pathway. *A) Boxplots showing expression levels of three representative members of the "Flagellar Assembly" pathway. Expression values have been log2 normalized as in Figure 2.9. B-D) Species contribution to the levels of flgE (B),fliC (C) and motA (D). Gene counts for each species are expressed as percentage and only the top 10 species for each gene are represented, whilst the remainder have been collapsed under the "other" category. Blank rows indicate absence of a certain gene in the corresponding sample. Empty rows correspond to samples that either had no reads annotated to the gene (motA, 55H and 14H; flgE 4 and 14H) or whose reads were entirely annotated to "k_unassigned" (fliC, 14H; motA, 14 and 40). Notably, every sample had a certain proportion of reads annotated to "k_unassigned" (Kingdom unassigned), and thus lacked a reliable annotation. These reads have been*

excluded from the dataset prior to plotting of fliC, flgE and motA. The percentage of unassigned reads for each gene, within every sample, is described in Appendix, Table 26.

Thus, results from our functional analysis support a role of Sibutramine treatment in positively affecting the gut microbiome. A decrease in genes related to flagellar assembly, previously associated with obesity-induced inflammation, as well as a higher contribution of anti-inflammatory species, such as *F.plautii*, to their differentially abundant gene counts suggests a shift towards a decreased inflammatory state, an opposite trend compared to patterns commonly associated with obesity.

2.4 Discussion

Our analysis of the gut microbiome of Sibutramine-treated samples revealed changes in the rats' microbiome concomitant with weight loss: an increase in α diversity, a distinct β diversity pattern and a shift in the Bacteroidetes to Firmicutes ratio towards a higher proportion of Bacteroidetes.

Analysis of differences in species composition between the beginning and the end of the treatment reveals a significant increase in 13 Bacteroides species, and in the Bacteroidetes phylum as a whole. In the context of obesity, this potentially suggests a more favourable microbiome. Indeed, *Bacteroides thetaiotamicron*, reportedly decreased in obese individuals and contributing to glutamate fermentation and short chain fatty acid production, is significantly increasing after treatment with Sibutramine (R. Liu et al., 2017). Furthermore, while several studies have found an increased presence of Firmicutes species, more efficient in extracting energy from nutrients and thus potentially favouring its excessive storage, compared to Bacteroidetes in obese individuals, we detect an increase of the latter by the end of Sibutramine treatment (M. J. Dalby, Ross, Walker, & Morgan, 2017; Turnbaugh et al., 2006). This pattern is, however, somewhat controversial, as there are contrasting reports in the literature, and Bacteroidetes and Firmicutes are the two most represented Phyla within the gut microbiome,

both containing members that have been linked to positive effects, and others with negative ones (Arumugam et al., 2011; Huttenhower et al., 2012; Ley, Turnbaugh, Klein, & Gordon, 2006; Turnbaugh et al., 2006).

Another potentially positive change induced by Sibutramine treatment is the decrease of genes related to flagellar assembly. Flagella are rod-like structures composed by flagellin and used by bacteria for locomotion. As host toll like receptor 5 (TLR5) recognises flagellin and is able to activate an immune response, increased presence of motile bacteria in the gut microbiota has been liked to a pro-inflammatory state in a range of diseases, including metabolic syndrome (Carvalho et al., 2012; Vijay-Kumar et al., 2010). Thus, a decrease of genes coding for this protein, along with other components of the flagellum might be indicative of a reduced inflammation in the gut of Sibutramine treated rats. As suggested by the significant decrease of genes involved in flagellar assembly, differentially abundant genes confirm positive changes to the gut microbiome, although results must be interpreted carefully. Indeed, while recent research in the field has highlighted microbial patterns associated with obesity at higher classification levels, such as the ratio of Bacteroidetes to Firmicutes and a lower α diversity, many of the species significantly changing are still largely unknown, making results interpretation difficult and often contrasting between published studies. Another complication is that, while over time metagenomic databases have expanded, there is still a consistent portion of reads that remains unassigned. For these reasons, results can be mainly descriptive, and while it is possible to associate certain species, genes or pathways to a disease condition of interest, little mechanistic insight can be provided without further experiments.

Existing reports on the effect of Orlistat and Naltrexone on the microbiome show several metabolic pathways being indirectly affected, mainly lipid metabolism for the former and secondary bile acids for the latter (Ke et al., 2020; F. Wang et al., 2018). Gsea results following Sibutramine treatment showed a statistically significant enrichment of genes belonging to the

"Flagellar Assembly" pathway towards the end of the ranked list, but metabolic pathways such as "Fatty Acid Metabolism" (Normalized Enrichment Score = 1.38, P.adj = 0.55, P.value = 0.06) and "TCA Cycle" (Normalized Enrichment Score = -1.4, P.adj = 0.55, P.value = 0.07), while ranking towards the top or bottom of the list, were not statistically significant.

Our analysis shows an influence of Sibutramine treatment on the microbiome of obese rats, but it cannot define whether this change is causing weight loss, or whether weight loss induced by Sibutramine is causing the changes in the gut microbiome. Perhaps a way to investigate this could have been to perform a faecal microbiota transplantation (FMT) from Sibutramine-treated rats at the end of the treatment, when their microbiome has already changed, to obese rats to evaluate potential effects on their weight or on metabolic markers, such as Insulin sensitivity.

Aside from the impact of Sibutramine on the microbiome, it's striking to note that many of the obesity treatments tested were not as efficient in inducing weight-loss, including Bupropion or Naltrexone, currently considered the clinical standard for obesity pharmacotherapy, although all treatment groups showed a high intra-group variability. Furthermore, despite existing reports on its impact on the microbiome, in the conditions tested here Naltrexone alone did not induce any significant changes to α or β diversity, while its combination with Bupropion significantly altered β diversity (Figure 2.5) (F. Wang et al., 2018). Sibutramine, however, has been recently discontinued in many countries due to its severe adverse cardiovascular effects (increased heart rate and blood pressure, fatal stroke), confirming the clinical need for an effective treatment of obesity (Srivastava & Apovian, 2017b).

Microbiome research is still a growing field, as many species remain largely uncharacterised and only recent advances in sequencing techniques have made it possible to bypass the limitation of culturing species in the lab. We will provide our complete shotgun metagenomics dataset to be publicly available as a resource, as more studies are needed to understand the functional role of specific microbial species and genes in metabolic diseases such as obesity.

IDH1: Linking chromatin and metabolism

3. Materials and Methods

3.1 Cell lines

Human glioblastoma cell line U-87 MG WT and its isogenic counterpart, IDH1 mutant U-87 MG, bearing the heterozygous R132H mutation were purchased from ATCC (England, UK). Cells were cultured according to the manufacturer's instructions in EMEM (ATCC, England, UK) supplemented with 10% heat-inactivated FBS (Gibco, Thermofisher, USA). Cells were split every two to three days in a 1:2 or 1:3 dilution in T175 flasks. Floating cells were collected by gentle centrifugation at 100g, 5 minutes, room temperature (RT) and added back to the culture. Cells were maintained in a 5% CO_2, 37°C humidified atmosphere. Checks for mycoplasma contamination were routinely performed according to Uphoff et al. (Uphoff & Drexler, 2014).

3.2 RNA Sequencing

3.2.1 48-hours treatment with GSK864 of U87 WT and R132H IDH1 cells

Human glioblastoma cell line U-87 MG WT and its isogenic counterpart, IDH1 mutant U-87 MG, bearing the heterozygous R132H mutation were purchased from ATCC (England, UK). Cells were cultured according to the manufacturer's instructions (see Appendix 4).

U87 WT and R132H IDH1 cells were plated at a density of 1 million in a T75 flask. The following day, cells were treated with 0.5 μM GSK864 (Sigma), a mutant IDH1 inhibitor (Okoye-Okafor et al., 2015) diluted in DMSO (Sigma) or DMSO only. After 48 hours of incubation with the drug, adherent cells were scraped in 2 ml Qiazol Lysis Reagent (Qiagen, Hilden, Germany), while floating cells were pelleted by centrifugation (5 minutes, 200g, RT) and added to the lysed sample. 2% Drosophila cells were added as spike-ins. Lysates were stored at -80°C, until all three biological replicates were ready for RNA extraction.

3.2.2 One week treatment with GSK864 of U87 WT and R132H IDH1 cells

Cells were plated at a density of 1 million in a T75. On the following day, 0.5 µM GSK864 or DMSO were added to the culture. Cells were kept in culture for a total of one week incubation with GSK864, splitting 1:2 and replenishing the drug every 48 hours, as shown in the diagram below. Harvesting was performed as described in paragraph 3.3.1.

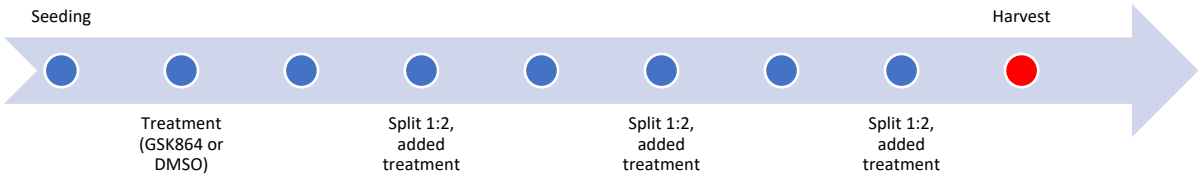


Figure 3.1. Timeline of one-week treatment with GSK864 experiments.

3.2.3 RNA extraction and library preparation

Cells were lysed in the appropriate amount of Qiazol Lysis Reagent (Qiagen), then phase separation was induced by addition of 160 µl chloroform. After centrifugation for 15 min at 4°C, the upper phase was transferred to a new tube and RNA was precipitated with 400 µl isopropanol (Sigma). RNA pellet was resuspended in RNAse-free water and treated with 2 U DNAse I (Qiagen) to remove residual DNA. After a 30 minute incubation at 37°C, RNA was precipitated with isopropanol and pelleted. Pellets were washed three times in 75% cold Ethanol. Finally, the pellets were air-dried 20 min at room temperature, then resuspended in 80 µl and stored at -80°C. 3'-mRNA sequencing libraries were prepared in three technical replicates for each sample using the QuantSeq 3'-mRNA-Seq library prep kit from Lexogen (Wien, Austria) following the kit instructions. Final libraries were quantified using a High-Sensitivity DNA Chip (Agilent, California, USA) on a 2100 Bioanalyzer (Agilent). Libraries were diluted to a final concentration of 250 pM and sequenced on an Ion Proton instrument.

3.2.4 Bioinformatic Analysis

Raw FASTQ files were downloaded from the Ion Torrent server, then uploaded to Galaxy (https://usegalaxy.org/) for preprocessing. Reads quality was assessed with FASTQC, then files were trimmed with trimmomatic using the options MINLEN = 20, SLIDING_WINDOW averaging over 4 base-pairs and average quality required = 20 (Bolger et al., 2014). Trimmed reads were loaded on the arc cluster (https://www.arc.ox.ac.uk/home), where samples were aligned to a custom human (hg38) and Drosophila (dm6) genome using Hisat2 with the --no-spliced-alignment option (D. Kim, Langmead, & Salzberg, 2015). Resulting bam files were filtered to retain only uniquely aligned reads and indexed using samtools, then counts tables were generated using HTSeq-count (Anders, Pyl, & Huber, 2015; H. Li et al., 2009). Quality control for successful 3'end amplification was conducted by visualization of bigwig files on the IGV genome browser (J. T. Robinson et al., 2011). All scripts were loaded on github (https://github.com/SilviaRaineri/RNASeq_analysis).

Exploratory analysis of data was done using NOISeq (Tarazona et al., 2015). Raw reads were normalized according to Drosophila spike-ins, trimmed mean of M values (TMM) or Deseq2 standard method. Normalization factors based on Drosophila spike-ins were calculated in base R as follows. The total number of human or Drosophila reads were calculated in each sample. The sample with the lowest number of human reads was selected as calibrator. The total number of Drosophila reads in the calibrator was divided by the total number of Drosophila reads in each sample, thus obtaining a normalization factor for each treatment group. Finally, each number in the raw gene counts table was multiplied by the appropriate normalization factor. Principal component analysis (PCA) plots before and after each normalization were examined to assess the presence of batch effects and evaluate the most appropriate normalization. Batch effects, if present, were removed using the ARSyNseq function from the NOISeq Bioconductor/R package. Differential expression analysis was carried out using and comparing

outputs from different R/Bioconductor packages, such as DESeq2, Limma and edgeR (Love et al., 2014a; M. D. Robinson, McCarthy, & Smyth, 2009; Smyth et al., 2015).

3.3 Reverse transcription-polymerase chain reaction (RT-PCR) and real time PCR

3.3.1 Reverse Transcription

One µg of RNA was reverse transcribed into cDNA with the SuperScript III Reverse Transcriptase kit, according to manufacturer's instructions (Invitrogen). The reaction mix used 50 ng/µl random hexamers, 10 mM dNTPs, 0.1 M dithiothreitol (DTT), 40 U/µl RNAse OUT, 200 U/µl Superscript III reverse transcriptase, DEPC-treated water to a final reaction volume of 20 µl. PCR program was the following: 5 min at 65°C (denaturation step); 1 min at 4°C; 5 min at 25°C, 45 min at 50°C (reaction step); 15 min at 75°C (enzyme inactivation step). Negative controls included samples with RT- mix (without the enzyme) and water with RT+ mix. At the end of the program, samples were treated with 1 µl RNAse H for 20 min at 37°C. As a quality control, 50-100 ng of starting RNA template, including an equal amount of negative controls, were subjected to a PCR for the ACTB transcript, according to the protocol described in the next paragraph, before proceeding with further experiments.

3.3.2 PCR and qPCR primer design

Unless otherwise specified, primers have been designed as follows. Human sequences of interest were downloaded from the Ensembl genome browser (http://www.ensembl.org/). The sequence was then given as an input on primer3 (http://primer3.ut.ee/), selecting a maximum T melting difference of 2 degrees Celsius and, where possible, an overlap junction to span between two exons. Primers specificity was then tested on Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). All primers were generated by Eurofins (Luxembourg).

Name	Sequence	$T_{melting}$

PDGFRA_RT_FW	GCTCAGCCCTGTGAGAAGAC	61.4
PDGFRA_RT_Rev	ATTGCGGAATAACATCGGAG	55
RPLP0_RT_FW	TCCCACTTGCTGAAAAGGTCA	57
RPLP0_RT_Rev	CCGACTCTTCCTTGGCTTCA	59
ACTB_RT_Fw	AGAAAATCTGGCACCACACC	57
ACTB_RT_Rev	AGAGGCGTACAGGGATAGCA	59
PDGFRA_std_fw	GACACTGGGAGATTCGGAGC	60
PDGFRA_std_rev	AAGCTGGCAGAGGATTAGGC	59
RAB3IP_rt_fw	CAACCCGTGTATTCATCCCC	58.32
RAB3IP_rt_rev	GCCTCTGTTATAGTAGCACCCT	59.03
DNAJC12_4_5_fw	CCCAAGCCCCTAGAGAAGTC	59.17
DNAJC12_4_5_rev	GCAGCATAGGGGACAGTCTT	59.4
DNAJC12_bis_fw	GCACTGGGTTGTCAGAGGTAA	60
DNAJC12_bis_rev	ACTTCTCTAGGGGCTTGGGT	60
KCRK3_last_fw	CATGACCATGAACGCCGAGG	61
KCRK3_last_rev	ACGAGCACGGAAGTGCAG	60
KCRK3_bis_fw	CTGCACTTCCGTGCTCGTG	60
KCRK3_bis_rev	GAGGGCGTGTCGCTGTAG	60
RAPGEF4_5_8_fw	TGGTGTTATGGAAACGGGCT	59
RAPGEF4_5_8_rev	GTGCTCGAGAGAGAATGGCA	59
BEX2_5'_fw	CGATCCTGAGGCTACGACC	59
BEX2_5'_rev	CCCGCAAACCACCATTTTCT	59
ARMCX4_5'_fw	CCTCGGATCATTACAGGTCCA	58
ARMCX4_5'_rev	GCTGGGGTAAAGAGAGTGGT	59

ARMCX4_ex2_fw	GAGTCAGGTCAGTGGAAGCT	59
ARMCX4_ex2_rev	CCACACCCCAGGAACCTATT	59
TCEAL7_fw	CGCACAGAGTTAAGCAGGAA	58
TCEAL7_rev	GCTGGCGTTCAAATTCTCCA	59
TCEAL7_ex3_fw	GGAGAATTTGAACGCCAGCA	59
TCEAL7_ex_rev	TCCATCTCATCTCCCTCCCT	58
L1Hs ORF1_fw	GAATGATTTTGACGAGCTGAGAGAA	59
L1Hs ORF1_rev	GTCCTCCCGTAGCTCAGAGTAATT	61
L1Hs ORF2_fw	CAAACACCGCATATTCTCACTCA	59
L1Hs ORF2_rev	CTTCCTGTGTCCATGTGATCTC	58
AluS-Fwd	GCCGAGGCGGGCGGATCACC	69
AluS-Rev	GCCTCCCGAGTAGCTGGGAT	63
AluY-Fwd	AGATCGAGACCATCCTGGCT	60
AluY-Rev	CCGCCTCCCGGGTTCACGCC	60
L1MD2_fw	GCTGAGAATTTTACCCACTGCA	59
L1MD2_rev	TGGGACATTTCTGAAATCACCT	57
L1MD2_2_fw	GGCAATGATGTCTGTTTTCACC	58
L1MD2_2_rev	TGTGTTTGTGTCATATTCTGGGA	58
HAL1_fw	TGCATAACTTTTCTTTCACCCAC	57
HAL1_rev	CAACCAAGGGCTCCAAGAAG	50
HAL1_2_fw	ACTTTTCTTTCACCCACTAGCT	57
HAL1_2_rev	GAACAACCAAGGGCTCCAAG	59
L1MA10_rev	GCCCAGAAGTAAACCCATGC	59
L1MA10_fw	GCTTAATGATTTCAGGTCCTGC	57

L1MA10_2_fw	TGGAGTACGAGGAAAACAGGA	58
L1MA10_2_rev	GGTGTTTGCCAAAATCAAGGT	57

Table 6. Primer design. *Primer sets PDGFRA_RT, RPLP0_RT and ACTB_RT were taken from Flavahan et al.* (W A Flavahan et al., 2016)*, while L1Hs_ORF1, L1Hs_ORF2, AluY and AluS were described in (Pereira et al., 2018). All other pairs were designed as described. Primers sequences and melting temperatures ($T_{melting}$) are indicated.*

3.3.3 PCR

50-100 ng of cDNA were given as an input for PCR amplification using *Taq* polymerase (Sigma). The PCR mix included 1x PCR buffer, 2.5 mM dNTPs, 0.4 µM Forward primer, 0.4 µM Reverse primer and 0.05 U/µl *Taq* DNA polymerase. The reaction was carried out with the following program: 5 min at 95°C; 25 cycles of 95°C, 30 seconds (s), 55°C, 30 s, 72°C, 30s; 7 min at 72°C. PCR products were then loaded on a 2% agarose gel for visualization.

3.3.4 Real time PCR

cDNA samples were diluted 1:100 prior to sample preparation for real time PCR. Reactions were set up mixing 3.5 µl of diluted cDNA, 4 µM of primers, 2x SensiMix™ SYBR® Hi-ROX (Bioline, United Kingdom) with the appropriate amount of milliQ water. Cycling profile was as follows: hold 10 min 95°C; 45 cycles of 95°C, 7s, 58°C, 12s, 72°C, 12s. Crossing point cycle (Cp) values were extrapolated with the Second Derivative Maximum Method (qRex software, QIAGEN), and relative expression of the genes of interest was calculated with the $2^{-\Delta\Delta ct}$ method, using the values of ACTB as reference. The specificity of the amplified product was assessed by melting curve analysis with the dedicated qRex software (QIAGEN). Primer design is detailed in Appendix 7.

3.4 Sample preparation for TAPS sequencing

U87 cells WT or R132H for IDH1 were plated and harvested as described in 3.1.2. DNA was extracted as follows. Cells were scraped in 1 ml ice-cold PBS, then pelleted by centrifuging 5 min, 200g at room temperature. Pellets were resuspended in 500 µl of DNA lysis buffer (2%

TritonX-100, 1% SDS, 100 mM NaCl, 10 mM Tris HCl pH 8.0, 1 mM EDTA), and incubated 1 hour at 37°C. Phase separation was induced by adding 500 µl of phenol:cholorform (Thermofisher), vortexing samples 10 min and centrifuging for 10 min, 13000 rpm, at 4°C. After transferring the upper phase to a new tube, DNA was precipitated with 500 µl Ethanol 100% (EtOH) and 100 µl 5 M NaCl, then incubated 15 min at -80°C. Samples were centrifuged for 10 min, 13000 rpm, at 4°C, then washed twice in 1 ml 70% EtOH. DNA pellets were air-dried 5 min at 65°C, then resuspended in 50 µl RNAse-free water. After treatment with RNAse A for one hour at 37°C, samples were stored at -20°C.

DNA from N=3 biological replicates was pooled and quantified by Qubit using a high sensitivity DNA kit (q32851, Thermofisher). 1.5 µg of each sample (WT control, WT GSK864, R132H control, R132H GSK864) were given to our collaborators for TAPS library preparation.

3.5 Chromatin Immunoprecipitation followed by sequencing (ChIPSeq)

U87 cells WT or R132H for IDH1 were plated and harvested as described in paragraphs 3.1.1 and 3.1.2, for a 48-hour or a week-long treatment with GSK864.

3.5.1 Immunoprecipitation

Cells were scraped in 1 ml ice-cold Phosphate-buffered Saline (PBS), then counted and 2 million cells were taken as an input for each sample. Chromatin was crosslinked with 37% formaldehyde (Thermofisher) for 5 min at RT, followed by 2.5 M Glycine, 5 min at RT. 0.5% Drosophila chromatin was added as a spike-in control. To further concentrate the samples, nuclei were extracted by incubation in a swelling buffer composed of 10 mM Tris-HCl pH 8.0, 10 mM NaCl and 0.2% NP-40. After gentle centrifugation (5 min, 3000 rpm at 4°C), nuclei were lysed in 2 ml lysis buffer (10 mM Tris-HCl, pH 8.0, 1% NP-40, 0.5% Na Deoxycholate and 0.1% SDS). Samples were sonicated for 2.30 or 3.30 hours on a Bioruptor sonicator set on High, 30 seconds on and 30 seconds off. Chromatin was immunoprecipitated overnight with 10 µg of the appropriate antibody together with 5 µg of α-Histone H2av (IP sample, see table below) or water (NO, negative control sample) and Protein-A-conjugated beads on a rotating

wheel at 4°C, while 200 µl of crude lysate was stored at -20°C as the Input fraction. After washing in TSE-150 (1% Triton-X 100, 0.1% SDS, 2 mM EDTA, 20 mM Tris-HCl pH 8.0 and 150 mM NaCl), TSE-500 (same composition of TSE-150, but with 500 mM NaCl), and LiCl solution (0.25 M LiCl, 1% NP-40, 1% Deoxycholate, 1 mM EDTA and 10 mM Tris-HCl pH 8.0), IPs and NO were incubated with the DNA Tagmentation enzyme (Illumina) for 10 min at 37°C. The reaction was stopped by adding 1 ml TSE-150 and incubating 5 min on ice. Samples were further washed with 2 ml TSE-150, followed by 2 ml TE. Chromatin from all samples (Input, NO, IP) was de-crosslinked by adding 5M NaCl and incubating 3 hours at 65 °C. Contaminant RNA was degraded by treating with RNAse A (Thermofisher) for 1 hour at 37°C. Finally, proteins were degraded by treating overnight with Proteinase K (Thermofisher) at 65°C. The following day, DNA was purified from all samples with the Qiagen MinElute PCR purification kit according to the manufacturer's instructions.

Antibody	Catalog number	Distributor
α-H3K9me3	ab8898	Abcam
α-H3K27ac	07-360	Upstate
α-CTCF	3418	Cell Signaling Technologies
α-Histone H2av	39715	Active Motif

Table 7 Antibodies used in ChIP Seq. *All antibodies were used in a quantity of 10 µg, with the exception of α-Histone H2av, which recognizes a Drosophila protein and due to its high efficiency was used at 5 µg.*

3.5.2 Quality Check: qPCR

During protocol optimization with a new antibody, a qPCR was performed as previously described (paragraph 6.4) as a quality check for immunoprecipitation. Primer design included examining, whenever possible, publicly available datasets on U87 cells from the WashU genome browser (http://epigenomegateway.wustl.edu/browser/) or publications to select positive and negative control loci (William A Flavahan et al., 2016; Daofeng Li, Hsu, Purushotham, Sears, & Wang, 2019). A list of the primers is provided below.

Primer	Sequence	IP control
PSMB1ctcf_FW	CCTTCCTAGTCACTCAGTAA	CTCF positive control

PSMB1ctcf_Rev	CAGTGTTGACTCATCCAG	CTCF positive control
SPG11ctcf_FW	CAGTACCAGCCTCTCTAG	CTCF positive control
SPG11ctcf_Rev	CTAAGCTAGGCCTTCAAG	CTCF positive control
ACTB_FW	TCACCCTGAAGTACCCCATC	Negative control
ACTB_Rev	CCACACGCAGCTCATTGTAG	Negative control
ACTB2_FW	GACCTGACTGACTACCTCATGA	Negative control
ACTB2_Rev	GCACAGCTTCTCCTTAATGTCA	Negative control
CDH13_fw	CGATGCTTGGCTAGAGGACG	H3K9me3 ChIP positive ctrl
CDH13_rev	CCGCTCGCAACTTCAGGC	H3K9me3 ChIP positive ctrl
LACT1BL1_fw	GTCTTCCAGCTGCACGTG	H3K9me3 ChIP positive ctrl
LACT1BL1_rev	CACTCTGTACGTGTTGAGGC	H3K9me3 ChIP positive ctrl
RPL18_fw	CCCAGTATCGGGAATTGCTCC	H3K27ac ChIP positive ctrl
RPL18_rev	GATAATCCGCTGCCATCCG	H3K27ac ChIP positive ctrl
SYN1_lastEx_FW	GACCAATGCCTTCAACCTTCC	H3K27ac ChIP negative ctrl
SYN1_lastEx_Rev	CTCAAGGGATTTGGAGACTCC	H3K27ac ChIP negative ctrl

Table 8 Primers for quality check of ChIPSeq. *Primer sets PSMB1 and SPG11 were used as described in Flavahan et al, 2016.*

3.5.3 Library Preparation and Sequencing

Libraries from immunoprecipitated chromatin were prepared using either 10 (IPs) or 6 µl (Input) of DNA. Following amplification with PCR using the Nextera DNA Library Prep kit (FC-121-1030, Illumina), libraries were size-selected and purified in three subsequent rounds of purification using Axyprep Mag PCR clean up beads (AX402, Appleton Woods, UK). Beads to DNA ratios used were 1.8x, 0.65x and a final round of 1x. Library concentration was determined by quantification with Qubit high sensitivity DNA kit (q32851, Thermofisher). Sequencing of ChIP libraries was performed on an Illumina NextSeq machine, using the NextSeq 500/550 High Output kit v2.5 (75 cycles, #20024906, Illumina). Samples were run on a High Sensitivity DNA Bioanalyser chip to determine their concentrations. Up to 8 libraries were diluted to a final concentration of 4.5 nM, then pooled. A minimum of N=2 repeats were sequenced for each ChIP experiment.

3.5.4 Bioinformatics Analysis

Raw paired-end fastq sequences were downloaded from BaseSpace, then loaded on Galaxy (https://usegalaxy.org/). Here, preprocessing of sequences entailed a run on FastQC for quality

control, followed by concatenation of files belonging to the same samples, to obtain two read files per samples, R1 and R2 (Andrews, n.d.). Paired-end reads were further processed by removing low quality sequences and trimming with Trimmomatic (SLIDINGWINDOW, number of bases to average across = 4, average quality required = 20; MINLEN = 30) (Bolger et al., 2014). After a second FastQC run to re-assess quality, trimmed sequences are aligned to a custom human (hg38, after removal of repetitive regions) and Drosophila (dm6) genome with Bowtie2 (Langmead & Salzberg, 2012a). As an additional quality control step, twenty unaligned FASTA sequences are randomly selected from each sample and given as an input in BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to exclude sample contamination.

Aligned bam files are further processed with Samtools to remove PCR duplicates (samtools markdup with the -r option to not only mark but actively remove duplicates) and filtered for high quality alignments (samtools view with MAPQ > 30) (H. Li et al., 2009). Immunoprecipitation efficiency for each biological replicate was assessed by a fingerprint plot using Deeptools plotFingerprint. Further differential analyses were carried out as described in Turcan et al, 2018. Briefly, N=3 (H3K9me3 and CTCF ChIPs at 48 hours) or N=2 biological replicates bam files were merged using samtools merge. As Drosophila spike-ins proved to introduce more noise in some of the dataset, in order to maintain consistency between ChIPSeq analysis of the different datasets, Drosophila reads were filtered out and broad (H3K9me3, using a cutoff of 0.01) or narrow (H3K27ac, CTCF) peaks were called using macs2 (Gaspar, 2018). Reads Per Kilobase of transcript, per Million mapped reads (RPKM) values were calculated for each Input and IP sample, and only peaks with RPKM (IP)/ RPKM (Input) > 0.5 were selected for further analysis. Bigwigs were produced by normalizing to 1x coverage with Deeptools bamCoverage, then the log fold change of IP/Input was calculated using bigwigCompare. Results were visualized in heatmaps using Deeptools computeMatrix followed by plotHeatmap (Ramírez, Ryan, Grüning, et al., 2016). Peaks were annotated using HOMER's AnnotatePeaks, then differentially enriched peaks amongst different conditions

were obtained with getDifferentialPeaks (HOMER) with default settings (Heinz et al., 2010). Since the software compares only two conditions at a time and outputs a log2 fold change value, the following pairs were analysed: WT control vs R132H control (to find peaks decreased in R132H); R132H control vs WT control (finds peaks enriched in R132H); WT control vs WT GSK864; WT GSK864 vs WT control; R132H control vs R132H GSK864 (finds peaks decreased in R132H GSK); R132H GSK vs R132H control (finds peaks enriched in R132H cells upon GSK treatment). Further analyses were carried out using R/Bioconductor packages. The IGV genome browser was used for visualization of the 1xCoverage-normalized bigwigs, were tracks from the same dataset (48 hours or one week) were grouped together and autoscaled to the same data-range (R. G. Verhaak et al., 2010).

3.6 Metabolic Assays: Glucose and Glutamine Uptake

One million U87 WT or R132H IDH1 cells were plated in a T75 flask. The following day, 0.5 μM GSK864 or DMSO were added to the medium. Media aliquots of 200 μl were taken 8, 24, 48 or 72 hours after plating. Aliquots were stored at -20°C until ready to be measured with the appropriate kit. Glucose levels in the medium were measured with the Glucose Colorimetric Detection kit (EIAGLUC, Thermofisher), performed following the manufacturer's instructions. Samples were diluted 1:16 in Assay buffer to ensure a more accurate quantification.

To determine Glutamine levels, samples were diluted 1:800 and processed with the Glutamine/Glutamate-Glo Assay kit (J8021, Promega), following the manufacturer's instructions.

3.7 Metabolic Flux of Glucose and Glutamine

Our metabolic flux aimed at following labelled carbons from 13C-Glucose or 13C-Glutamine within U87 WT Control (DMSO only), WT treated with GSK864, R132H cells Control or

treated with GSK864. One million cells were plated in unlabeled medium (ATCC EMEM 30-2003) in T75 flasks, in a total of five technical replicates per experimental group, as described in the diagram below.

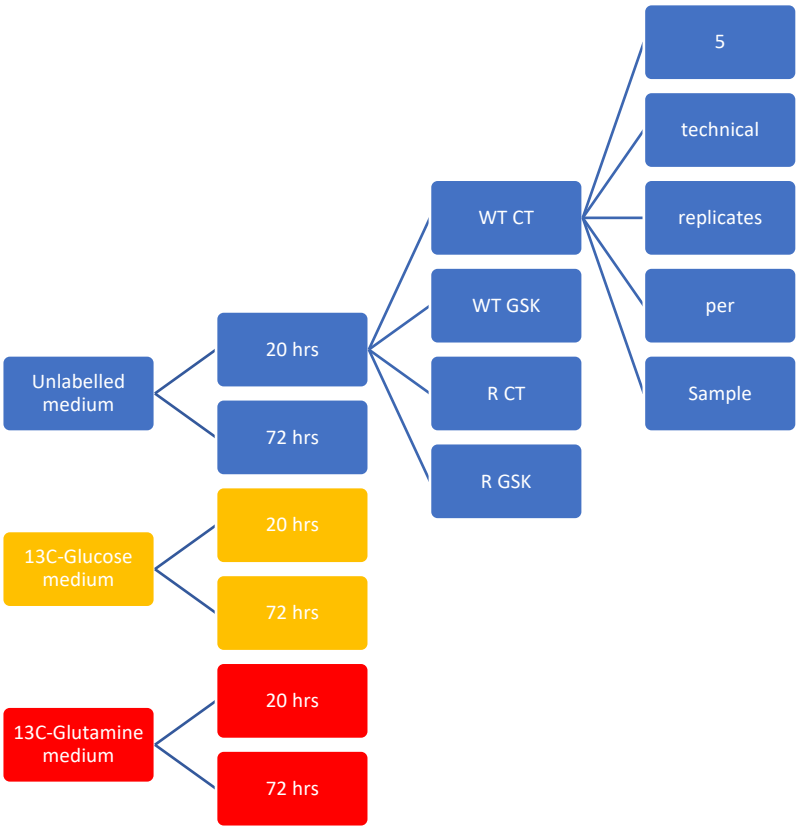


Figure 3.2. Diagram of metabolic flux samples collected*The experiment included three different labelling groups: ^{13}C-Glucose (yellow), ^{13}C-Glutamine (red), or unlabelled (blue) medium. Each of these conditions further included two different timepoints (20 or 72 hours), composed of four different groups (WT controls, WT treated with GSK864, R132H controls, R132H treated with GSK864).*

After 8 hours, medium was changed as appropriate to 13-C labelled Glucose medium (Nacalai Tesque 09848 EMEM supplemented with 1 g/l 13C6-Glucose (389374-1G, Sigma), 0.11 g/l Sodium Pyruvate (S8636, Sigma) and 10% FBS (10500064, Gibco)), 13C-labelled Glutamine (EMEM M5650 (Sigma), supplemented with 0.292 g/l 13C5-Glutamine (605166-500MG, Sigma), 0.11 g/l Sodium Pyruvate and 10% FBS) or unlabeled medium.

Twenty hours after plating, a first timepoint was collected by removing the medium, washing in PBS, and adding 1 ml of extraction solution (80:20 methanol:water), prechilled at -80°C. After 15 min at -80°C, cells were scraped and centrifuged for 10 min, 2000 g at 4°C to pellet

cell debris. After collecting the supernatant, two subsequent rounds of metabolite extraction were performed. Supernatant from each sample were kept at -80°C until ready to be shipped to our collaborators' laboratory.

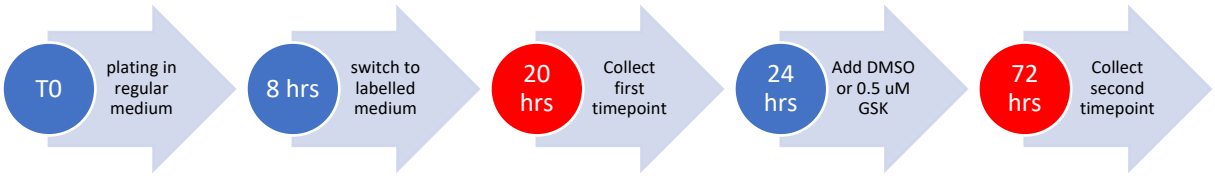


Figure 3.3 Timeline of Metabolic flux experiment.

Twenty-four hours after the initial plating, 0.5 µM GSK864 or DMSO were added to the cells. 72-hours after plating, corresponding to a 48-hour incubation with the drug, the second timepoint was collected as described above.

3.7.1 Sample preparation for Mass Spectrometry

Detailed acquisition methods for GC/MS and LC/MS runs have been uploaded into a dedicated github repository (https://github.com/SilviaRaineri/Thesis). To make up for liquid evaporation from the samples due to storage at -80°C, all methanol was evaporated using N_2 in a vacuum manifold. Samples were then frozen at -80°C for one hour and freeze-dried overnight in a Virtis benchtop K freeze dryer. The following day, lyophilized samples were resuspended in 420 µl 80:20 Methanol:Water + labelled internal standard (Glycerol). To precipitate insoluble debris, samples were incubated at -20°C for one hour, then centrifuged for 10 min, 4°C, at maximum g in a Hereaus Megafuge 40R centrifuge. For our liquid chromatography followed by mass spectrometry (LC/MS), 90 µl of sample were taken and stored at -80°C. An aliquot of 290 µl of each sample was taken for gas chromatography followed by mass spectrometry (GC/MS), then stored at -20°C. For each experiment, a quality control, made up of 10 µl from every sample, and a blank, composed of 290 µl 8:2 MetOH:Water + labelled glycerol, were prepared.

Metabolite peaks from GC/MS were extracted using the Isoscan package in R (https://github.com/jcapelladesto/isoSCAN, Capellades et al., manuscript in preparation), then each peak was normalized to the peak of the internal standard (labelled Glycerol). LC/MS metabolite peaks were manually integrated using MassHunter software (Agilent), obtaining an area under curve (AUC) value for each peak. Further exploratory and statistical analyses were carried out using R/Bioconductor packages. Total levels of each metabolite were calculated by summation of AUC values for each isotopologue, then statistical significance was assessed with ANOVA (~ Genotype * Treatment), followed by Tukey's post-hoc test.

GC/MS

Levels of intermediates from the tricarboxylic acid cycle, glycolysis and pentose phosphate pathway were analyzed with GC/MS. Cell lysates or Media samples were dried out using N_2 in a vacuum manifold for approximately 30 min, then stored at -80°C. Prior to injection on the instrument, all samples were derivatized as follows. Samples and 6 mg Metoxiamine (Sigma) were dried for one hour in the freeze dryer. Afterwards, Pyridine (Sigma) up to 2 ml was added into Metoxiamine, then 40 µl of this solution were pipetted into the samples. After a brief incubation (5 min, 40°C, 250 rpm in shaker (Environmental Shaker Incubator ES20, Biosan)), all samples were heated for 45 min 60°C in an oven, then 25 µl TMS (Sigma) were added to the mix. Finally, 65 µl of derivatized samples were inserted in glass vials and stored at room temperature until injection. GC/MS was performed on an Agilent 7000 QQQ instrument, using a Agilent J&W GC HP5MS column (Agilent). Detailed acquisition method for this run have been uploaded into a dedicated github repository (https://github.com/SilviaRaineri/Thesis). Metabolite peaks were extracted using the Isoscan package in R (https://github.com/jcapelladesto/isoSCAN, Capellades et al., manuscript in preparation), then each peak was normalized to the peak of the internal standard (labelled Glycerol). Further statistical analysis was carried out using R/Bioconductor.

Metabolite	m/z	Retention Time (min)
Pyruvate	190	3'48
Lactate	235	3'56
Fumarate	261	6'76
Succinate	263	6'38
Isoleucine	276	5'93
Leucine	276	6'19
Glycine	292	6'33
Glycerol (labelled)	312	5'98
Serine	322	7'01
Aspartate	350	8'83
Malate	351	8'48
Tryptophan	355	14'01
Glutamate	364	10'167
Phosphoenolpyruvate	385	9'75
Aconitate	391	11'18
Glyceraldehyde-3P	416	11'004
DiOHAcetone-P	416	11'20
Citrate	481	11'86
Glucose	570	12'51
Ribose-5P	620	13'76
2-HG	365	9'44
Fructose-6P	722	14'74
Glucose-6P	722	14'86
Glutamine	363	11'52

Table 9. Metabolites monitored in GC/MS experiment. *Metabolites detected in our GC/MS experiment included panel of intermediate from Glycolysis, TCA cycle, Pentose phosphate pathway and selected aminoacids. Molecules were monitored at the m/z stated above.*

LC/MS/MS

Due to their molecular size, the levels of some metabolites of interest were detected with LC/MS/MS with electrospray ionization source (ESI), using a Poroshell 120 HILICZ-P InfinityLab column (Agilent) on a Ultra High Performace Liquid Chromatography system (Agilent 1290 Infinity UHPLC) connected to a Agilent 6490 QQQ LC/MS mass spectrometer.

The experiment was divided into two different runs: one for unlabeled α-Ketoglutarate, NADP, NADPH, NAD, NADH, Acetyl-CoA, ATP, AMP, Methionine, Homocysteine and Isocitrate, and one for labelled Acetyl-CoA and SAM.

Compound	Transition	Collision Energy	Retention Time (min)	Polarity
NADP_1	744 → 604	12	7'7	Positive
NADP_2	744 → 622	12	7'7	Positive
NADPH_1	746 → 729	15	7'5	Positive
NADPH_2	746 → 302	30	7'5	Positive
NAD_1	664 → 428	24	6'86	Positive
NAD_2	664 → 136	60	6'86	Positive
NADH_1	666 → 137	52	6'86	Positive
NADH_2	666 → 136	40	6'86	Positive
SAH_1	385 → 250	60	6'28	Positive
SAH_2	385 → 136	24	6'28	Positive
SAH_3	385 → 88	4	6'28	Positive
Methionine_1	150 → 104	12	5'53	Positive
Methionine_2	150 → 56	6	5'53	Positive
Homocysteine_1	136 → 90	8	6'95	Positive
Homocysteine_2	136 → 56	16	6'95	Positive
α–Ketoglutarate_1	145 → 101	4	6'87	Negative
α–Ketoglutarate_2	145 → 57	8	6'87	Negative
Acetyl_CoA_3	810 → 428	25	6'98	Positive
Acetyl_CoA_1	810→303	36	6'98	Positive
Acetyl_CoA_2	810→136	60	6'98	Positive
ATP_1	508→136	44	7'33	Positive
ATP_2	508→97	40	7'33	Positive
SAM_3	399→298	4	7'29	Positive
SAM_1	399→250	12	7'29	Positive
SAM_2	399→97	32	7'29	Positive
AMP_1	348→136	15	6'82	Positive
AMP_2	348→97	30	6'82	Positive
Isocitrate_1	191→117	10	7'55	Negative
Isocitrate_2	191→73	20	7'55	Negative

Table 10. Transitions of unlabelled metabolites monitored in LC/MS/MS. *Levels of unlabelled metabolites were monitored through the transitions described above. Detailed acquisition method for*

this run have been uploaded into a dedicated github repository (https://github.com/SilviaRaineri/Thesis). Ionization source parameters for electrospray ionization prior to sample injection are as follows: gas temperature 270°C, gas flow 15 l/min; nebulizer 35 psi; sheath gas heater 400; sheath gas flow 11; capillary 3000V; V charging 1000 (positive polarity) or 1500 (negative polarity).

The instrument was kept at 30°C for the duration of the linear gradient. Mobile phase was composed of 994 ml Water + 6.6 ml Ammonium Acetate 7.5M + 1 ml Agilent Additive (Medronic Acid). Standards were used to establish fragmentation patterns for all metabolites except labelled SAM and AcetylCoA, measured as described in Chapter 4, sections 4.6 and 4.7. Each sample was injected once, with QC samples composed of 10 µl from all samples injected once every five samples.

Detailed acquisition method for labelled and unlabeled samples run have been uploaded into a dedicated github repository (https://github.com/SilviaRaineri/Thesis). Resulting metabolite peaks were manually integrated using MassHunter software (Agilent), obtaining an area under curve (AUC) value for each peak. Further exploratory and statistical analyses were carried out using R/Bioconductor packages. Total levels of each metabolite were calculated by summation of AUC values for each isotopologue, then statistical significance was assessed with ANOVA (~ Genotype * Treatment), followed by Tukey's post-hoc test.

3.8 Immunofluorescence on H3K9me3, H4K5ac and H3K27ac levels in WT or R132H IDH1 cells

Immunofluorescence of several histone modifications (see Table below for antibodies) was performed as described in Kraus et al, 2017 (Kraus et al., 2017). WT or R132H U87 cells were seeded at an appropriate confluency on pre-cleaned coverslips 24h prior to fixation. The following day, cells were fixed in 2% formaldehyde in PBS solution for 10 min at RT. Fixative was carefully washed by exchanging it with washing buffer (PBST; 0.02% Tween 20 in PBS), for a total of three times. Coverslips were incubated in a quenching solution (20 mM Glycine in PBS) for 15 min at RT, washed twice in washing buffer, then permeabilized in 5% (vol/vol)

Triton X-100 in PBS for 10 min at RT. Samples were blocked in 100 µl MAXblock blocking medium (Active Motif) for 30 min, followed by incubation with the appropriate primary antibody for an hour in a humidified chamber.

Table 11. Schematic of Immunofluorescence staining for WT and R132H IDH1 U87 cells. *WT or R132H cells were stained for immunofluorescence according to the combinations described below.*

Primary antibody **[1 h., RT]**	α H3K9me3 (mouse, 1:1000)	α H3K9me3 (mouse, 1:1000)	α H3K27ac Alexa 488 conjugated (mouse, 1:1000)	α H3K9me3 (mouse, 1:1000) α H4K5ac (rabbit, 1:1000)	α H3K9me3 (mouse, 1:1000) α H3K27ac Alexa 488 conjugated (mouse,1:1000 – after secondary Ab)	
Secondary antibody **[1 hr, RT]**	D α mouse Alexa488 (1:500)	D α mouse Alexa594 (1:500)		D α mouse Alexa488 (1:500) D α rabbit Alexa594 (1:500)	D α mouse Alexa594 (1:500)	α mouse Alexa488 (1:500)
Counterstain	DAPI	DAPI	DAPI	DAPI	DAPI	DAPI

Coverslips were washed four times in washing buffer, then incubated for an hour with the appropriate secondary antibody. After washing four times with washing buffer, a post-fixation with 4% formaldehyde in PBS was performed, incubating for 10 min at room temperature. Cell nuclei were stained with 2 µg/ml DAPI in PBST, then coverslips were mounted on slides using Vectashield mounting medium (Vector Laboratories, United Kingdom). Slides were stored at 4°C until ready for image acquisition using the 3D-SIM method, in collaboration with Dr. Lothar Schermelleh.

Table 12. List of antibodies used in immunofluorescence sample preparation for 3D-SIM

Antibody	Catalog number
α H3K9me3	Ab7604, Abcam
α H3K27ac Alexa 488 conjugated	07-449-AF488, Millipore
α H4K5ac	04-688, Millipore
D α mouse Alexa488	A32766, Thermofisher
D α rabbit Alexa594	A-21207, Thermofisher

3.9 Bioinformatics analyses

Results tables from the main bioinformatics analyses have been uploaded on a dedicated github repository at the following URL: https://github.com/SilviaRaineri/Thesis .

4. Metabolic characterization of U87 WT and R132H cells

4.1 Introduction[1]

Metabolism and epigenetics are highly interconnected. Several proteins involved in metabolic pathways also participate to chromatin remodelling and gene regulation by producing co-factors or substrates used by epigenetic writers (Wellen et al., 2009b). One such example is Isocitrate Dehydrogenase (IDH). IDH enzymes metabolize Isocitrate to α-Ketoglutarate, either in the mitochondrion as a step of the Krebs cycle (IDH2) or in the cytoplasm (IDH1) (M. Gagné, Boulay, Topisirovic, Huot, & Mallette, 2017b). α-Ketoglutarate produced by this reaction serves as a co-factor for several α-Ketoglutarate-dependent dioxygenases, notably the ten-eleven translocation (TET) family of DNA demethylases and the Jumonji (Jmj) family of histone demethylases (Tahiliani et al., 2009; Tsukada et al., 2006). Interestingly, *IDH* mutations are common in several types of cancers, including ~80% of glioblastomas, ~40% chondrosarcomas, 20% of acute myeloid leukemias (AML), sinonasal undifferentiated carcinoma and prostate cancer (Abeshouse et al., 2015; Adam, Yang, Soga, & Pollard, 2014; Amary et al., 2011; X. Liu et al., 2013; Pansuriya et al., 2011). These heterozygous mutations can be found in substrate binding residues of both IDH1 (R132H) and IDH2 (R140Q, R172K) (Yan et al., 2009). While IDH1 mutations are more common in gliomas (80%) and AML (20%), IDH2 mutations occur more frequently in AML (20%) and cholangiosarcomas (20%) (Mondesir, Willekens, Touat, & de Botton, 2016). Both mutations have been associated with a relatively better prognosis and induce a gain of function that causes further processing of α-Ketoglutarate into 2-Hydroxy-Glutarate (2HG) (Dang et al., 2009), an oncometabolite linked with tumour progression (Weller, Wick, & von Deimling, 2011). Given the large number of studies on these mutations and their impact on cancer progression, several targeted inhibitors of the mutant form of IDH1 or IDH2 have been developed and have now reached the clinical trial stage (DiNardo et al., 2016; Popovici-Muller et al., 2018; Yen et al., 2018, 2017). Despite

the many similarities between mutations in IDH1 and IDH2, our focus will primarily be on IDH1.

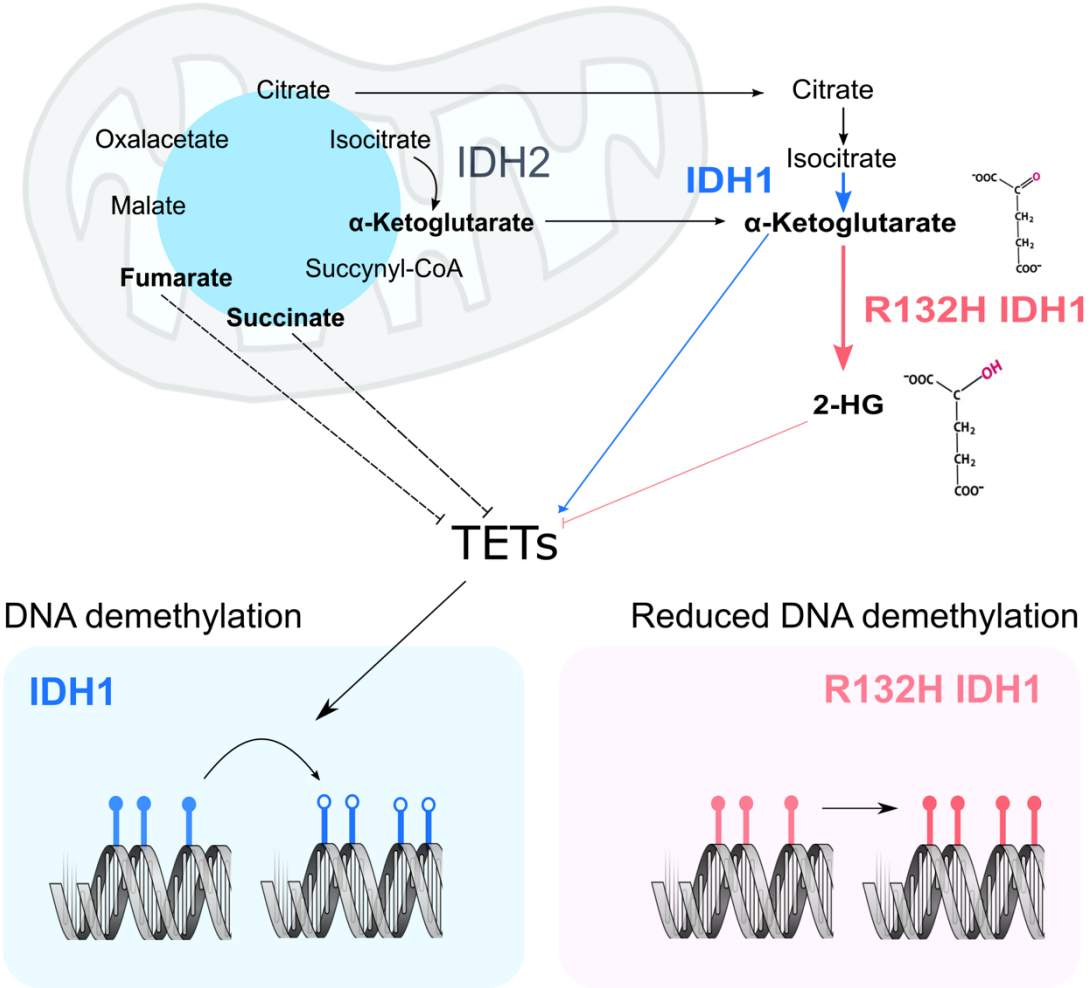


Figure 4.1.Current model of the impact of IDH1 mutation on chromatin remodelling*. In a WT cell, IDH1/2 (blue and black) metabolise Isocitrate into α-Ketoglutarate. Upon mutation of IDH1 (pink panel), α-Ketoglutarate is processed into 2HG. In other cancer settings, mutations in Fumarate hydratase and Succinate dehydrogenase result in the accumulation of Fumarate and Succinate, which may inhibit TETs. α-Ketoglutarate acts as co-factor in the nucleus for both DNA demethylases (TETs) and histone demethylases (KDMs). Their activity ensures the correct levels of DNA (bottom panel) and protein methylation in the cell. More specifically, TETs act on methylated DNA sequences (filled lollipops), starting a reaction chain that will ultimately lead to methyl group removal (empty lollipops). Image from Raineri and Mellor (2018).*

According to the current model, the structural similarity between 2HG and α-Ketoglutarate causes inhibition of both histone and DNA demethylases, inducing an increased methylated state in the nucleus which leads to gene expression deregulation and promotes cancer

development (Xu et al., 2011) (Figure 4.1). This inhibition is also achieved upon mutations in other enzymes of the tricarboxylic acid cycle, mainly *Fumarate hydratase (FH)* and *Succinate dehydrogenase (SDH)* (Xiao et al., 2012). These loss of function mutations induce excessive accumulation of their respective substrates, Fumarate and Succinate, which then act as competitive inhibitors of α-Ketoglutarate-dependent dioxygenases. This model suggests a metabolic basis for the changes observed in chromatin as a result of the *IDH1* mutations. Despite a growing body of evidence, the exact molecular mechanism and consequences of 2HG production are still largely unknown.

In this study, we aim at understanding the impact of a mutation in the metabolic enzyme IDH1 on the epigenetic state of the cell and from this infer how it might impact on tumorigenesis. To achieve this, we will dissect each step of the current model for IDH1 mutation-driven cancer development (Figure 4.1), by examining metabolic, epigenetic and gene expression differences between U87 WT or R132H IDH1 human glioblastoma cells. To understand whether rescuing levels of 2HG, the primary metabolic defect, will be sufficient to correct the epigenetic or gene expression alterations, we will treat cells with an IDH1 inhibitor (GSK864), reportedly showing high affinity for the mutated copy of IDH1 (Okoye-Okafor et al., 2015). In this chapter, the metabolic changes are explored.

4.2 Discrepancies between literature studies on the metabolic alterations in IDH1 mutant cells

IDH1 is a key metabolic enzyme and many efforts have been put into describing not only its epigenetic consequences, but the metabolic alterations caused. A plethora of studies have used different techniques, ranging from proton-magnetic resonance spectroscopy (H-MRS), liquid or gas chromatography followed by mass spectrometry (LC or GC/MS) or isotope tracing for metabolic flux to describe the metabolic alterations upon IDH1 mutation, sometimes with contrasting results (Dang et al., 2009; Izquierdo-Garcia et al., 2014; Izquierdo-Garcia,

Viswanath, Eriksson, Cai, et al., 2015; McBrayer et al., 2018; Reitman et al., 2011; Salamanca-Cardona et al., 2017).

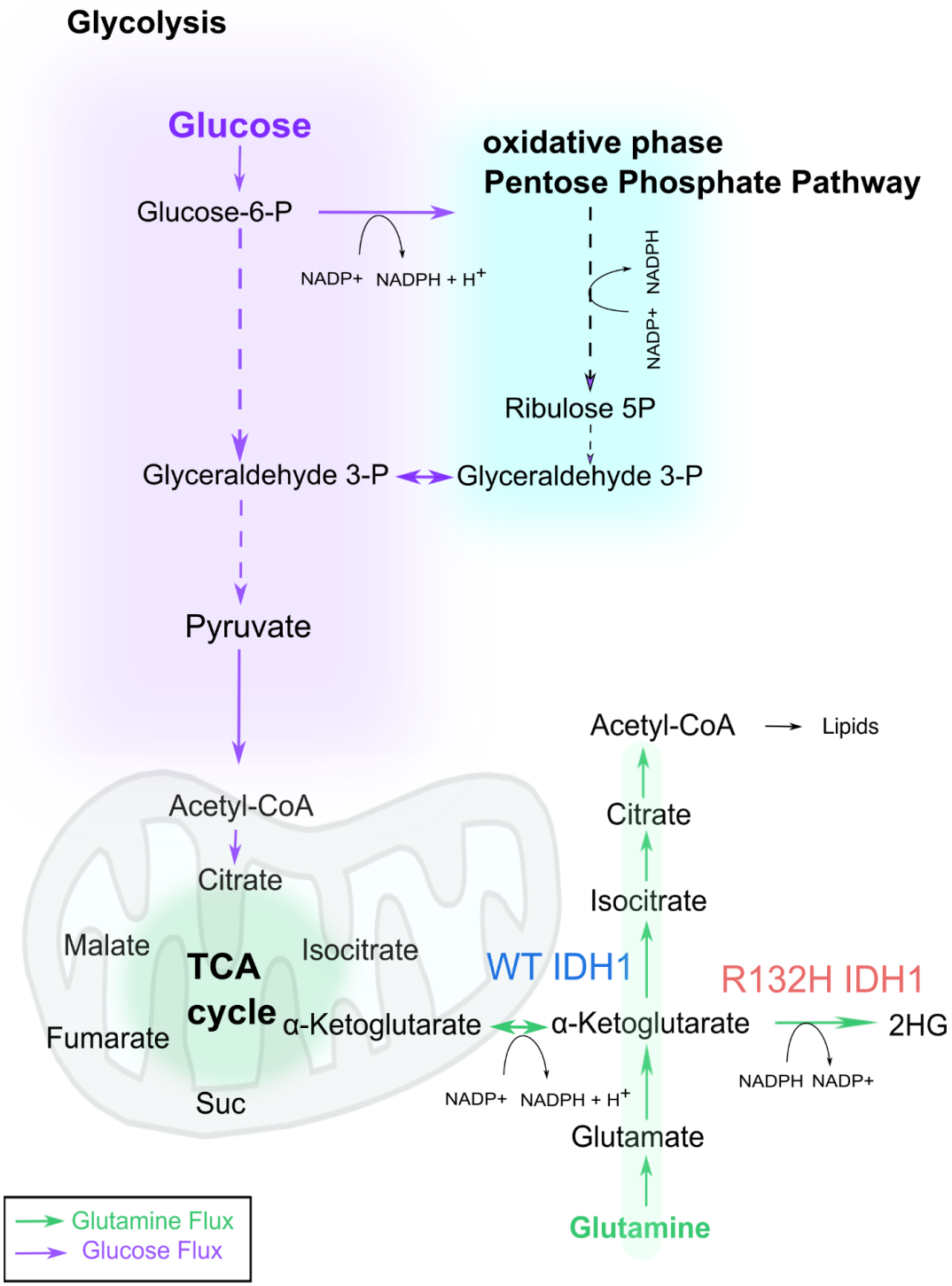


Figure 4.2. Diagram of main metabolic pathways potentially affected by IDH1 mutation. *Literature studies described metabolic alterations, albeit with some discrepancies, mainly in the TCA cycle (green) and Glutaminolysis (green arrows), the conversion of Glutamine into Glutamate, then α-Ketoglutarate (Ohka et al., 2014a; Salamanca-Cardona et al., 2017). Furthermore, some studies have suggested an activation of the Pentose Phosphate Pathway (light blue), as a way for the cell to produce more NADPH, following its increased consumption in the IDH1 R132H-mediated transformation of α-Ketoglutarate into 2HG (Badur et al., 2018; Gelman et al., 2018). Whether this mutation impacts Glycolysis is less clear, although studies have suggested impairments in Pyruvate*

metabolism (Izquierdo-Garcia, Viswanath, Eriksson, Cai, et al., 2015; Izquierdo-Garcia, Viswanath, Eriksson, Chaumeil, et al., 2015)

All studies describe the increase in 2HG levels and a decrease of Glutamate, generally coupled with an increase in Glutamine levels, except for one study, reporting its consistent decrease (Ohka et al., 2014a). Levels of tricarboxylic cycle (TCA) intermediates, such as α-Ketoglutarate, Acetyl-CoA, Fumarate, Malate or Isocitrate are reported to be decreased (Dang et al., 2009) or unchanged (Ohka et al., 2014b; Reitman et al., 2011). These results might be misleading as they only provide a static picture of the metabolome. Low or unchanged levels of these intermediates could actually just reflect the fact that the TCA cycle is active in these cells.

The current model proposes that the conversion of Isocitrate to α-Ketoglutarate promoted by IDH enzymes concomitantly reduces NADP+ (IDH1 and 2) or NAD+ (IDH3) to NADPH or NADH. Upon mutation, R132H IDH1 would further use NADPH to fuel the conversion of α-Ketoglutarate to 2HG, disrupting NADP/NADPH homeostasis (Badur et al., 2018).

Whether this mutation also impacts on Glycolysis is less clear. Ohka et al. describe unchanged levels of Glycolysis intermediates in a global metabolome analysis of 30 patients-derived gliomas, 13 with WT IDH1 and 20 with mutated IDH1 (Ohka et al., 2014a). However, other studies in engineered glioma cells (U87) or astrocytes (NHA) have found a reduction in pyruvate metabolism (Izquierdo-Garcia, Viswanath, Eriksson, Cai, et al., 2015; Izquierdo-Garcia, Viswanath, Eriksson, Chaumeil, et al., 2015).

Despite the growing interest in the characterization of metabolic alterations following IDH1 mutation, the variety of methods and cellular models used in different studies makes it difficult to define common patterns, causing discrepancies between reports. Nevertheless, dissecting the metabolism of different tumours bearing the IDH1 mutation provides important information to better understand the impact of metabolism on the many proposed epigenetic consequences of this mutation. Using metabolic flux analysis provides the most rigorous method, as besides

quantifying the levels of each metabolite of interest, it provides insight into what energy source is being used by the cells, and what metabolites it is producing. Indeed, using hyperpolarized magnetic resonance imaging to follow labelled glutamine (HP [1-^{13}C]Glutamine), Salamanca-Cardona et al. describe an activation of the TCA cycle and Glutamine dependence in patient-derived chondrosarcoma cells bearing mutations in *IDH1* or *IDH2* (Salamanca-Cardona et al., 2017). Furthermore, metabolic flux in colorectal cancer cells (HCT116) with or without R132H IDH1 mutation has described the activation of the oxidative pentose phosphate pathway (oxPPP) in cells bearing the IDH1 mutation as a way to restore the correct NADP/NADPH balance (Gelman et al., 2018) (Figure 4.2).

Given IDH1's central role in cell metabolism, we first set out to characterize metabolic differences between U87 WT and IDH1 mutated R132H cell lines in a metabolic flux experiment. This entails adding 13C-labelled Glucose (bearing the heavier 13C on all its six carbons, 6C) or Glutamine (fully labelled on all five carbons, 5C) to the cell culture media, so that by measuring the proportion of each labelled metabolite we are able to track how these cells are using the carbons derived from either nutrient. An advantage of using this technique over a regular metabolomics experiment is that on top of being able to measure the level of each metabolite of interest, we can also understand what the primary resource of energy for these cell lines is. This is achieved by measuring levels of each metabolite's isotopologues, molecules identical to the original metabolite in their chemical formula and arrangement of atoms, but differing in weight due to the presence of one or more 13C carbons derived from the labelled nutrient added to the medium. These results were obtained while on secondment in Dr. Oscar Yanes's lab (Universitat Rovira I Virgili, Reus, Spain).

4.3 Metabolic Flux Optimization: Glucose and Glutamine Uptake

Design of a metabolic flux experiment requires choosing an appropriate timepoint for cell harvesting after the addition of labelled media, in order to maximise the amount of labelling in the secondary metabolites and to follow for as long as possible the labelled carbons derived from the original nutrients. We first examined the uptake levels of Glucose and Glutamine in U87 WT or IDH1 R132H cells (from here on referred to as R132H cells), treated with the mutant IDH1 inhibitor GSK864 or with DMSO (Ctrl) over time (Figure 4.3).

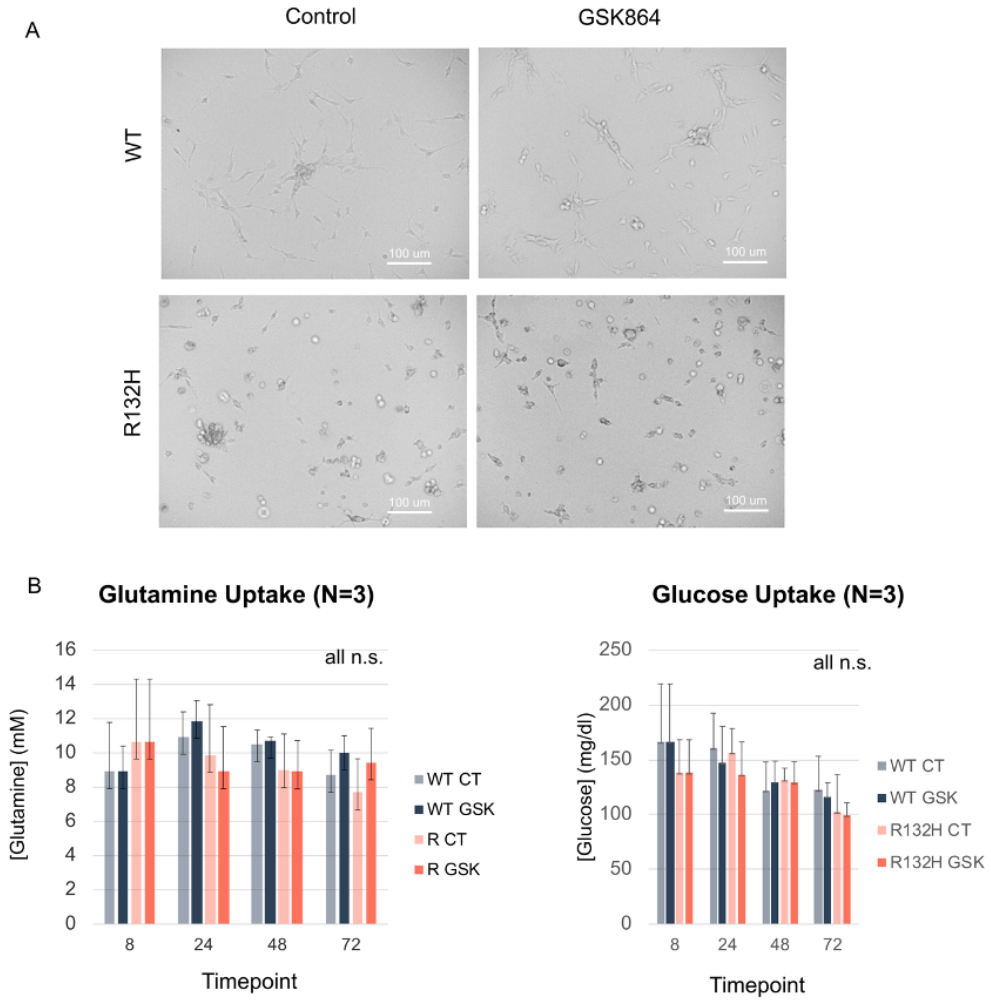


Figure 4.3 Glucose and Glutamine uptake unchanged in WT vs R132H IDH1 cells. *A) Bright field microscopy images of U87 WT or R132H treated with GSK864 or untreated. Images were taken at the end of a 48 hours treatment with a Biorad ZOE imager. Despite the low quality of the images, GSK864 treatment lack of effect on cell viability or morphology has been confirmed in following experiments: indeed, when harvesting for RNASeq, ChIP-Seq and metabolic flux, both cell lines had an equal number of cells. B) Barplots showing Glutamine (left panel) or Glucose (right panel) concentration in the medium of samples over time, as measured by fluorescence (Glutamine) or colorimetric (Glucose) assay.*

Error bars represent standard deviation in N=3 biological repeats. Statistical significance was assessed with ANOVA, but none of the comparisons were significantly different.

GSK864 treatment did not seem to impact on cell morphology or viability (Figure 4.3A). In both cases, the first differences in nutrient uptake are seen 24 hours after seeding, although there is a high variability amongst biological replicates (Figure 4.3B). Glucose uptake remains somewhat comparable between the two cell lines, progressively decreasing over time and showing the greatest difference after 72 hours from seeding. Glutamine assay results suggest some differences between WT and IDH1 mutated cells, albeit not statistically significant. GSK864 treatment doesn't seem to impact Glucose uptake, but might affect to a greater extent Glutamine, especially after 72 hours (which corresponds to a 48 hour treatment, as described in the Methods section). In terms of biological interpretation, little can be stated from these preliminary experiments: these colorimetric assays are often imprecise in their measurements, and are mainly testing the action of Glucose and Glutamine transporters, not the actual cellular composition of metabolites. In spite of the little biological significance, these assays were still useful in guiding our metabolic flux experimental design, as we choose 24 hours and 72 hours as harvesting time points.

4.4 GSK864 reduces levels of Glutamine-derived 2HG in R132H cells

The direct consequence of IDH1 R132H mutation is an increase in production of 2-hydroxyglutarate (2HG), an oncometabolite derived from the further processing of α-Ketoglutarate by mutated IDH1. Indeed, U87 R132H Control cells show a significant (*** $p<0.001$) increase in 2HG levels compared to the WT, and a 48 hours treatment with GSK864 successfully brings its levels down to the WT baseline (Figure 4.4).

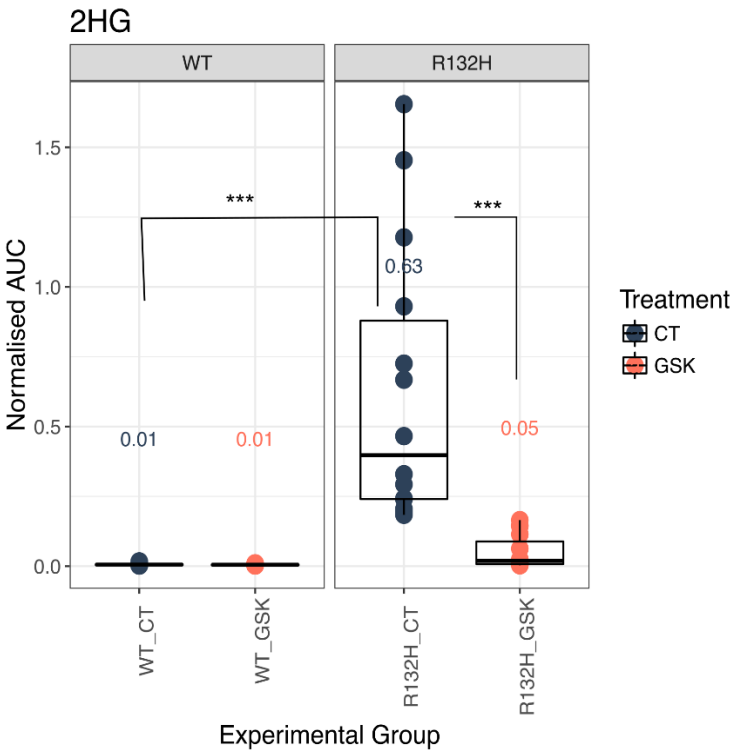


Figure 4.4 Treatment with GSK864 reduces 2HG levels in R132H cells. *Boxplots of the area under curve (AUC) value for 2HG levels in WT (left panel) or IDH1 R132H cells (right panel), treated with 0.5 μM GSK864 (here referred to as GSK, orange), or DMSO (Ctrl), as measured by GC/MS. Total levels of 2HG were calculated by summing up the AUC values of each isotopologue, as described in the methods section (3.7.1). The average AUC value is printed on top of each boxplot. Statistical significance was assessed by ANOVA, followed by Tukey's post-hoc test. Significance levels = ns P > 0.05; * P < 0.05; ** P < 0.01; *** P < 0.001. (N=10 technical replicates per group).*

Metabolic flux analysis reveals that the majority of labelled 2HG derives directly from Glutamine, as the most abundant labelled isotopologue is M+5, corresponding to the Mass of 2HG + 5 13C-labelled carbons (Figure 4.5, pink bar). While the M+5 isotopologue can be seen from the 20 hour timepoint onwards, others emerge in smaller proportions at 72 hours, such as M+3, derived from different rounds of the TCA cycle (Figure 4.5, light blue and violet bars).

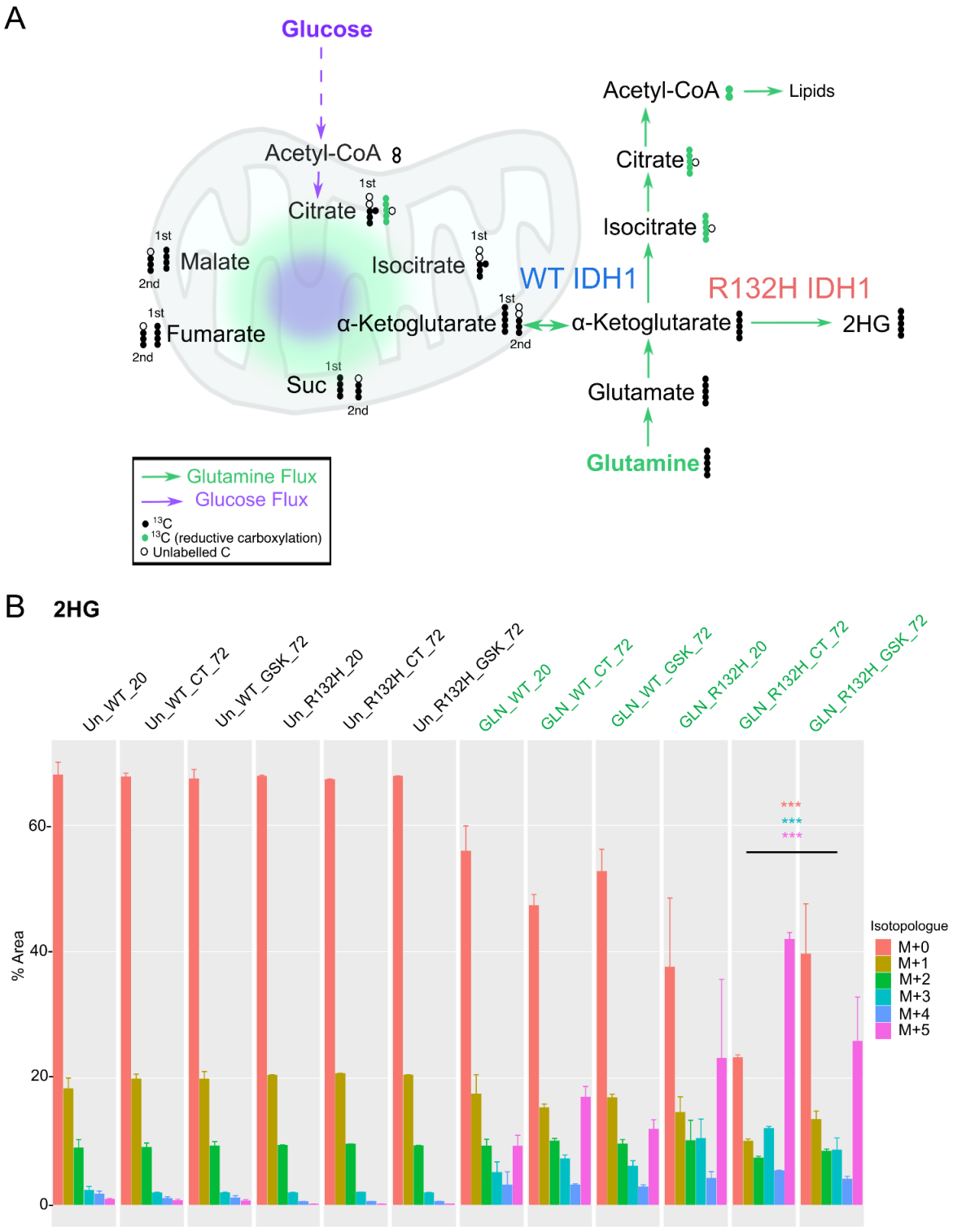


Figure 4.5 Glutamine labelling patterns of 2HG. *A) Diagram showing the labelling patterns of 2HG, Glutamine and main TCA cycle intermediates. Green arrows indicate the Glutamine flux, which can either feed into the TCA cycle (green circle) or the cytosolic production of Isocitrate via reductive carboxylation. B) Bar plots showing the area under curve of GC/MS spectra for Glutamine-labelled (GLN, green) or unlabeled (Un, black) samples, expressed in percentage. Compared to the unlabeled panel, Glutamine labelling progressively increases over time (20 or 72 hours) for both WT and R132H samples, although 2HG absolute levels are greatly increased only in R132H cells (see Figure 4.4). Notably, isotopologue M+5 represents a 2HG molecule whose five carbons are derived directly from Glutamine. Levels of isotopologues M+5, M+3 and M+0 are all significantly different in R132H cells treated with GSK compared to their untreated control. Statistical significance between samples at 72*

*hours was assessed by ANOVA, followed by Tukey's post-hoc test. Significance levels = ns P > 0.05; * P < 0.05; ** P < 0.01; *** P < 0.001. (N=5 per group). Stars color corresponds to the statistically different isotopologue,*

4.5 Methionine cycle-related metabolites levels remain unchanged in IDH1 mutant cells

2HG is thought to affect DNA and histone methylation levels by inhibiting α-Ketoglutarate dependent dioxygenases, a large family of proteins which comprises TETs and KDMs demethylates. The main methyl donor in the cell is S-Adenosyl Methionine (SAM), whose methyl group is donated to acceptor molecules upon its conversion to S-adenosyl Homocysteine (SAH). This is processed into Homocysteine, which is then remethylated to form Methionine, completing the Methionine cycle (Figure 4.6).

Thus, we choose to monitor SAM, SAH, methionine and homocysteine. While data collected for SAH and Homocysteine did not pass quality control, we were able to detect levels of Methionine and labelled SAM. Quantification of both metabolites shows that neither are significantly different in WT compared to R132H cells, or in control compared to GSK864 treated cells.

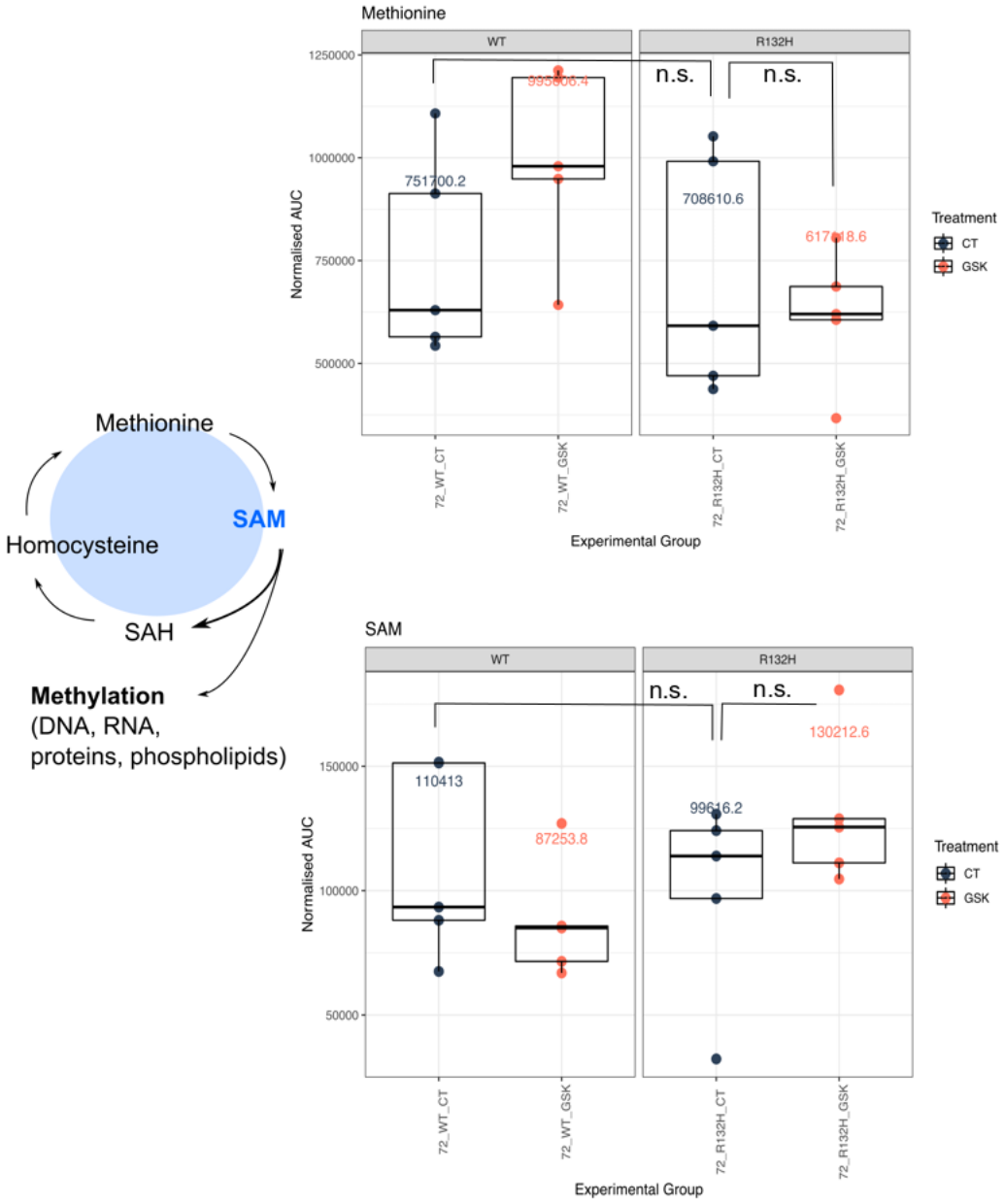


Figure 4.6 Intermediates of the Methionine cycle show no change between WT or R132H cells*. Left-side panel: diagram of Methionine cycle, depicting the conversion of Methionine to SAM by Methionine Adenosyltransferase. During substrate methylation, SAM is converted to SAH, whilst the methyl group is transferred to its acceptor, ranging from DNA and RNA to proteins, such as histones, or phospholipids. SAH hydrolase then catalyzes the hydrolysis of SAH into homocysteine, which is finally converted back to Methionine by addition of a methyl group (Sanderson et al., 2019). Right-side panel: boxplots showing Methionine levels (top) or SAM levels (bottom) in WT compared to R132H cells. Average AUC value for each group is represented on top of the boxplot. Total levels of each metabolite were calculated by summing up the AUC of all the isotopologues, as described in Chapter 3. ANOVA test revealed no significant differences either in genotype or before and after GSK864 treatment (N=10 technical replicates per group).*

Detection of labelled SAM proved to be challenging due to the molecule's high molecular weight. As detailed in Figure 4.7, SAM is composed of Methionine, Ribose and Adenine (the last two molecules together forming Adenosine). We assumed the molecule could be

fragmented in three points: the bond between Adenine and Ribose, the one between Ribose and Sulphur from Methionine, and the bond between Sulphur and the Carbon from Methionine. This would result in three main fragments: one containing Ribose alone, one composed of Ribose, Adenine and the Methyl group (referred to as Adenosine-Methyl) and Adenosine alone.

We then predicted that out of all the carbons in SAM, those that could end up labelled are two Carbons from Adenine (A), the 5-Carbon ring from Ribose (B) and the methyl group from Methionine (C). We then calculated all the possible combinations of A, B and C fragments, which resulted in a list of transitions to monitor in our LC/MS/MS experiment (Appendix Table 27). Our analysis suggests that while the biggest proportion of fragments remain unlabelled, the main source for SAM is Glucose, rather than Glutamine. Indeed, most labelled fragments are the 5C Ribose (50% at 72 hours, dark green, Figure 4.7C) or the 5C isotopologue of the Adenosine-Methyl fragment, which still corresponds to the Ribose ring (purple, 15% at 72 hours) (Figure 4.7C).

No significant differences were detected between the two cell lines or between control and GSK864 treated cells, suggesting that in spite of the general increase in methylation, SAM levels remain unchanged. This could be explained by the fact that the methylation increase isn't due to an increased activity of methylases, but rather a defect in proteins responsible for demethylation.

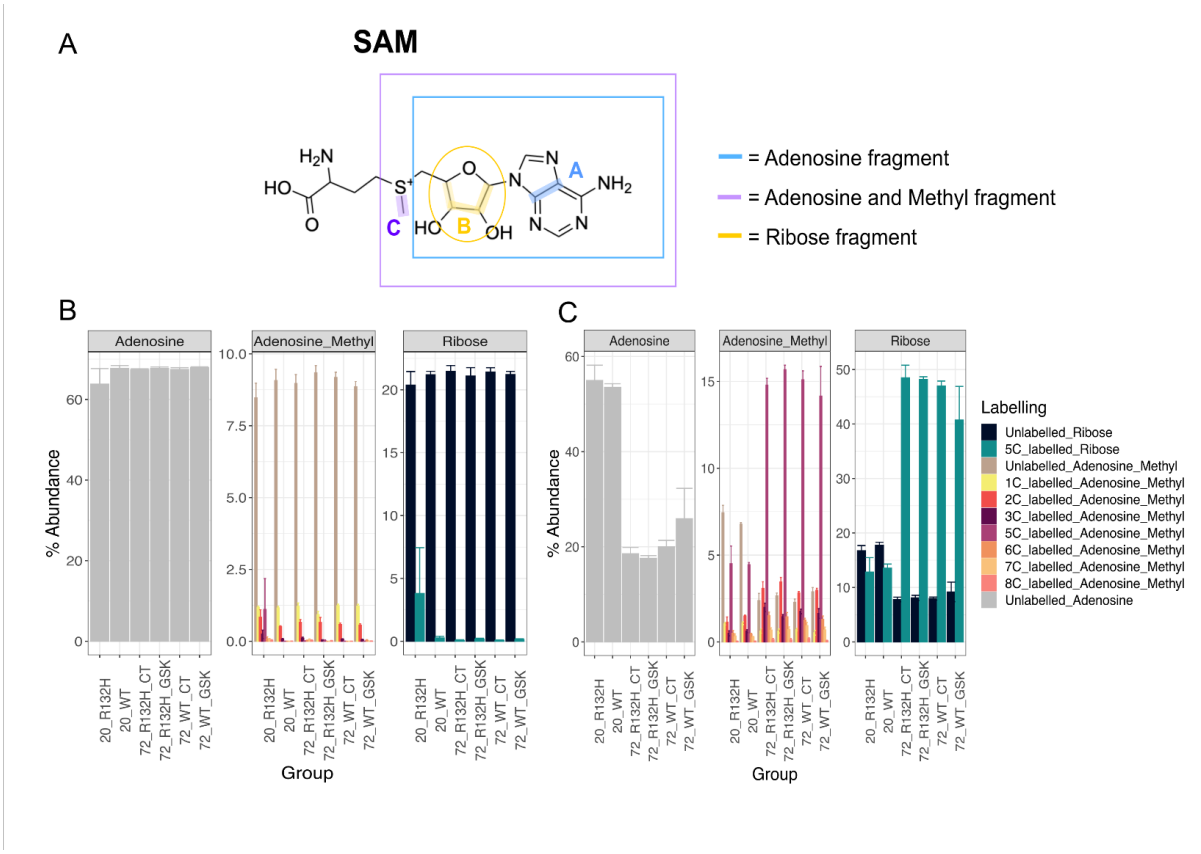


Figure 4.7 SAM labelling patterns reveal no differences between WT or R132H cells. *A) Strategy for detection of labelled SAM in LC/MS/MS. We assumed the generation of three different fragments: one containing Adenosine (light blue), one containing Ribose alone (yellow) and one composed of Adenosine and a methyl group (purple, referred to as Adenosine_Methyl). We then considered which carbons could be labelled in each fragment: up to seven carbons in the Adenosine fragment (five from Ribose, two from Adenine, B and A respectively), five in Ribose, and up to eight in Adenosine_Methyl (seven from Adenosine and one from the carbon in the methyl group). We then monitored all the possible transitions that could result from the different labelling of these fragments. Bottom: Glucose (C) or Glutamine (B) labelling patterns of SAM. Data is represented as percentage of the abundance of a specific labelled fragment. Error bars represent standard deviation of each group (N=5 data points per group). ANOVA analysis showed no significant difference between genotype, or following GSK864 treatment, hence significant levels were not annotated on the plot.*

4.6 Increased levels of TCA cycle-related metabolites in R132H cells

Given the central role of IDH1 in α-Ketoglutarate production, we expected other metabolites besides 2HG to be altered in R132H cells. Results show a significant two-fold increase in several TCA cycle intermediates, such as α-Ketoglutarate, Malate, Fumarate, Isocitrate and Acetyl-CoA (Figure 4.8).

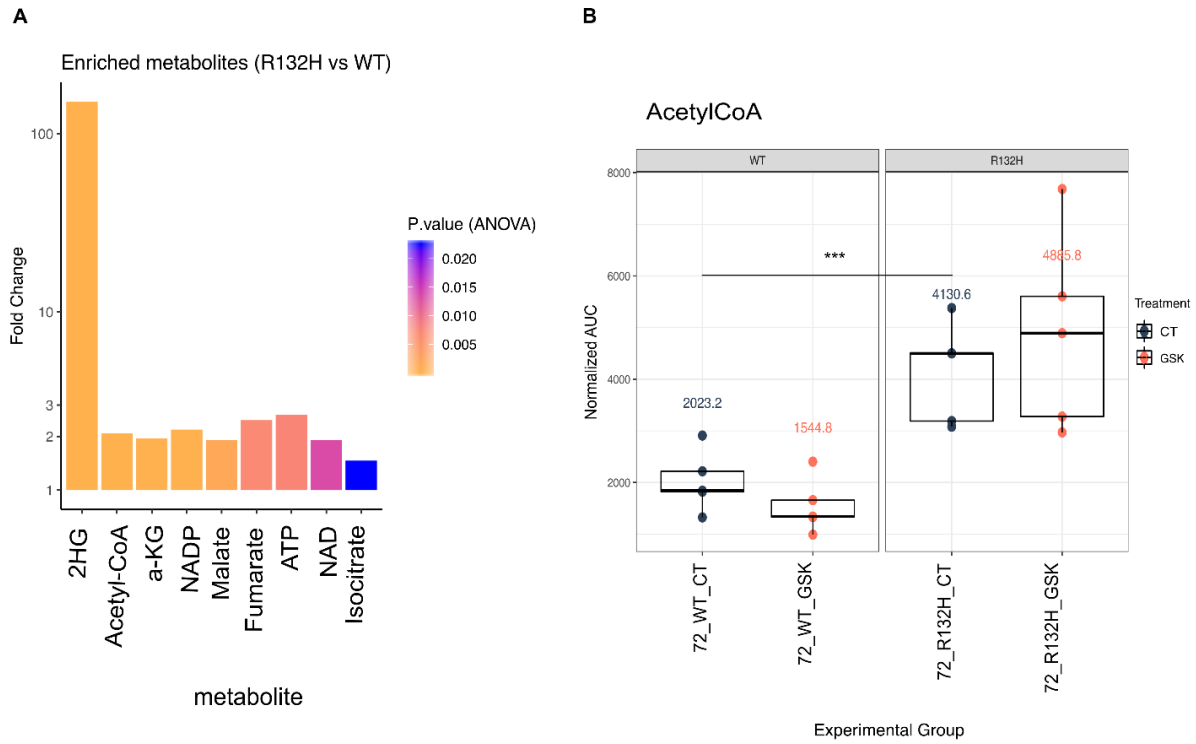


Figure 4.8 R132H IDH1 induces increase in metabolites other than 2HG. *A) Barplot showing the statistically significant fold change in R132H over WT cells of a panel of metabolites, many of which participating to the TCA cycle. Bar colors represent the P.value of each change, as calculated on the 72-hour WT Ctrl or R123H control samples via ANOVA. B) Boxplot of Acetyl-CoA total levels at 72 hours. Average normalized AUC value for each experimental group is represented on top of the boxplot. Statistical significance between samples at 72 hours was assessed by ANOVA, followed by Tukey's post-hoc test. Significance levels = ns P > 0.05; * P < 0.05; ** P < 0.01; *** P < 0.001. (N=10 per group).*

The IDH1-catalysed reduction of α-Ketoglutarate to 2HG is mediated by NADPH, and increased levels of both NADP and NAD were detected, whilst NADPH and NADH data did not pass quality control. In spite of a significant increase of these metabolites in our mutated cell line compared to its WT counterpart, treatment with GSK864 did not affect levels of any of them.

Analysis of labelled Acetyl-CoA posed the same challenges as SAM, and a similar strategy was followed. This molecule could result in three fragments: the Adenine portion by itself, which could be labelled on two carbons (A), Adenosine (Adenine and Ribose) and two phosphate groups, potentially bearing five labelled carbons on the Ribose (B), and finally, pantothenate

and beta-mercaptoethylamine, only labelled on the two carbons of the acetyl group (C) (Figure 2.8A). A table of all the possible combination of A, B and C fragments monitored is available in the Appendix (Table 26).

Glucose-derived labelling of the Acetyl group (C) is significantly increased in WT cells (Figure 4.9B). In contrast, the Acetyl group (C) shows increased Glutamine labelling in R132H cells, mirrored in a significant decrease in its unlabelled counterpart (Figure 4.9C).

As with SAM, the majority of Ribose in the Adenosine fragments presents labelling from Glucose (Figure 4.9B, light pink bars), although its proportions are the same in both WT and R132H. Thus, Acetyl-CoA levels are significantly increased upon IDH1 mutation, and the molecule presents different labelling patterns on the Acetyl group, deriving mainly from Glutamine in R132H cells and Glucose in the WT.

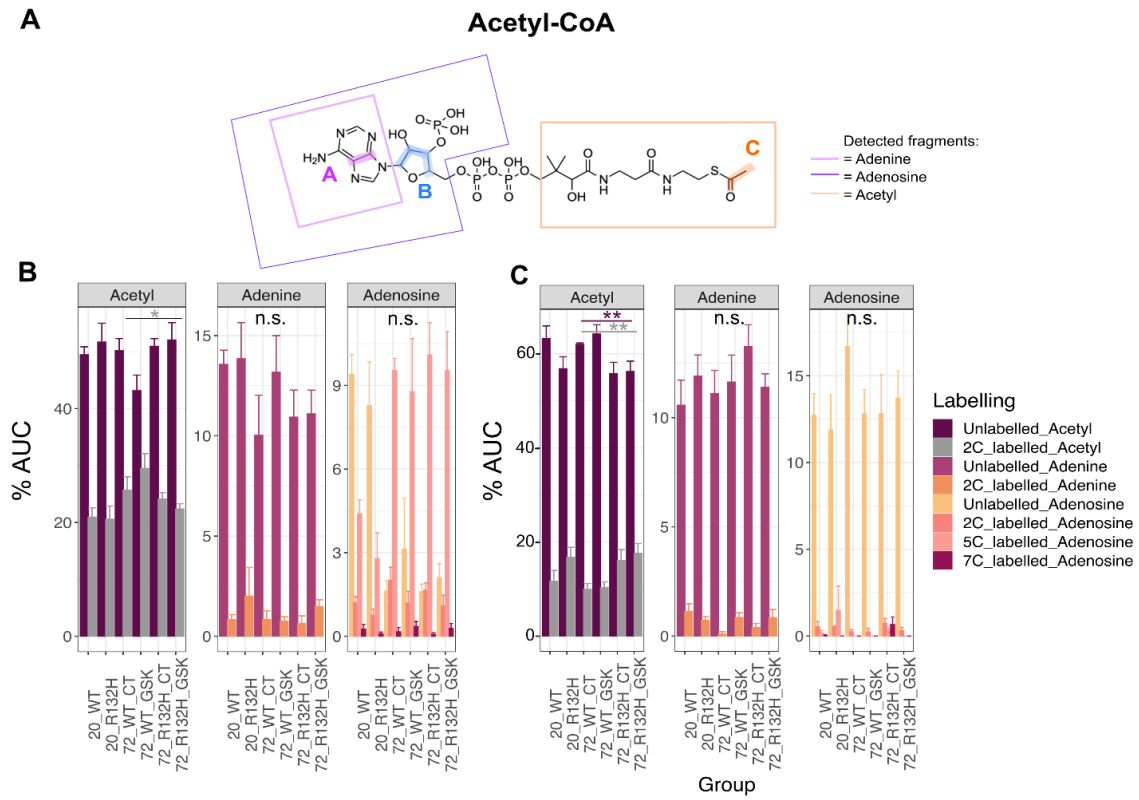


Figure 4.9 Labelling patterns of Acetyl-CoA suggest increased Glutamine-labelling of acetyl group in R132H cells. *A) Strategy for labelled Acetyl-CoA fragments detection in LC/MS/MS. Acetyl-CoA structure could generate three possible fragments, here referred to as Adenine (pink, potentially bearing two labelled carbons (A)), Adenosine (purple, showing up to seven labelled carbons (A and B)) or Acetyl (orange, only labelled on two carbons (C)). Bottom: Glucose (B) or Glutamine (C) labelling patterns.*

*Each panel represents one of the three main fragments described in Figure 4.9A, and each bar represents the percentage of the abundance of each labelling combination, by experimental group. Statistical significance was calculated by ANOVA followed by Tukey's post hoc test. Significance levels = ns P > 0.05; * P < 0.05; ** P < 0.01; *** P < 0.001. (N=5 per group). Stars color corresponds to the statistically different isotopologue.*

4.7 Glutamine fuels both TCA cycle and reductive carboxylation

We next focused on investigating how these cells use their energy. Results suggest that while both lines are Glutamine-dependent, Glutamine flux is increased in R132H cells. Indeed, labelling patterns in TCA intermediates such as Malate and Fumarate hint at a role of Glutamine in fuelling the TCA cycle through Glutaminolysis, a sequence of reactions that catalyse its conversion into Glutamate first, then α-Ketoglutarate. As shown in Figure 4.10B, Fumarate bears significantly increased labelling on the M+4 isotopologue, which derives from the first round of TCA cycle (Figure 4.10A).

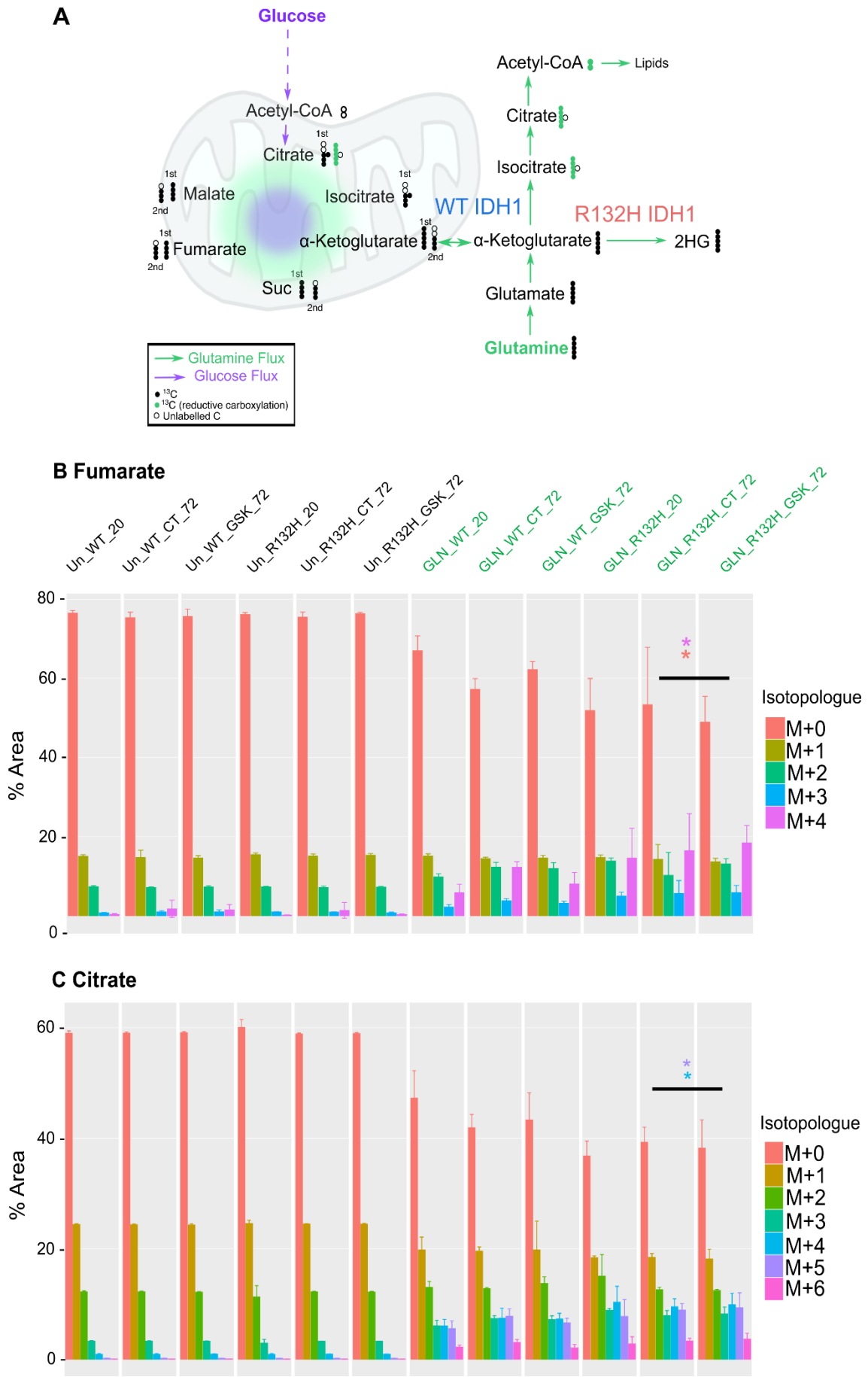

A
Glucose
Acetyl-CoA
Citrate
Malate
Isocitrate
Fumarate
α-Ketoglutarate
Suc
Acetyl-CoA
Lipids
Citrate
Isocitrate
WT IDH1
R132H IDH1
α-Ketoglutarate
2HG
Glutamate
Glutamine
Glutamine Flux
Glucose Flux
13C
13C (reductive carboxylation)
Unlabelled C
B Fumarate
Un_WT_20
Un_WT_CT_72
Un_WT_GSK_72
Un_R132H_20
Un_R132H_CT_72
Un_R132H_GSK_72
GLN_WT_20
GLN_WT_CT_72
GLN_WT_GSK_72
GLN_R132H_20
GLN_R132H_CT_72
GLN_R132H_GSK_72
% Area
Isotopologue
M+0
M+1
M+2
M+3
M+4
C Citrate
% Area
Isotopologue
M+0
M+1
M+2
M+3
M+4
M+5
M+6

Figure 4.10 Glutamine labelling patterns of WT and R132H cells reveal Glutamine dependency and activation of reductive carboxylation in R132H cells. *A) Diagram of labelling patterns for reductive carboxylation (green circles), glutaminolysis and TCA cycle intermediates (black circles). Fully labelled Glutamine (5C) is introduced in the cell and converted to Glutamate, then α-Ketoglutarate (Glutaminolysis). α-Ketoglutarate can subsequently enter the TCA cycle in mitochondria or be processed into Isocitrate/Citrate in the cytosol, in a reaction mediated by IDH1. B) Labelling patterns for Fumarate confirm Glutamine as Energy source for TCA cycle. A statistically significant increase in its M+4 isotopologue corresponds to Fumarate production in a first round of TCA cycle. C) Labelling patterns of Isocitrate suggests Glutamine is fueling both the TCA cycle (isotopologue M+4, light-blue bar) and reductive carboxylation (M+5, purple bar). Barplots represent the percentage of each labelling pattern in a specific experimental group. Error bars represent standard deviation in N=5 technical repeats Statistical significance amongst 72 hours samples was calculated via ANOVA, followed by Tukey's post-hoc test. Significance levels = ns $P > 0.05$; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. (N=5 per group). Asterisk color corresponds to the statistically different isotopologue.*

Moreover, specific labelling patterns in Citrate also indicate activation of reductive carboxylation (Figure 4.10C). Indeed, a significant increase in two of its isotopologues reflects the dual presence of the TCA cycle (M+4, derived from the first round of the cycle) and reductive carboxylation (M+5, where the five labelled carbons are those converted from Glutamine). This pathway is present in cancer cells with defective mitochondria and allows cytosolic Acetyl-CoA production, starting with IDH1-mediated production of isocitrate (Mullen et al., 2012). While these isotopologues are also present in WT cells, their increase is significant only in cells bearing the IDH1 mutation.

In contrast with changes seen in Glutamine labelling, Glucose labelling bears little to no difference between WT or R132H cells. Nevertheless, presence of labelled Fumarate (M+2, green, Figure 4.11B) suggests that Glucose is still feeding the TCA cycle, although to a lesser extent compared to Glutamine. Furthermore, as already found in labelling patterns of Acetyl-CoA and SAM, Ribose 5-Phosphate is, as expected, derived from Glucose (M+5, pink bar, Figure 4.11D).

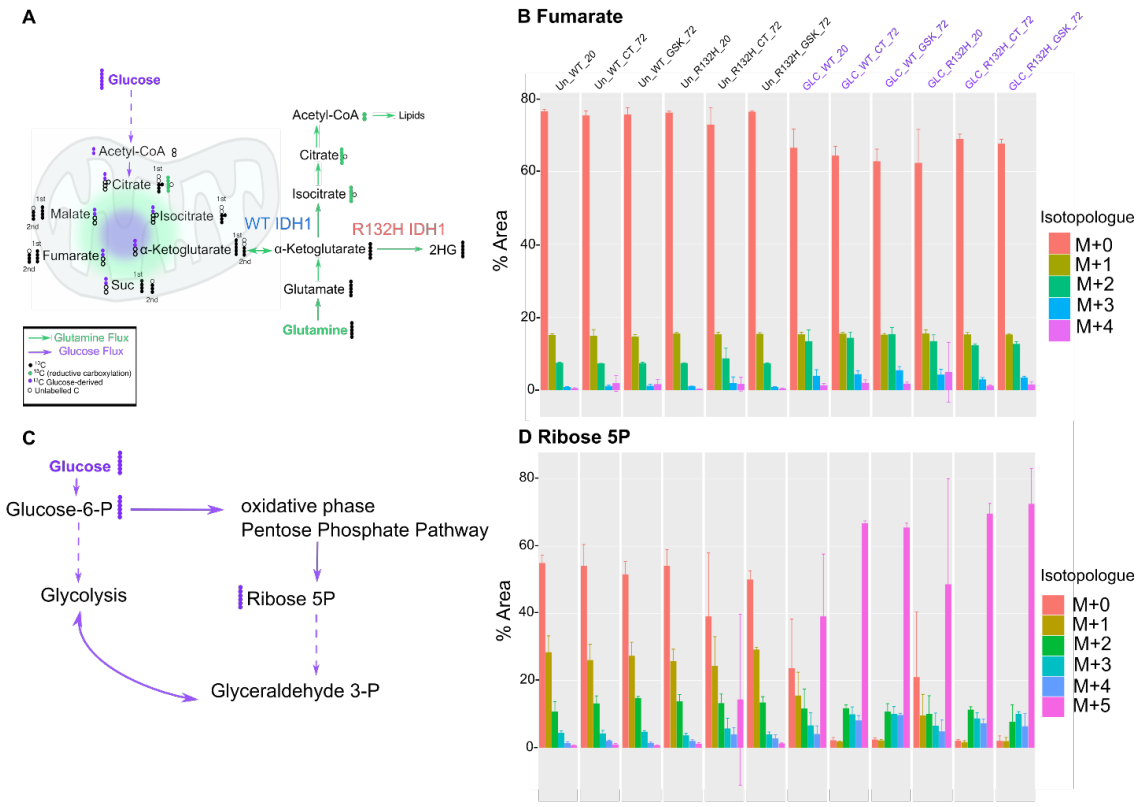


Figure 4.11 Glucose labelling patterns remain largely unchanged between WT and R132H cells. *A) Diagram showing glucose labelling patterns for TCA cycle intermediates (purple circles). When glucose is used as the main energy source, AcetylCoA produced at the end of Glycolysis bears two labelled carbons, which enter the TCA cycle. Thus, most of the TCA intermediates during the first round of the cycle will bear only two 13C carbons, while in a Glutamine labelling setting they would be carrying four to five labelled carbons (purple circles compared to black ones). B) Barplots represent the percentage of each Fumarate labelling pattern in a specific experimental group. X axis labels are common between B) and D), with black representing the unlabeled samples and violet their labelled counterpart. Error bars indicate the standard deviation for N=5 technical replicates. C) Diagram of Glucose labelling pattern between Glycolysis and Pentose Phosphate Pathway intermediates. In the first step of Glycolysis, fully labelled Glucose (six carbons) is converted into Glucose-6-phosphate. This intermediate can be either converted into Fructose-6-phosphate and continue with Glycolysis or enter the first phase (oxidative) of the Pentose Phosphate Pathway (PPP), where it will be converted to 6-phosphogluconate. Further reactions in this pathway will lead to the formation of Ribose-5-phosphate, which will bear five 13C upon labelling. Subsequent steps in the PPP produce Glyceraldehyde 3-phosphate, which can then re-enter Glycolysis. D) Barplots showing the percentage of distribution for each Ribose 5-phosphate isotopologue in labelled (purple, see panel 8B) or unlabeled (black, 9B) samples. Statistical significance amongst 72 hours samples was assessed for both 11B and 11D with ANOVA, but no differences were found between genotype or following GSK864 treatment. Error bars indicate the standard deviation for N=5 technical replicates.*

Thus, both WT and R132H cells are Glutamine-dependent, although Glutamine flux is increased upon IDH1 mutation. Glucose is still being used as an Energy source in equal measure in both cell lines, although not as the primary one.

4.8 Discussion

Our analysis revealed alteration of several metabolites in IDH1 mutated cells, including a group of TCA intermediates. Both WT and R132H cells show Glutamine dependence, but its flux is increased in IDH1 mutated cells. GSK864 treatment successfully rescued 2HG levels, but was ineffective in reducing levels of other compounds, suggesting a high specificity of action and raising the interesting question as to how the IDH mutation affects metabolism more broadly.

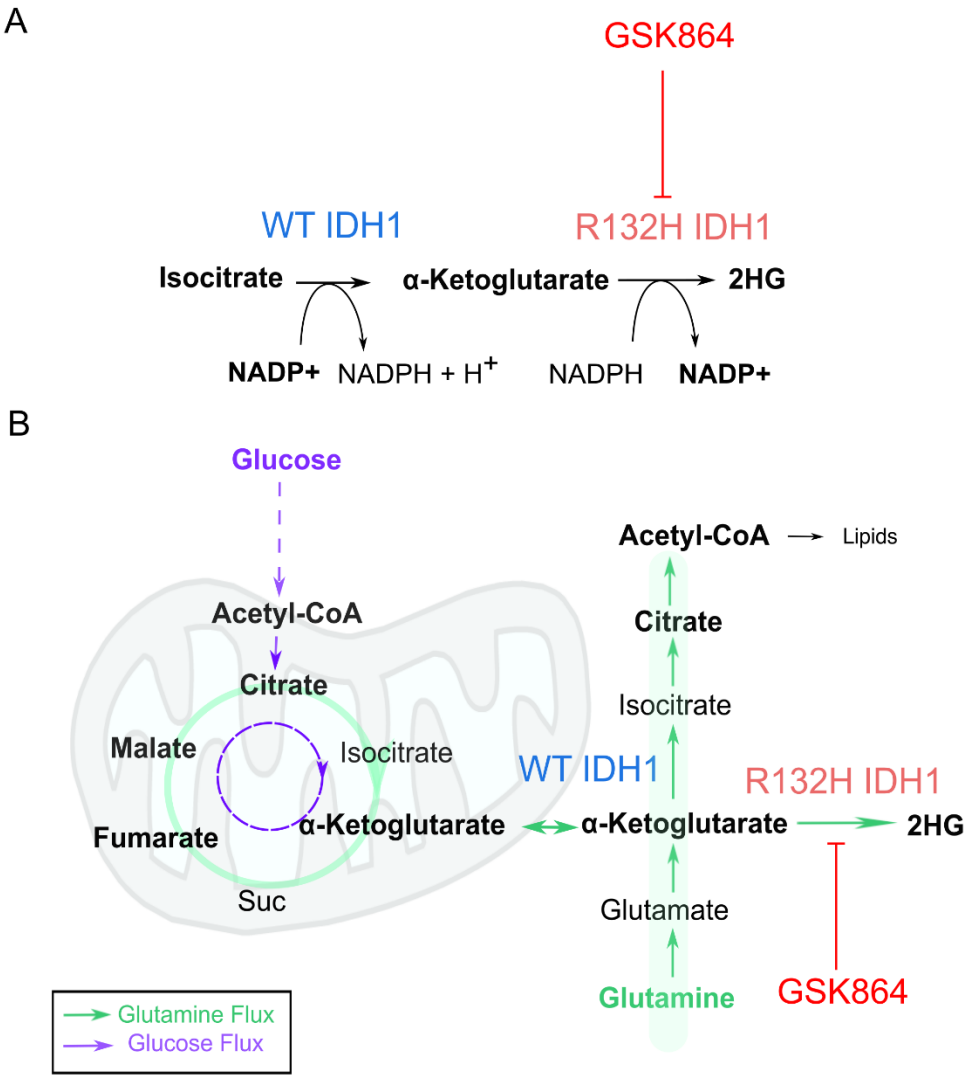


Figure 4.12 R132H IDH1 cells show increased 2HG levels and Glutamine flux, together with broad metabolic changes. *A) IDH1 and R132H IDH1-mediated reaction. As expected, IDH1 mutated cells show increased concentration of 2HG, successfully rescued by a 48-hours treatment with IDH1 inhibitor GSK864. B) Summary of other metabolites affected by IDH1 mutation (bold text). R132H cells showed increased Glutamine flux, as well as alterations in levels of several TCA cycle intermediates. Glutamine was the main source of energy, feeding both into the TCA cycle and reductive carboxylation. Despite*

cells preferably using Glutamine as the main energy source, Glycolysis was also fueling the TCA cycle, albeit to a similar extent in both WT and R132H IDH1 cells.

Indeed, many studies have measured 2HG and, to some extent, TCA cycle intermediates levels in cells bearing the IDH1 mutation, reporting a robust increase in the former and contrasting results regarding the latter (section 4.2). In our study, the detection of an increase in TCA cycle intermediates, coupled with higher concentrations of cofactors, such as NADP and NAD, and ATP suggests that the metabolic alterations induced by IDH1 mutation are not restricted to metabolites within its own biochemical pathway, but might affect more broadly energy consumption, an aspect that should be further investigated.

In our study, we have used metabolic flux to follow over time the consumption of labelled glucose or glutamine in our cellular model. We report that, while both WT and R132H cells are Glutamine-dependent, this flux is increased in cells bearing the IDH1 mutation. Activation of reductive carboxylation in these cells might hint at the presence of defective mitochondria, as could be expected in cancer cells. Furthermore, as the IDH1-catalysed conversion of Isocitrate to α-Ketoglutarate is mediated by NADP, increased production of 2HG could induce an imbalance of the NADP/NADPH ratio, potentially impairing other NADP-dependent pathways, such as redox homeostasis and lipogenesis (Badur et al., 2018). Finally, while IDH1 mutated cells show a similar phenotype and growth rate to their WT counterpart, an increased Glutamine flux could hinder their survival under stressful conditions, such as starvation, hypoxia or redox stress. This effect could be verified by evaluating differences in cell survival upon culturing in starvation medium, and explored further by repeating a metabolic flux in the same conditions. Given that IDH1 mutation is present in 80% of glioma patients, exploiting these cells' increased sensitivity to metabolic stress might help developing a new therapeutic strategy for this subset of patients.

The study of IDH1 often focuses on the epigenetic consequences that stem from this metabolic impairment. Indeed, IDH1 mutation typically results in increased methylation levels, both on DNA and histones, as a consequence of the 2HG-mediated inhibition of demethylases. Nevertheless, we did not detect any major differences in SAM levels, which is the main methyl donor in the cell. This could be due to the fact that the impairment affects this mark's erasers, rather than its writers.

Even though the most investigated epigenetic mark in relation to IDH1 mutation is methylation, this analysis also highlights a two-fold significant increase of Acetyl-CoA in R132H cells, which could be reflected in an increase of protein acetylation, including histones. We addressed this hypothesis in following experiments, aimed at assessing the impact of IDH1 mutation and subsequent GSK864 treatment on epigenetic marks such as histone methylation and acetylation, DNA methylation and gene expression.

5. Characterization of gene expression in R132H U87 cells using RNASeq

5.1 IDH1 mutation in Glioma patients

The study of IDH mutations has a high clinical relevance for glioma, as it occurs in 80% of the patients. Gliomas have been historically classified based on their histological characteristics into four different grades (I -IV, Figure 5.1), each associated with a progressively worse clinical outcome. Grade I gliomas (pilocytic astrocytomas) are slow-growing cancers, generally curable if surgically resected. Lower grade gliomas (II-III: astrocytomas, oligodendriogliomas and oligoastrocytomas) show an intermediate clinical course and can evolve into grade IV tumours (glioblastomas). Glioblastomas are very aggressive malignancies associated with poor prognosis, as the survival is around 15 months (D. N. Louis et al., 2016). Further stratification of patients based on their IDH1 mutation status has provided a better way to diagnose the disease, choose an appropriate therapy and better predict the outcome (Brat et al., 2015a; Suzuki et al., 2015).

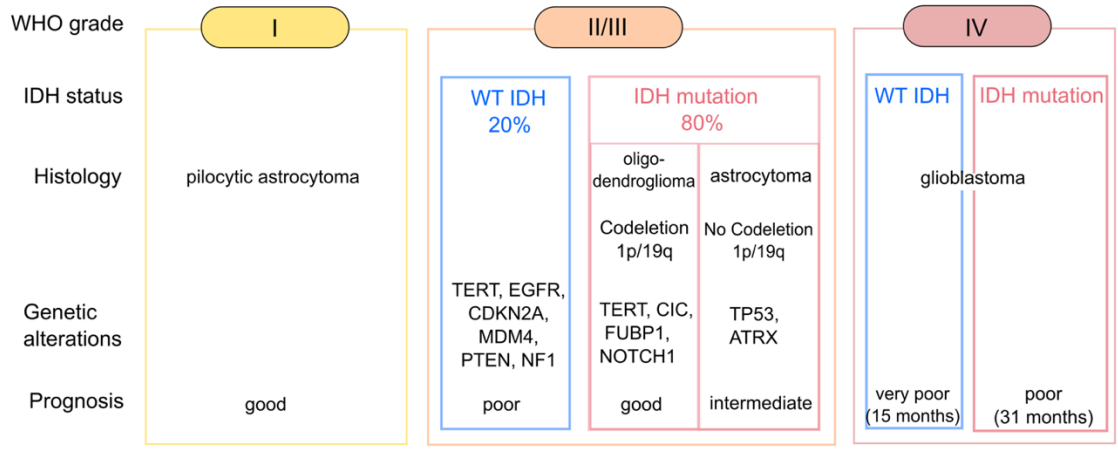


Figure 5.1. Schematic classification of gliomas*. Summary of the most recent WHO classification of gliomas, stratified by IDH mutation status, co-deletion of 1p/19q and other genetic alterations. Data adapted from* Louis et al., 2016 and Wen and Reardon, 2016.

Indeed, presence of the IDH1 mutation, with or without a concomitant codeletion of chromosome 1p and 19q has been associated with a better prognosis in the grade II/III gliomas (Brat et al., 2015a; Hartmann et al., 2010; Myung et al., 2014). The exact reason for this advantage is still unclear, and while some biological features of cells bearing the IDH1 mutation could help explain these observations, there are contrasting reports in literature. Increased production of 2HG inhibits α-Ketoglutarate-dependent dioxygenases, including a group of demethylases involved in DNA repair. Indeed, while ALKBH DNA demethylases repair DNA damage caused by alkylation, histone demethylases KDM4A and 4B participate to homologous recombination. Thus, inhibition of these proteins following IDH1 mutation would sensitize cells to chemotherapeutic agents, and this has been shown recently following treatment with PARP inhibitors (Sulkowski et al., 2020). However, an opposite effect was reported in a mouse model bearing the IDH1 R132H together with loss of function mutations of TP53 and ATRX: IDH1 R132H expression was still associated with better survival, even though the DNA damage response was upregulated and tumour cells were resistant to radiotherapy (Núñez et al., 2019). Thus, while it is true that presence of IDH1 mutation seems to be somewhat protective, the genetic context in which this is expressed might be important in understanding the exact molecular mechanism behind this advantage.

Other studies have reported that IDH1 mutation-induced hypermethylation of DNA and histones halts cell differentiation (Figueroa et al., 2010; Lu et al., 2012). Whether this can be linked to IDH1 mutation having a better prognosis is unclear. The presence of IDH1 R132H would keep glioma cells in a less differentiated state, subjecting them to more proliferation and the development of other genetic lesions, such as codeletion of 1p/9q or mutations in p53. On the other hand, this could be exploited as a therapeutic strategy: by inducing differentiation, malignant cells would be forced into a specialized, post-mitotic state and slow down, if not stop, tumour progression. Indeed, several IDH1 inhibitors have been shown to induce cell differentiation (Amatangelo et al., 2017; Okoye-Okafor et al., 2015; Rohle et al., 2013).

Standard glioma treatment usually involves a combination of surgical resection, radiotherapy and chemotherapy (D. N. Louis et al., 2007). While the first tends to be the most successful in terms of clinical outcome, it is not always a viable option, as many glioma subtypes are diffuse, rather than contained in a single area of the brain. Studies testing different combinations of radio- and chemotherapy based on the patients' IDH1 mutation status have highlighted a better survival rate following a more aggressive regime of radiotherapy combined with PCV (procarbazine, lomustine and vincristine) chemotherapy in patients bearing the mutation (Cairncross et al., 2013).

Recent advances in the understanding of R132H IDH1 role in cancer progression have fostered the development of novel therapeutic approaches, many of which are currently being tested in clinical trials (Table 13). A first strategy is represented by IDH inhibitors, targeting either IDH1 or IDH2. Notably, the IDH2 inhibitor AG-221 (Enasidenib) has been shown to decrease 2HG levels by 98%, promote cell differentiation and has since been approved by FDA for treatment of refractory or relapsed AMLs (Amatangelo et al., 2017). While many of the inhibitors are being tested for hematopoietic cancer therapy, early phase clinical trials are ongoing on glioma patients.

A second approach is to use PARP inhibitors, exploiting the IDH mutation-induced genomic instability. Indeed, after observation of a greater susceptibility to the alkylating agent temozolomide (TMZ) in IDH1 tumours with an impaired PARP DNA repair pathway following treatment with its inhibitor, olaparib, a combination treatment with both drugs has been tested in a Phase I clinical trial on patients with relapsed glioblastoma and results are awaited.

Finally, a few immunotherapy strategies are being developed. Programmed cell-death protein 1 (PD1) is a membrane protein of T-cells, which upon recognition of its ligand, PDL1, usually on normal cells, inhibits T cell activation and consequently stops the immune response. Many cancers exploit this mechanism by overexpressing PDL1 on their surfaces in a process called

immune evasion. A series of therapeutic strategies aim at preventing this mechanism by inhibiting PD1, thus encouraging T cell activation and the development of an immune response. Indeed, two clinical trials are testing the efficacy of a PD1 inhibitor alone or in combination with radiotherapy for the treatment of glioblastoma (Table 8). A different approach involves a peptide vaccine against R132H IDH1. Results in preclinical models have confirmed the triggering of an immune response coupled with a slower tumour growth (Pellegatta et al., 2015; Schumacher et al., 2014), and a Phase I clinical trial has been completed in late 2018, although results have not been made publicly available at the time of writing.

The IDH1 mutation has a high clinical relevance in the treatment of glioma, and recent advances in the studies of the molecular consequence of this mutation have been informing not only the diagnosis, but the therapeutic approach to use, improving their clinical outcome.

Table 13 Current targeted therapies for IDH1/2-mutant tumours

Drug	Target	Disease	Effect	Clinical Stage	Clinical Trial ID
AG-221	Mutant IDH2	Refractory or relapsed AML	Suppression of 2HG production, induction of cell differentiation	Approved by FDA	NCT01915498
AG-120	Mutant IDH1	Advanced hematologic malignancies	Suppression of 2HG production	Phase 1 (recruiting)	NCT02074839
IDH305	Mutant IDH1	Advanced malignancies	Suppression of 2HG production and cell proliferation	Phase 1	NCT02381886
AG-881	Mutant IDH1 and 2	Advanced hematologic malignancies	Suppression of 2HG production, induction of cell	Phase 1 (Ended)	NCT02492737

			differentiation		
Olaparib + TMZ	PARP	Relapsed Glioblastoma	Inhibition of DNA repair pathway and sensitization to alkylating agents	Phase I (Ended)	NCT01390571
Pembrolizumab	PD1	Recurrent Malignant Glioma With Hypermutator Phenotype	Inhibition of PD1, activation of immune response	Interventional (currently recruiting)	NCT02658279
Avelumab With Hypofractionated Radiation Therapy	PD1	IDH1 mutant glioblastoma	Inhibition of PD1, activation of immune response	Phase 2	NCT02968940
IDH1 Peptide Vaccine	IDH1	IDH1 mutant grade III and IV gliomas	Activation of immune response against IDH1 mutant cells	Phase 1 (Ended)	NCT02454634

5.2 Assessment of gene expression differences following GSK864 treatment in U87 WT or R132H cells

Our metabolomics experiment has highlighted a strong increase in 2HG levels, as well as other TCA cycle intermediates, including AcetylCoA, in R132H cells. Furthermore, treatment with GSK864 greatly reduced 2HG levels in R132H cells after a 48 hours treatment. We next asked whether lowering 2HG concentration affected gene expression. To this end, we performed 3' end RNA sequencing (RNASeq) of WT and R132H cells, treated with 0.5 μM GSK864 or DMSO (Ctrl) for 48 hours or one week.

5.3 RNA Sequencing of U87 cells reveals little effect of GSK864 treatment on R132H cells gene expression

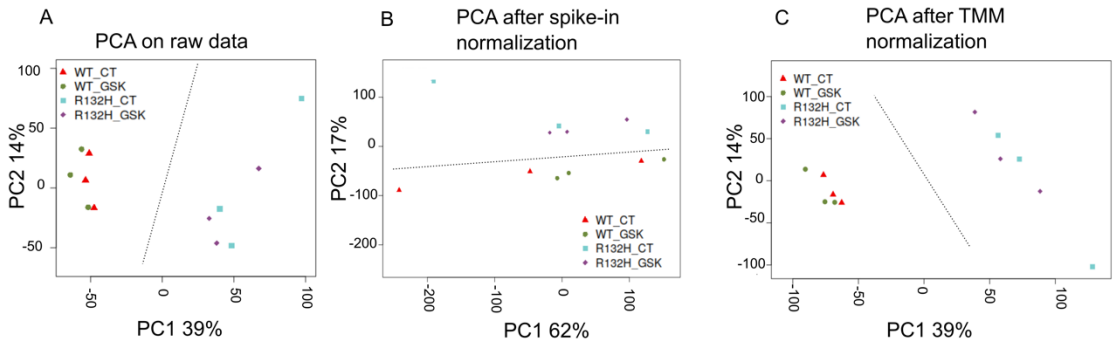


Figure 5.2. Principal component analysis of RNASeq data distribution before and after normalization. *A) Principal component analysis (PCA) of raw data, colored by treatment group. The first component (PC1) separates data according to genotype (dashed line): WT samples (red or green) cluster away from R132H (purple, light blue) and explains 39% of the variance, whilst PC2 explains 14%. B) PCA of RNASeq data after normalization with Drosophila spike-ins. Here, samples are separated across PC1 by an unknown variable, while PC2 clusters samples based on their genotype, as R132H samples and WT samples are on different sides of the dashed line. C) PCA of samples after normalization with the trimmed mean of M values (TMM) method and batch effect removal. PCAs, TMM normalization and batch effect removal were performed with separate functions from the NOISeq R/Bioconductor package* (Tarazona, Nueda, & Ferrer, 2016). *Normalization factors based on Drosophila spike-ins were calculated in base R as follows. The total number of human or Drosophila reads were calculated in each sample. The sample with the lowest number of human reads was selected as calibrator. The total number of Drosophila reads in the calibrator was divided by the total number of Drosophila reads in each sample, thus obtaining a normalization factor for each treatment group. Finally, each number in the raw gene counts table was multiplied by the appropriate normalization factor.*

Sequencing reads have been normalized either with Drosophila spike-ins or the Trimmed Mean of M values (TMM), and a batch effect due to biological repeats was removed using the ARSynSeq function from the NOISeq R/Bioconductor package (Tarazona et al., 2016). PCA analysis of data before or after each normalization was used to assess the most appropriate normalization method (Figure 5.2). From a first look, PCA analysis reveals a good separation between WT and R132H samples in the raw data, as well as after normalization (Figure 5.2, dashed line). However, GSK864 treatment does not seem to have a strong impact, as treated samples cluster very closely to their untreated counterparts (Figure 5.2, green and red (WT) or blue and purple samples (R132H)). Thus, the expectation would be to see a lot more differentially expressed genes between the two genotypes, and fewer changes upon GSK864 treatment. Following this analysis, TMM normalization was deemed the most appropriate, as

spike-in normalization is separating the samples according to an unknown variable, introducing noise in the dataset (Figure 5.2B).

Analysis of differentially expressed genes was carried out with edgeR using a linear model of Genotype + Genotype:Treatment, calculating gene expression differences amongst genotypes and the changes based on the interaction between genotype and treatment (Table 14) (M. D. Robinson et al., 2009; Smyth et al., 2015).

Table 14 Number of differentially expressed genes after 48-hours treatment with GSK864. *Analysis of differentially expressed genes was carried out as described in the Methods section. Rows describe the number of statistically significant genes with P.value or Benjiamini-Hochberg adjusted P.value (P.adj). Different comparisons considered are detailed in the first column. Notably, the model of the interaction between genotype and treatment (from here on referred to as "interaction") represents the genes that respond to the treatment differently in mutant compared to WT cells. It is calculated as the difference of differences, that is the difference between genes that are changing in R132H cells upon treatment and genes that are changing in the WT upon treatment: (R132H GSK – R132H Ctrl) – (WT GSK – WT Ctrl).*

Model: ~ Genotype + Genotype:Treatment (TMM normalization)		
	P.adj <= 0.05	P.value <= 0.05
R132H vs WT (Ctrl)	2262	5576
R132H GSK vs R132H Ctrl	19	1772
WT GSK vs WT Ctrl	18	1536
Interaction	0	2048

The analysis revealed 2262 genes changing between R132H and WT Control cells, but only a few genes varying after treatment with GSK864, in either WT or R132H cells (18 and 19, respectively, Table 14). Thus, GSK864 treatment has little effect on gene expression, despite the change in 2HG levels. As expected, a greater number of differentially expressed genes are found when considering P.value, rather than adjusted P.value alone. Genes with a significant adjusted P-value should be more robust, as the regular P-value in large datasets may be biased by multiple testing, meaning that a statistical test repeated for as many genes in a dataset is bound to show a significant p-value just by chance. To avoid false positives, p-values can be adjusted with a series of correction methods, in this case Benjamini-Hochberg. Adjusted p-values, however, might be introducing false negatives and genes of interest showing a significant p-value, but a non-significant adjusted p.value should be examined more carefully,

for example by checking their counts before and after normalization, or by plotting their expression levels in all samples. Levels of differentially expressed genes of interest have been examined by qPCR for further validation (Chapters 8 and 9).

Despite its action on 2HG levels, GSK864 treatment after 48 hours did not affect gene expression, with only 19 genes changing between R132H control and R132H GSK864-treated cells (Figure 5.3). The small number of genes makes further characterization difficult, as gene ontology enrichment does not find any terms significantly associated with this group of genes. Out of 19, five are novel transcripts (AC004223.4, AC009560.1, AC005332.4, AC008115.3, AL355377.2). AC092115.2 and NAP1L4P3 are pseudogenes, LNC-LBCS is a long non-coding RNA related to bladder and prostate cancer while NPAS2-AS is an antisense transcript to NPAS2, coding for a circadian clock component (X. Chen et al., 2019; P. Gu et al., 2019; Hoffman, Zheng, Ba, & Zhu, 2008). Amongst the remaining protein-coding genes, we find representatives involved in different cellular functions: some are involved in metabolism (leukotriene C 4 synthase, LC4S; chymotrypsin C, CTRC; dimethylarginine dimethylaminohydrolase 1, DDAH1, Parashuraman and D'Angelo, 2019; Szmola and Sahin-Tóth, 2007), others are related to immune activities (zeta chain of T cell receptor associated protein kinase 70, ZAP70; CD52, Chan et al., 1992) or involved in cellular adhesion (adhesion G protein-coupled receptor L2, ADGRL2, White et al., 1998). Thus, in spite of correcting the metabolic alteration, RNASeq results suggest only a small number of genes are robustly (adjusted p.value <= 0.05) changing upon GSK864 treatment after 48 hours.

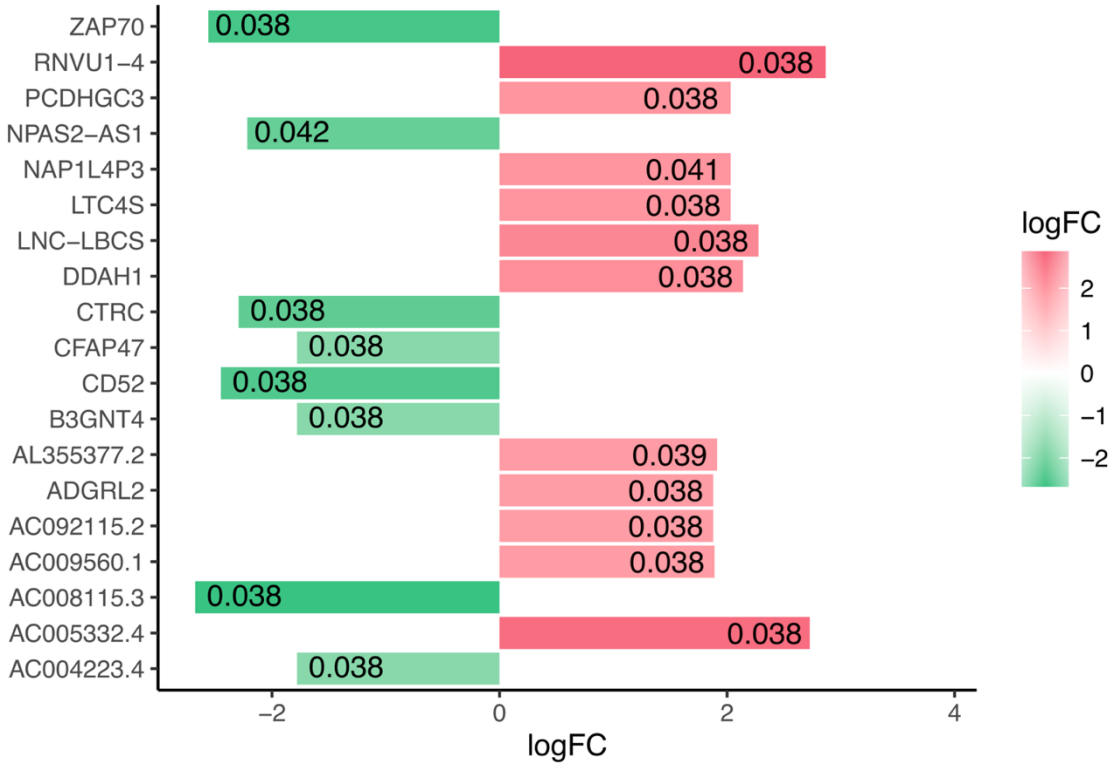


Figure 5.3. Expression levels of 19 genes change after treatment with GSK864 in R132H cells. *Bar plot of Log2 Fold Change between R132H control and R132H GSK864-treated cells of the 19 genes responding to GSK864 treatment. In spite of lowering 2HG levels down to the WT baseline, treatment with the IDH1 inhibitor GSK864 after 48 hours only changes the expression of 19 genes between R132H control and R132H GSK864-treated cells. Adjusted p-values for each gene, corrected with the Benjamini-Hochberg method, are printed at the end of each bar.*

5.4 Gene expression differences between R132H and WT cells

We then focused on characterizing the differences between R132H and WT U87 control cells. To highlight differential pathways between the two, we performed a gene set enrichment analysis (GSEA). This method aims at describing which sets of genes belonging to the same pathway are enriched in a particular dataset. In contrast with a Gene Ontology enrichment analysis, it considers the levels of all genes detected in an RNASeq experiment, rather than focusing only on the ones that are differentially expressed. Results show 40 pathways significantly different (p.value <= 0.05) in R132H cells compared to WT, 24 upregulated and 16 downregulated (Figure 5.4).

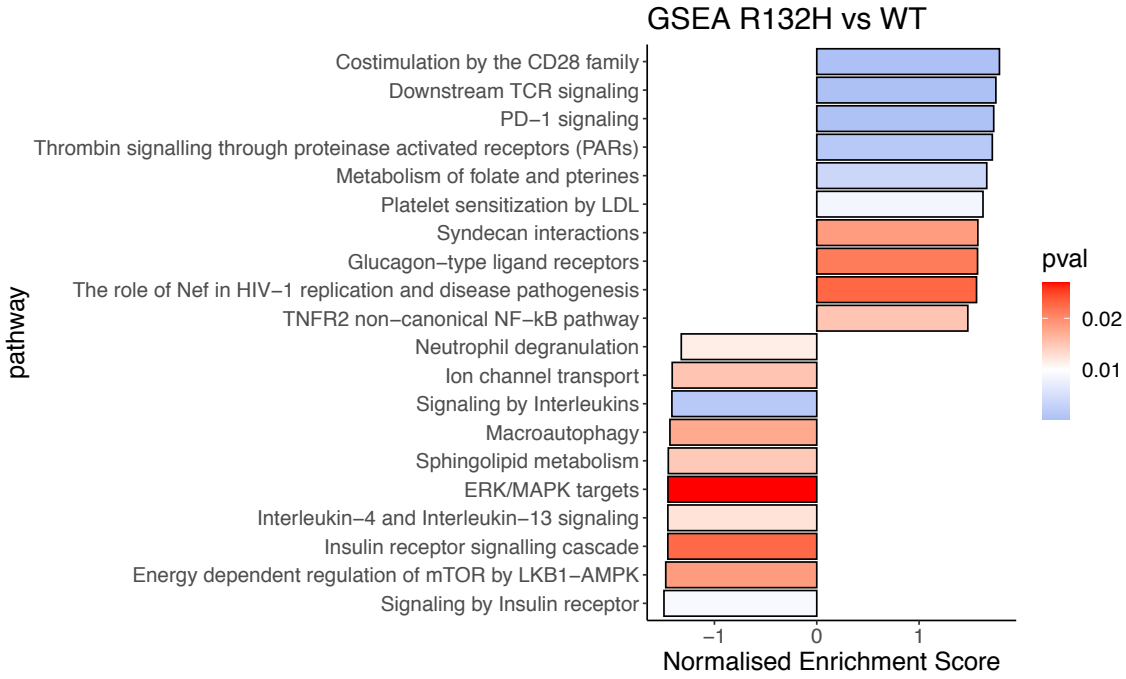


Figure 5.4. Gene set enrichment analysis suggests increase in immune response, folate metabolism and decreased Insulin receptor mediated signaling in R132H cells. *Top ten up- or downregulated pathways in IDH1 mutant cells. Gene set enrichment analysis was performed with the fgsea package (Sergushichev, 2016) on the full list of genes detected in RNASeq ranked by Log2FC multiplied by -log10 P.Value and using the Reactome pathway annotation. Bar colors indicates significance level: red <= 0.05, white = 0.01, blue < 0.01. All pathways shown have P.value <= 0.05, but a non-significant P.adjusted.*

Notably, many of the top twenty upregulated pathways hint at the activation of an immune response in R132H cells: “PD-1 signaling”, “downstream TCR signalling” and “costimulation by the CD28 family” pathways all show a significant enrichment (p.values: 0.0002, 0.002 and 0.0004, respectively). Core genes from these sets include HLA genes (HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB2, HLA-DRB1, HLA-DBR5), coding for MHC II class antigen presenting molecules, all significantly enriched in R132H control cells (Figure 5.5A).

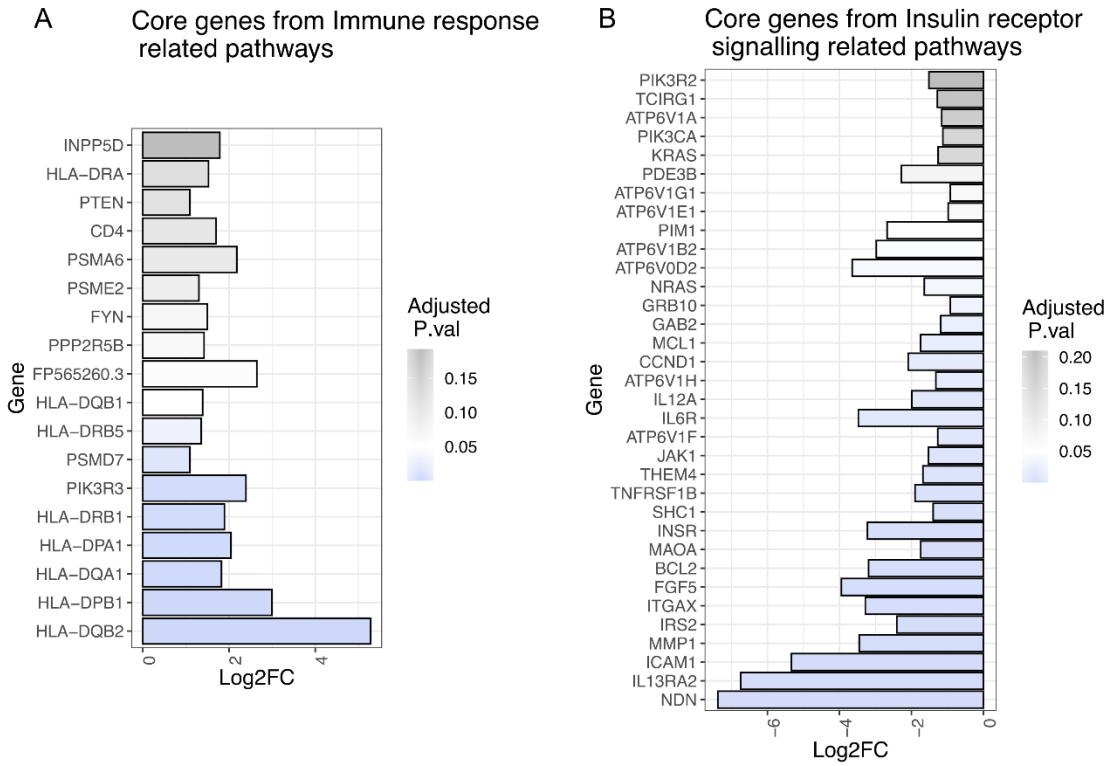


Figure 5.5 Gene expression levels of core genes from immune response- or Insulin signaling-related pathways. *A) Bar plots of the Log2 fold-change (x axis) of the core common genes from three immune response-related GSEA pathways ("Costimulation by the CD28 family", "PD-1 Signaling", "Downstream TCR signaling"). Bar colors reflect the significance level (grey > 0.05, white = 0.05, light-blue <= 0.05). B) Bar plots showing the log2 fold-change of core genes from Insulin signaling-related pathways ("Insulin receptor signaling cascade", "Interleukin 4 and Interleukin-13 signalling", "Signaling by Insulin receptor"). Bar colors reflect the significance level (grey > 0.05, white = 0.05, light-blue <= 0.05).*

Interestingly, the insulin receptor-related pathways "Insulin receptor signalling cascade", "interleukin-4 and interleukin-13 signalling" and "signalling by Insulin receptor" are amongst the gene sets significantly decreased (p. values = 0.009, 0.01 and 0.02, respectively) (Figure 5.4B). This signalling cascade can be activated by a series of cytokines, amongst which IL4 and IL13, through JAK phosphorylation. Upon activation, Insulin tyrosine kinase receptors are phosphorylated on their tyrosine residues, forming binding sites for their second messengers. Downstream signalling can lead to activation of different cellular pathways, such as KRAS/MAPK pathway, promoting cell proliferation, or phosphoinositide-3 kinase (PI3K), which stimulates lipid synthesis, mRNA translation and autophagy (Janku, Yap, & Meric-Bernstam, 2018). Interestingly, PI3K is often mutated in glioblastomas, causing an aberrant

activation of its signalling pathway, which contributes to cancer progression (McLendon et al., 2008). Thus, a decrease of this pathway in R132H cells could contribute to the protective effect of IDH1 mutations in gliomas (Brat et al., 2015b; Hartmann et al., 2010; Myung et al., 2014). Results also suggest an increase in Folate metabolism (Figure 5.6A).

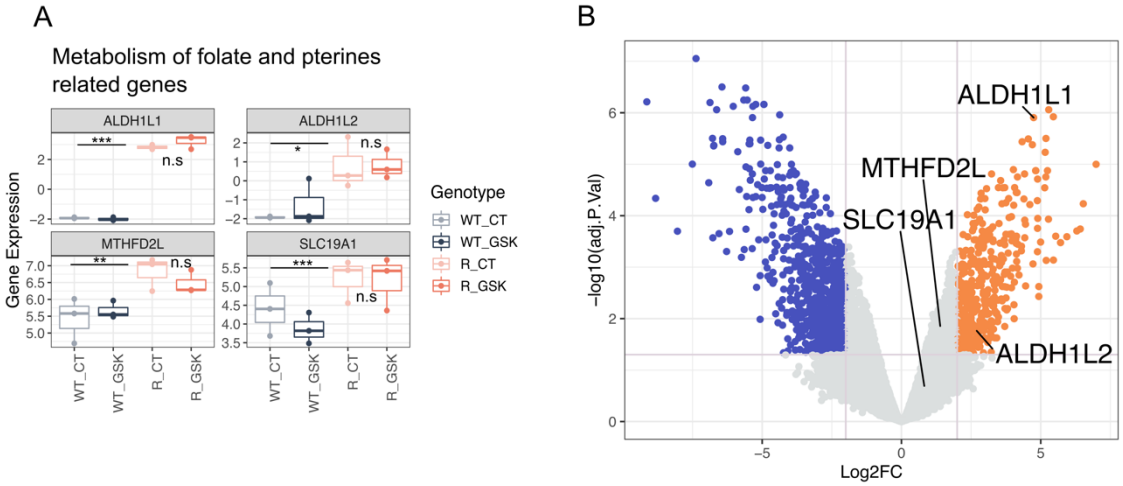


Figure 5.6 Gene set enrichment analysis suggests increased Folate metabolism in R132H cells. *A) Boxplots showing gene expression levels of core members of the "Metabolism of folate and pterines" pathway. Values correspond to the voom-normalized TMM value of each gene's expression as calculated by EdgeR package* (M. D. Robinson et al., 2009). *Significance levels for the comparisons WT CT vs R132H CT and R132H CT vs R132H GSK864 are indicated. B) Volcano plot showing the log2 Fold change in expression (x axis) over the -log10 of the adjusted p-values (y axis) for all genes detected in RNASeq. Blue dots correspond to genes with a significant p-value and a negative log2 fold change. Orange dots represent statistically significant genes with a positive log2 fold-change. Arrows indicate points corresponding to the "Metabolism of folate and pterines" related genes. Significance levels = ns P > 0.05; * P < 0.05; ** P < 0.01; *** P < 0.001. (N=3 per group).*

This pathway could be of interest, as Folate is involved in the methionine cycle, and thus contributes to the production of methyl moieties. Three out of the four representative genes belonging to this group, ALDH1L1, its mitochondrial isoform ALDH1L2, and MTHFD2L, code for enzymes catalysing subsequent steps in the Folate cycle, and are significantly upregulated in IDH1 mutant cells. However, in all cases treatment with GSK864 does not affect their expression levels (Figure 5.6A).

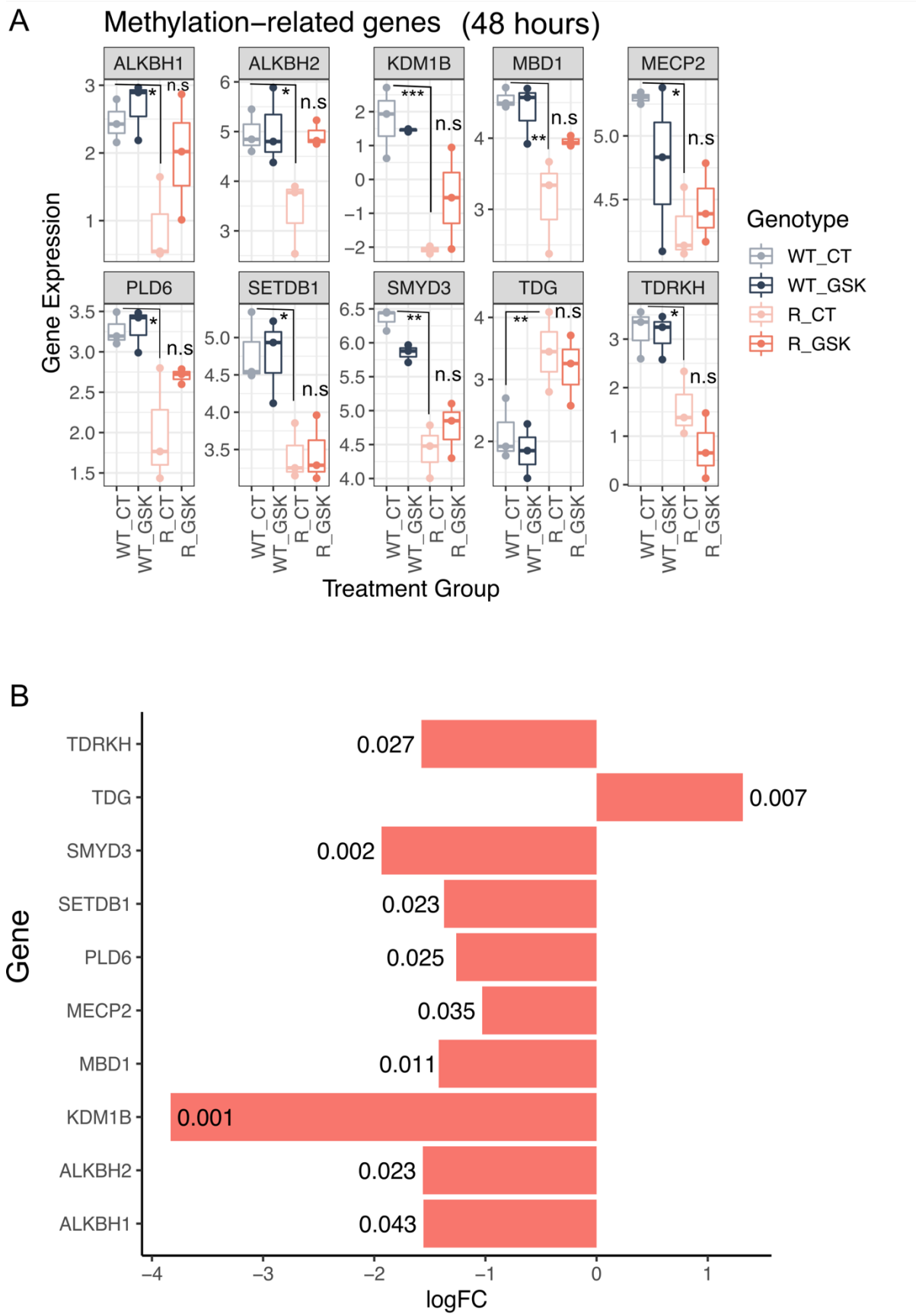


Figure 5.7 Methylation related genes show distinct pattern in IDH1 mutant cells. *A list of Gene Ontology (GO) terms related to methylation was obtained by searching the GO.db for children terms of GO:0006306 (methylation)(Carlson M, 2019). Then, the whole dataset of genes detected with RNASeq*

was scanned for genes annotated with any of these GO terms, using biomaRt (Durinck S, Spellman P, Birney E, & Huber W, 2009). *TMM normalized levels of such genes were plotted in the boxplots above. Significance levels for the comparisons WT CT vs R132H CT and R132H CT vs R132H GSK864 are indicated as following: Not significant = ns; P > 0.05; * P < 0.05; ** P < 0.01; *** P < 0.001. (N=3 per group). B) Barplot indicating log2 fold-change of methylation-related genes. Adjusted p-values are printed at the end of each data point. All genes are statistically different (p.adj <= 0.05) in R132H Ctrl vs WT Ctrl. Upon GSK864 treatment, all genes show a p.adjusted greater than 0.05, but ALKBH2, KDM1B and MBD1 have a significant p-value (ALKBH2 p.value = 0.0017, p.adjusted = 0.31; KDM1B p.value = 0.016, p.adjusted = 0.47 ; MBD1 p.value = 0.017, p.adjusted = 0.48).*

A well described consequence of IDH1 mutation is the indirect inhibition of DNA and histone demethylases (Lu et al., 2012; Turcan et al., 2012; Xu et al., 2011). We next asked whether any differentially expressed genes were coding for proteins involved in histone or DNA methylation and demethylation. Results show ten genes related to methylation, all of which are significantly different in the IDH1 mutant. Thymine-DNA Glycosylase (TDG) is the only gene whose expression is higher in R132H cells compared to WT, and codes for an enzyme involved in demethylation that removes Thymine from T/C or T/T mispairings (Neddermann et al., 1996). Amongst the nine downregulated genes, we find three α-Ketoglutarate-dependent dioxygenases with demethylases activity: ALKHB1 and 2, involved in DNA and RNA repair, and KDM1B, whose main target is Lysine 4 of Histone 3 (H3K4) (Ciccone et al., 2009; D.-H. Lee et al., 2005). The KDM family of lysine demethylases counts more than twenty members, so while detecting three of them might not be enough to claim significant changes in histone demethylases as a group, it is still an encouraging result. Interestingly, treatment with GSK864 increases their levels (p. value < 0.05, p.adjusted = n.s.), suggesting that lowering the concentration of 2HG might be sufficient to rescue their expression (Figure 5.7A). However, since this increase is not statistically significant, further validation is required.

Other downregulated genes in the IDH1 mutant are DNA methylation reader MECP2 and histone methyltransferases SETDB1 and SMYD3, but unlike α-Ketoglutarate-dependent dioxygenases, their expression levels do not change after GSK864 treatment (Amir et al., 1999; Hamamoto et al., 2004; D. C. Schultz, Ayyanathan, Negorev, Maul, & Rauscher, 2002) (Figure 5.7A).

Thus, analysis of the pathways enriched in R132H cells highlighted the activation of an immune response, together with an increase in the expression of enzymes involved in folate metabolism, potentially contributing to the production of methyl moieties. Furthermore, downregulation of insulin receptor signaling could negatively affect cellular proliferation and protein synthesis. In line with literature findings, we detect the downregulation of a small set of enzymes carrying out either demethylase or methyltransferase activities.

5.5 1 week treatment with GSK864 has greater impact on gene expression

Despite correcting the metabolic defect, a 48 hour treatment with GSK864 had only a modest impact on gene expression. We hypothesised a longer treatment would be needed in order to induce gene expression or epigenetic changes. Thus, the same experiment was repeated after a week-long treatment with 0.5 µM GSK864 or DMSO (Ctrl), refreshing the drug every 48 hours, as detailed in the Methods section.

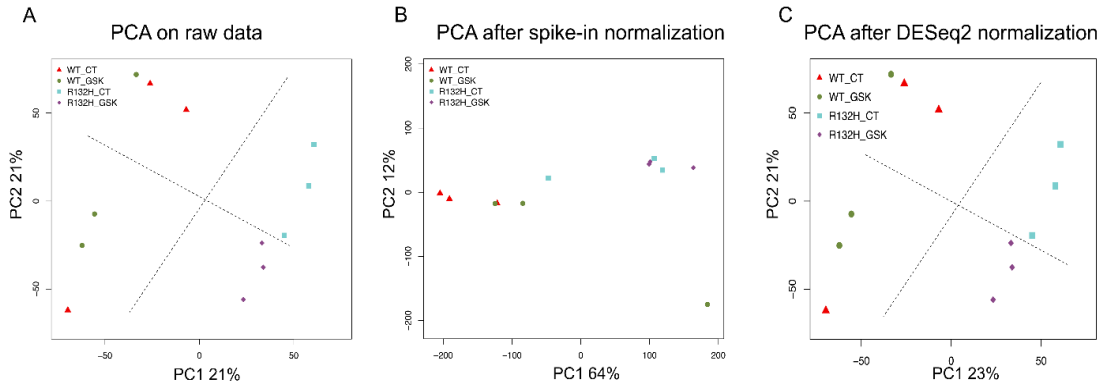


Figure 5.8 PCA showing data distribution before or after normalization in the 1 week dataset. *A) PCA showing distribution of raw data. While the first component (PC1) separates samples by genotype (dashed line), grouping WT(red or green) on one side and R132H samples on the other (violet or light blue), PC2 further clusters by treatment. B) PCA showing data distribution following normalization with Drosophila spike-ins. This method introduces an unknown variable separating samples in both PC1 and PC2, adding noise to the dataset. Spike-in normalization factors were calculated as described in Figure 1. C) PCA on data as normalized by DESeq2, accounting for library size. All PCA plots were obtained with the NOISeq R/Bioconductor package (Love et al., 2014a; Tarazona et al., 2016).*

Data distribution before or after normalization with either Drosophila spike-ins or standard analysis with DESeq2 was assessed with principal component analysis to choose the most appropriate normalization method (Figure 5.8). As with the 48 hours dataset, spike-in normalization introduced noise in the dataset, separating the samples according to unknown

variables (Figure 5.8B). In contrast with our previous RNASeq, no batch effect was detected, and differential gene expression was calculated using DESeq2 with the linear model Genotype + Genotype:Treatment, evaluating both genes changing in the two genotypes and genes responding differently to treatment with GSK864 in R132H compared to WT cells (as represented by the interaction model Genotype:Treatment) (Love et al., 2014a).

Table 15. Number of differentially expressed genes after 1 week of GSK864 treatment. *Sequencing reads were normalized either with Drosophila spike-ins or directly used as an input for DESeq2, an R package that normalizes data according to library size. PCA analysis of data after each normalization suggested DESeq2 was the most appropriate method. Analysis of differentially expressed genes was carried out as described in the Methods section, Rows describe the number of statistically significant genes with P.value or Benjiamini-Hochberg adjusted P.value (P.adj). Different comparisons considered are detailed in the first column. As in Table 1, "Interaction" models the interaction between genotype and treatment, and represents the genes that respond to the treatment differently in mutant compared to WT cells. It is calculated as the difference of differences: the difference between genes that are changing in R132H cells upon treatment and genes that are changing in the WT upon treatment: (R132H GSK – R132H Ctrl) – (WT GSK – WT Ctrl).*

Model: ~ Genotype + Genotype:Treatment (DESeq2 normalization)		
	P.adj <= 0.05	P.value <= 0.05
R132H vs WT (Ctrl)	451	1797
R132H GSK vs R132H Ctrl	275	1997
WT GSK vs WT Ctrl	0	190
Interaction	0	385

Results after one week treatment detect more genes with differential expression in R132H GSK compared to its untreated counterpart, but show a smaller number of robust changes (P.adjusted <= 0.05) between the R132H and WT control cell lines: only 451 genes are differentially expressed, against the 2262 found after 48 hours (Table 15). Moreover, we find little overlap between 451 genes significantly changing between R132H Ctrl and WT Ctrl and the 275 genes changing in R132H cells upon GSK864 treatment, as only 13 genes are in common (Figure 5.9A), suggesting that GSK864 treatment is also altering levels of genes that would normally be expressed at similar levels in R132H and WT cells. In spite of being all significantly different between R132H and WT cells, and all having their expression levels successfully recovered after GSK864 treatment, little commonalities were found amongst these 13 genes, coding for

proteins involved in immune reactions (CD68), cell cycle progression (S100A13, PKNI) or pseudogenes (AC026462.1, RPL10P3). Since these genes do not show any common function, other factors might explain the link between them, such as their position in the genome or the epigenetic marks around them.

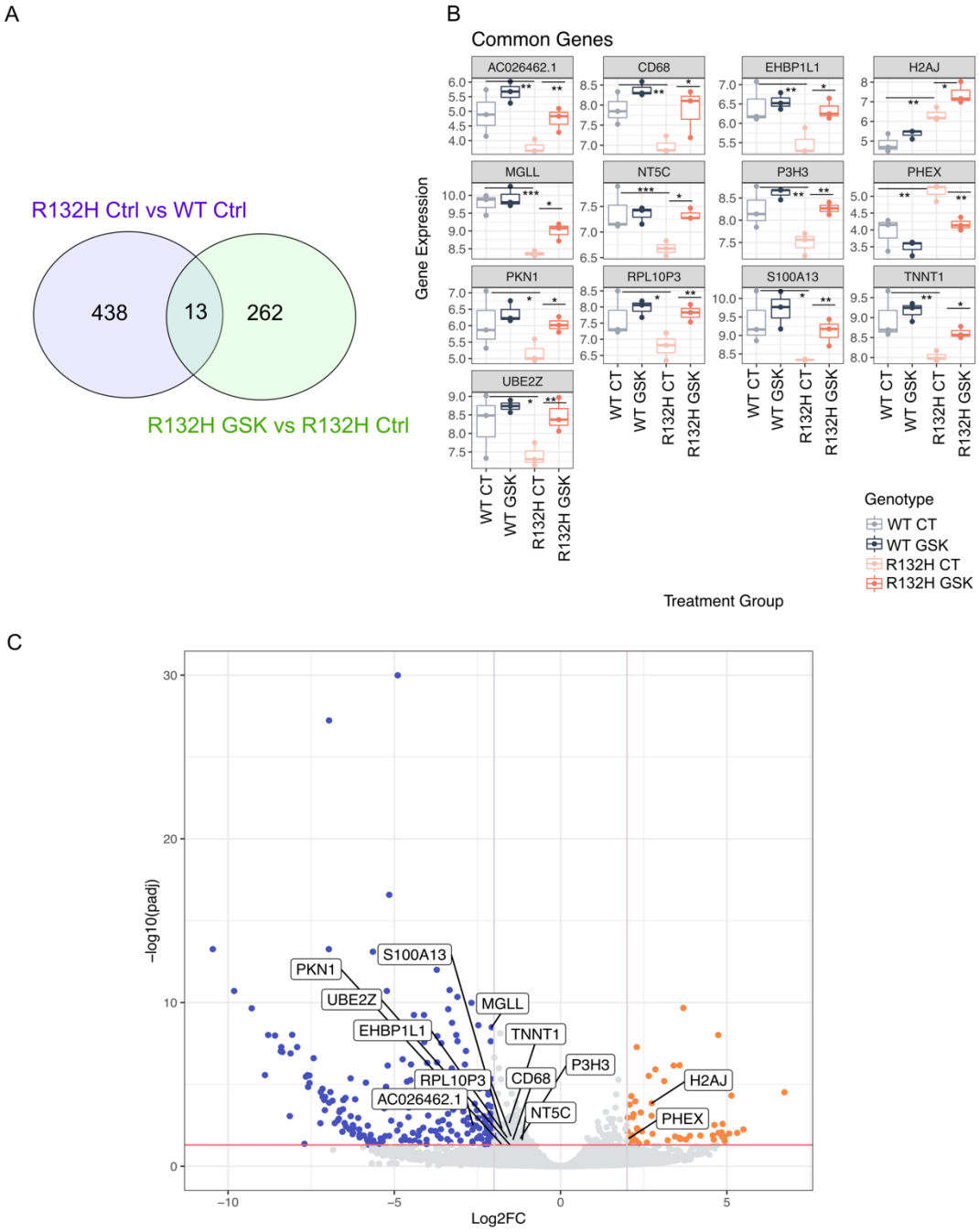


Figure 5.9. Common genes between Genotype (R132H Ctrl vs WT Ctrl) and R132H treatment (R132H GSK vs R132H Ctrl). *A) Venn diagram showing the 13 genes in common between the two categories of interest. B) Boxplots showing the DESeq2-calculated regularized log (rlog, a transformation applied to raw gene counts into the log2 scale that also normalizes by library size) gene expression levels for each of the 13 common genes* (Love, Huber, & Anders, 2014b). *Significance levels for the comparisons WT*

*CT vs R132H CT and R132H CT vs R132H GSK864 are indicated as following: Not significant = ns; P > 0.05; * P < 0.05; ** P < 0.01; *** P < 0.001. (N=3 per group). C) Volcano plot showing the log2 fold-change (x axis) and -log10 of the adjusted p-value (y axis) for the one week RNASeq dataset. All data points above the red threshold indicate genes with a significant adjusted p-value. Genes with a log2 fold-change of at least two-fold are colored in orange, while those with a log2 fold-change smaller than -2 are colored in blue. Labels indicate the position of the 13 common genes changing between R132H control and WT control, as well as R132H GSK864 compared to R132H control.*

Comparisons between R132H control and GSK864-treated cells showed a greater effect of the IDH1 inhibitor after one week, as 275 genes were significantly different compared to the 19 genes detected in the 48 hours dataset. Gene Ontology enrichment analysis of these 275 genes was performed using the GeneOntology online tool, powered by Panther and executed according to the website's instructions (http://geneontology.org/docs/go-enrichment-analysis/). Results suggests the treatment might affect cellular metabolism ("aspartate family aminoacid metabolic process", "acyl-CoA metabolic process", "thioester metabolic process"), as well as protein synthesis ("translation", "ribonucleoprotein complex biogenesis", "peptide biosynthetic process") and DNA transcription and repair ("transcription coupled nucleotide-excision repair") (Figure 5.10). As GSK864 has been shown to push AML cells towards differentiation, this enrichment in terms related to protein synthesis and potential cell proliferation might be reflecting a similar effect (Okoye-Okafor et al., 2015).

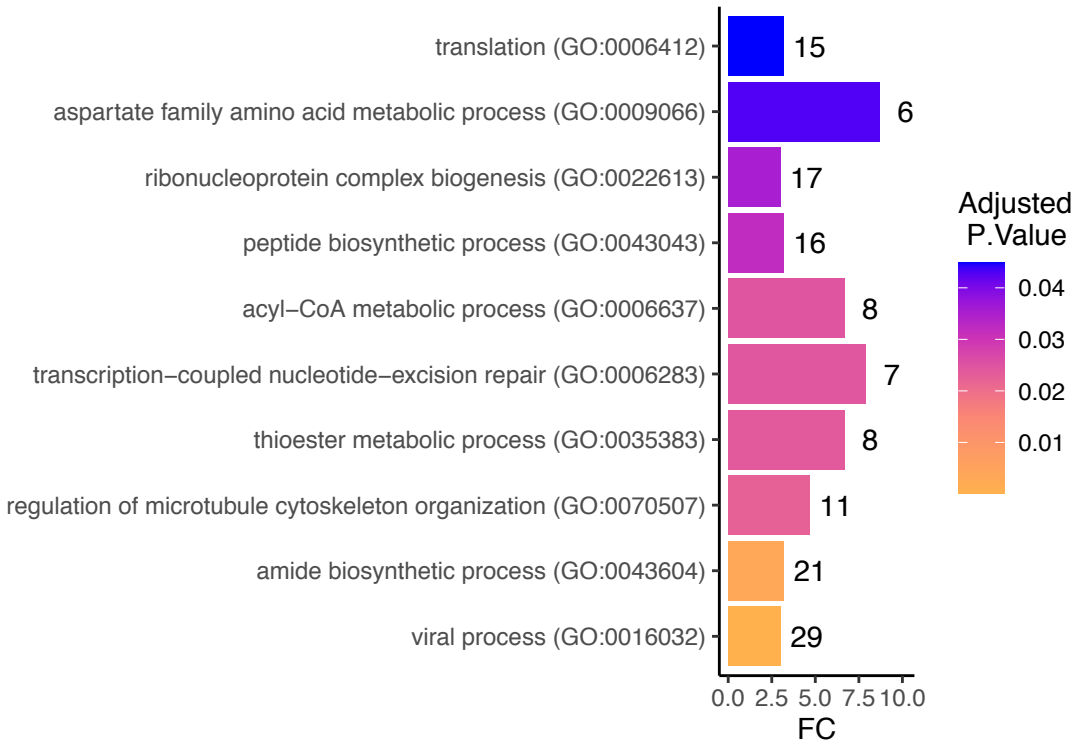


Figure 5.10 Top 10 Gene Ontology categories enriched in genes changing upon GSK864 treatment in R132H cells*. Bar plot showing the fold enrichment (FC) of each gene ontology. The number of genes in each category is printed at the top of the bar, while the color represent the adjusted p-value associated to each term. Gene Ontology enrichment analysis was performed using the GeneOntology online tool (http://geneontology.org/), powered by PANTHER* (Mi, Muruganujan, Ebert, Huang, & Thomas, 2018)

Enrichment of genes coding for enzymes involved in acyl-CoA metabolism is of particular interest, as we detected a two-fold increase in Acetyl-CoA levels in R132H cells (Chapter 4). While a 48 hours treatment with the IDH1 inhibitor was not decreasing its levels, maybe the longer treatment is having a greater effect on Acetyl-CoA, which would explain the expression changes in some of the enzymes involved in its metabolism (Figure 5.11A).

In our dataset, seven genes belonging to the "transcription coupled nucleotide-excision repair" GO significantly change following treatment with GSK864 (Figure 5.11B). DNA repair, specifically by homologous recombination, is reportedly impaired in IDH1 mutant cells, improving their sensitivity to chemotherapeutic drugs such as alkylating agents and PARP inhibitors (Rohle et al., 2013; Sulkowski et al., 2017, 2020). Thus, an increase in DNA repair-related genes upon IDH1 treatment would not necessarily be beneficial in terms of therapy, even though a rescue of the DNA repair defect could help in reverting cancer cells to a non-

transformed phenotype, if tumour development is still at an early stage. However, most genes associated to this category are more related to transcription than DNA excision repair: POLD1 and POLR3B are polymerase subunits, CCNH is a transcriptional regulator, while PPIE and PRPF19 are involved in splicing (Bertram et al., 2017; Hu et al., 2002; K. K. Kim, Chamberlin, Morgan, & Kim, 1996; Maga et al., 2013). *ERCC2* and *LIG1*, respectively, code for a member of the transcription factor II H complex, which, aside from taking active part to transcription, separates the double strand DNA in proximity of a damaged area, and a ligase that seals nicks in the double helix (Barnes, Tomkinson, Lehmann, Webster, & Lindahl, 1992; Taylor et al., 1997). Notably, the fact that a group of genes belongs to the same GO category does not mean that their expression levels will change in the same direction: in this case, two out of seven genes decrease upon GSK864 treatment, even though the category as a whole is enriched.

In conclusion, a week-long treatment with an IDH1 inhibitor has more impact on gene expression compared to a 48-hours one. GSK864 seems to affect genes involved in a variety of cellular processes, activating both transcription, translation and protein synthesis, and anabolic pathways such as AcylCoA metabolism.

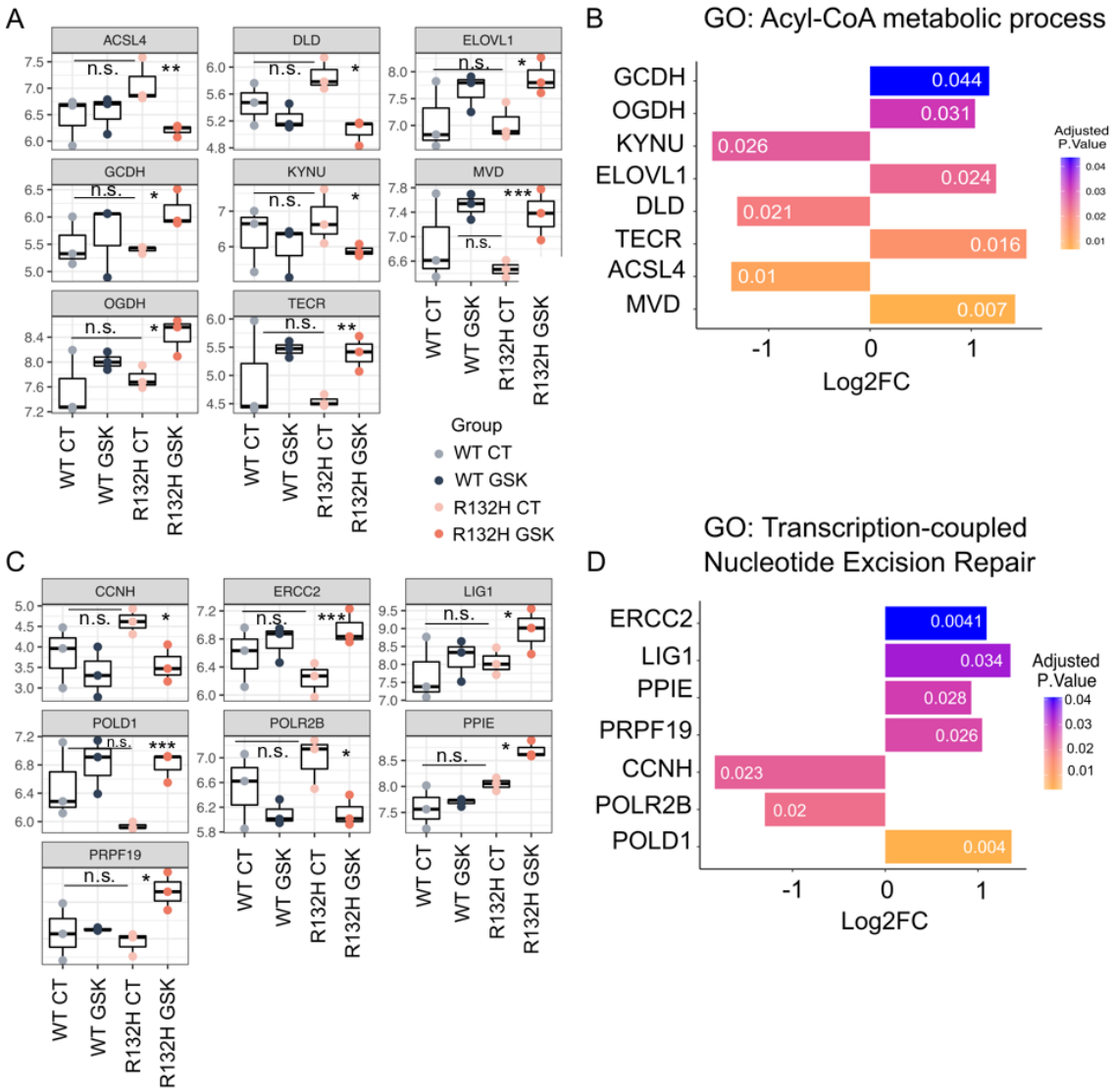


Figure 5.11. Example GO categories enriched in R132H cells treated with GSK864*. A) Boxplot showing normalized levels (rlog) of genes representing the "Acyl-CoA metabolic process" category. While the category as a whole is significantly enriched in R132H GSK864-treated cells, the single genes can be decreased or increased upon GSK864 treatment. Significance levels for the comparisons WT CT vs R132H CT and R132H CT vs R132H GSK864 are indicated as following: Not significant = ns; P > 0.05; * P < 0.05; ** P < 0.01; *** P < 0.001. (N=3 per group). B) Bar plot showing log2 fold-change of expression and adjusted p-value for each gene in the "Acyl-CoA metabolic process" group. C) Boxplots of core genes for the "Transcription-coupled Nucleotide Excision Repair" GO. D) Bar plot describing the log2 fold-change and adjusted p-value for each gene representing the "Transcription-coupled Nucleotide Excision Repair" category.*

5.6 Gsea analysis of 1 week dataset hints at increased Fatty AcylCoA Biosynthesis

To further investigate gene expression changes between Control R132H and WT cells, we repeated our GSEA analysis on the 1 week dataset. We found 68 pathways significantly

changing between the two cell lines, of which 14 upregulated and 54 downregulated (Figure 5.12).

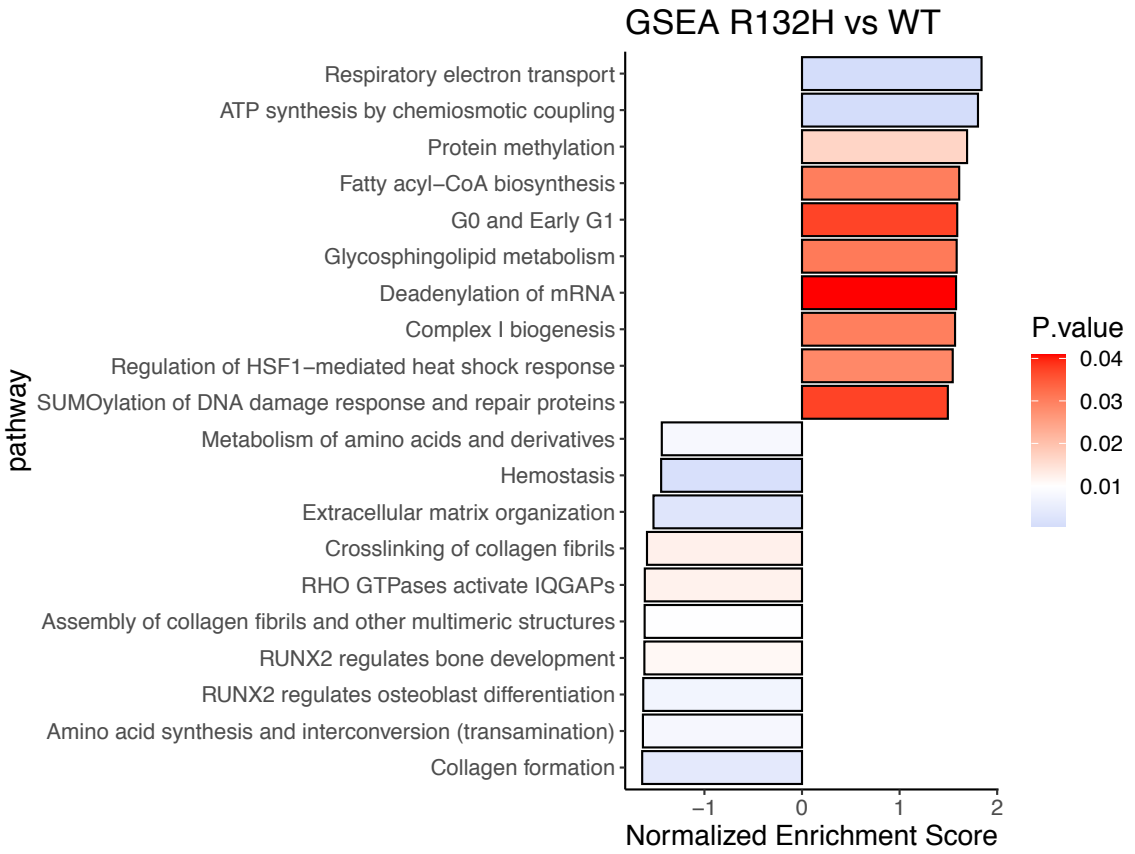


Figure 5.12 Top 20 gene sets enriched or decreased in one week dataset. *Top 10 statistically significant up- or down- regulated pathways. Gene set enrichment analysis was performed using the fgsea R/Bioconductor package using the full list of genes detected in RNASeq ranked by Log2FC multiplied by -log10 P.Value and Reactome pathway annotation.(Sergushichev, 2016). The normalized enrichment score for each of the most significant up- or down- regulated pathways is represented on the x axis, while the color indicates the significance level (red <= 0.05, white = 0.01, blue < 0.01) . All pathways shown had a significant p-value, but the respective p-adjusted was greater than 0.05.*

Amongst the top 20 enriched pathways appears an increase of "Fatty AcylCoA biosynthesis", confirmed by expression levels of genes belonging to this particular set (Figure 5.13). This is in line with the two-fold increase in AcetylCoA levels detected in our metabolomics experiment, and could imply an activation of genes coding for enzymes involved in fatty acid metabolism to make up for the excess AcetylCoA availability. Interestingly, GSK864 treatment in R132H cells is indeed changing expression levels of genes involved in AcylCoA metabolism, further hinting at a potential connection between IDH1 mutation and AcetylCoA levels (section 5.5).

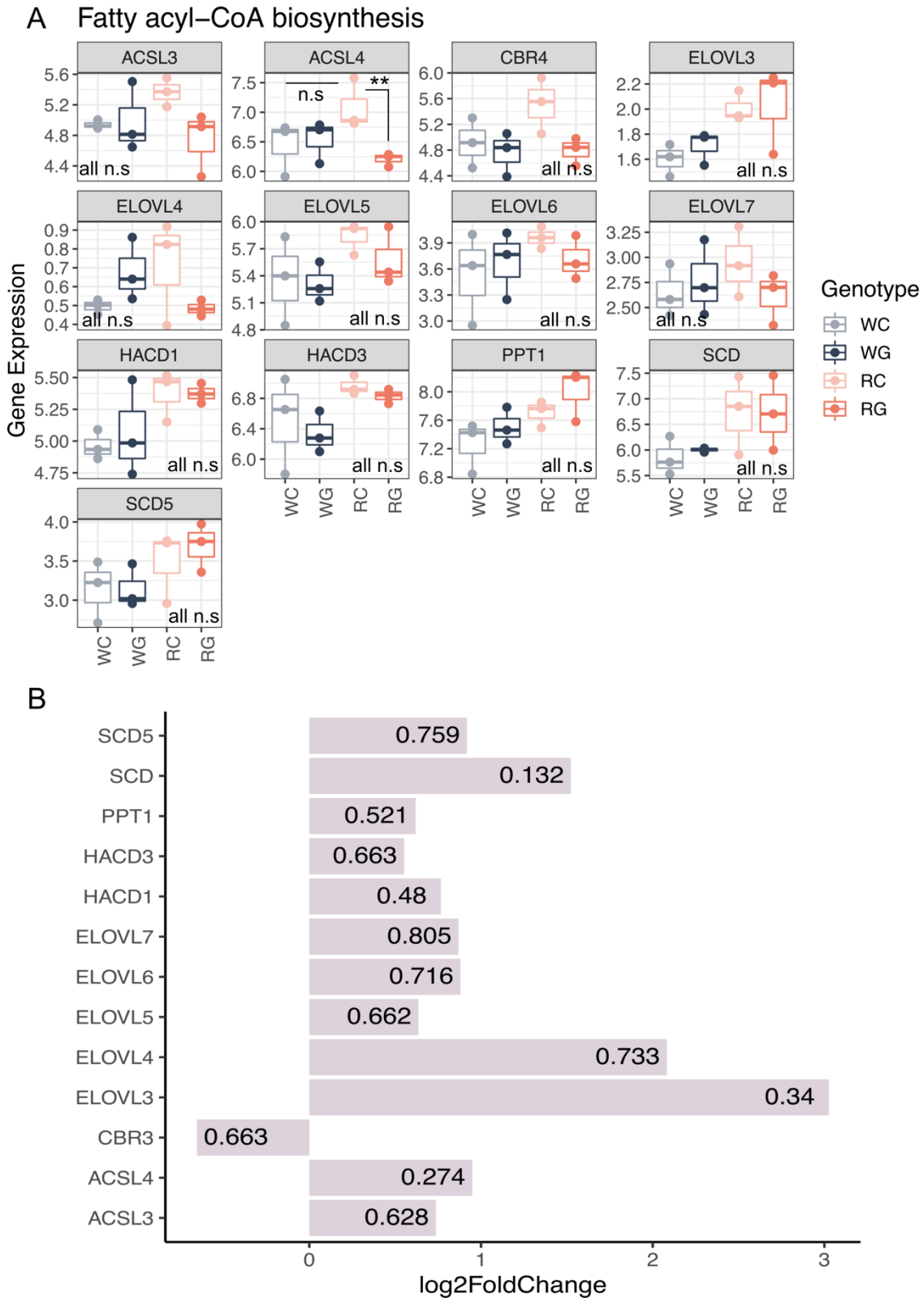


Figure 5.13 Expression levels of core genes from the "Fatty Acyl-CoA biosynthesis" set. *A) Boxplots showing the normalized (rlog) levels of each gene in all the treatment groups, for N=3 biological repeats. Significance levels for the comparisons WT CT vs R132H CT and R132H CT vs R132H GSK864 are indicated as following: Not significant = ns; P > 0.05; * P < 0.05; ** P < 0.01; *** P < 0.001. (N=3 per group). B) Barplot of the log2 fold-change of every gene, together with the associated adjusted p-value, printed at the end of each bar, as calculated in the differential expression analysis with DESeq2. (Love et al., 2014a). As GSEA considers all genes in a dataset, as opposed to the differentially expressed ones, to find pathways that might be over-represented as a whole, even though the levels of each single*

gene might be only slightly changed, it is not surprising that none of the genes in this set have a significant adjusted p-value.

GSEA results also show an increase in the "SUMOylation of DNA damage response and repair proteins" pathway (Figure 5.14). A small group of genes, *SUMO1*, *PIAS1* and *NSMCE2*, codes for proteins involved in SUMOylation, the attachment of small ubiquitin-related modifier to target proteins. This post-translational modification can alter protein-protein interactions, induce changes in activity of its substrate or in its subcellular localization (Gareau & Lima, 2010). SUMOylation has been linked with the DNA Damage response (DDR). Indeed, following DNA damage, several molecular players in the homologous recombination pathway are SUMOylated and defective SUMOylation hinders DNA repair (Cremona et al., 2012; Galanty et al., 2009; Silva et al., 2016). Four genes in this panel code for proteins directly involved in different types of DNA repair: SMC5 and WRN both participate to homologous recombination, XRCC4 to non-homologous end joining and CETN2 to nucleotide excision repair (Compton, Tolun, Kamath-Loeb, Loeb, & Griffith, 2008; Decorsière et al., 2016; Macrae, McCulloch, Ylanko, Durocher, & Koch, 2008; Nishi et al., 2005). Representative genes from this category are all increased in the IDH1 mutant, even though the increase is not statistically significant and the log2 fold-change is in most cases below 1 (Figure 3.14B). Notably, the expression of these genes is not significantly affected by GSK864 treatment (P.adjusted <= 0.05), even though boxplot of the expression levels of *NUP42* (P.value = 0.04), *NSMCE2* (P.value = 0.03) , *PHC1* (P.value = 0.07), *PIAS1* (P.value = 0.02), *STAG1* (P.value = 0.17) , *SUMO1* (P.value = 0.003), *WRN* (P.value = 0.1) and *SEH1L* (P.value = 0.05) show a reduction in R132H GSK864 cells compared to their untreated counterpart (Figure 5.14A). Thus, genes coding for proteins participating in SUMOylation and DNA damage response are generally enriched in R132H cells compared to the WT counterpart, hinting at a potential alteration of either pathways following IDH1 mutation. However, given that SUMOylation targets proteins in several cellular processes, it is difficult to pinpoint the exact molecular consequences of its

increase, although a deregulation in such a key post-translational modification would contribute to the general altered state of the cell.

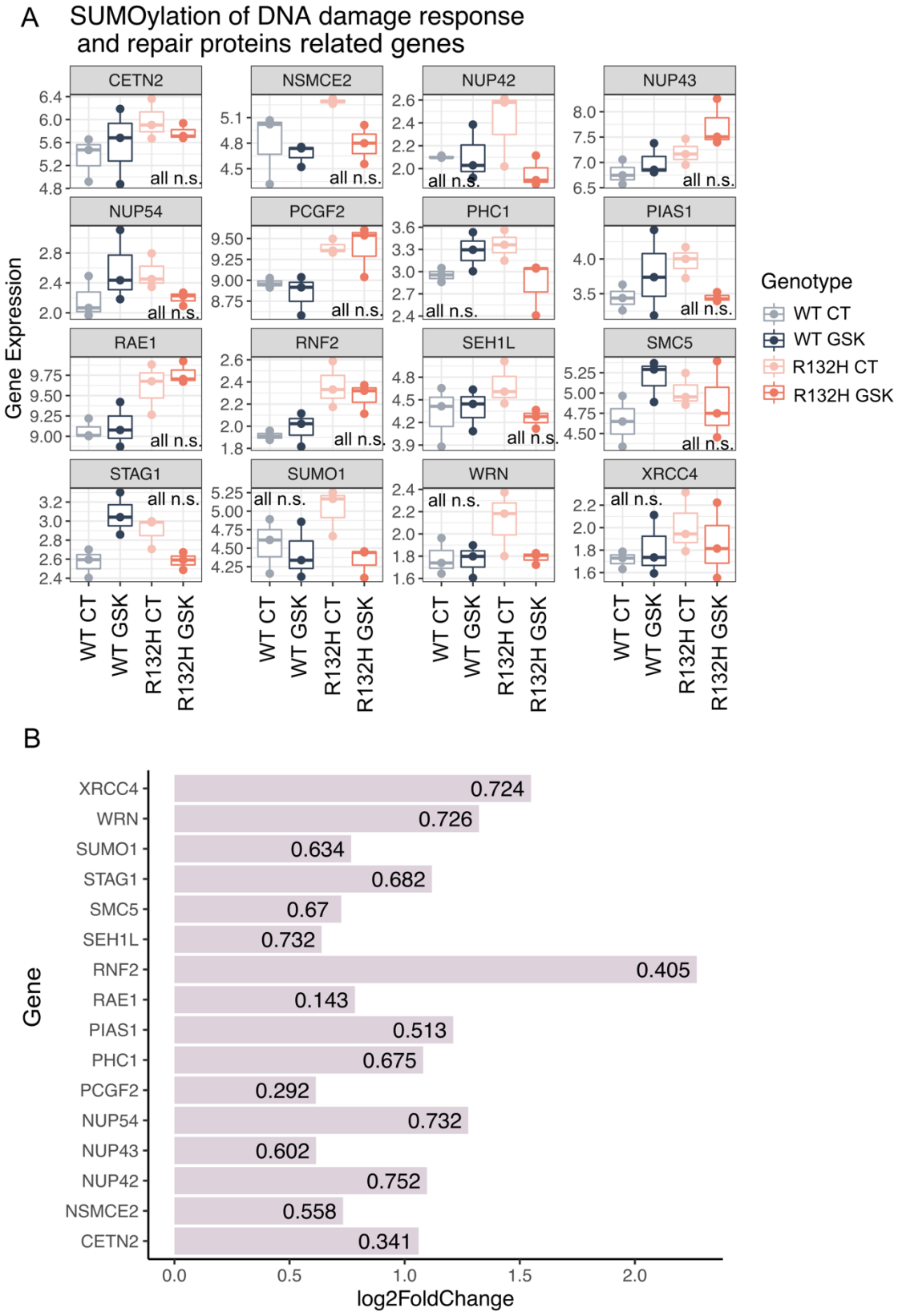


Figure 5.14 Expression levels of core genes from the "SUMOylation of DNA damage response and repair" set*. A) Boxplots showing rlog-normalized gene expression of core genes from the*

*"SUMOylation of DNA damage response and repair" pathway. GSEA analysis considers all genes detected in RNASeq, as opposed to only those differentially expressed. As a result, despite the general increase in expression in the IDH1 mutant, their difference is not statistically significant. Significance levels for the comparisons WT CT vs R132H CT and R132H CT vs R132H GSK864 are indicated as following: Not significant = ns; P > 0.05; * P < 0.05; ** P < 0.01; *** P < 0.001. (N=3 per group). B) Bar plot of the log2 fold-change for each gene, with associated adjusted p-value printed at the top of each bar, as calculated in the differential expression analysis with DESeq2. (Love et al., 2014a)*

In contrast to findings from the 48-hours dataset, only three methylation-related genes were differentially expressed after a week-long treatment (Figure 5.15). In line with previous results, levels of a methylation reader, MBD1, and a histone methyltransferase, KMT5A, are significantly downregulated in R132H cells, but their expression is not rescued after treatment with GSK864. Expression of a metallopeptidase involved in Lamin A maturation, ZMPSTE24, shows an opposite trend and its levels are lowered, but not significantly, after treatment (p.value = 0.04; adjusted p.value = 0.2).

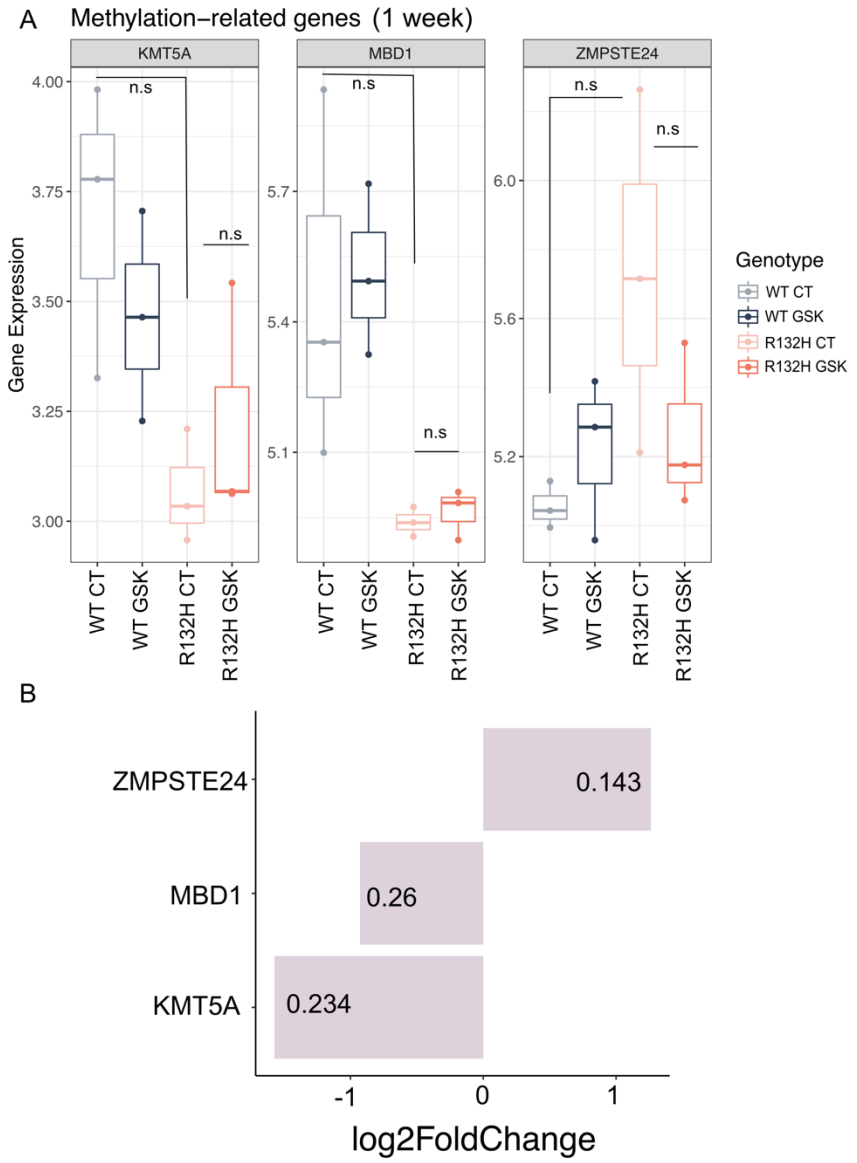


Figure 5.15 Methylation-related differentially expressed genes in the one-week dataset. *A list of Gene Ontology (GO) terms related to methylation was obtained by searching the GO.db for children terms of GO:0006306 (methylation)(Carlson M, 2019). Then, the whole dataset of genes detected with RNASeq was scanned for genes annotated with any of these GO terms, using biomaRt (Durinck S et al., 2009). Normalized levels of such genes (rlog) were plotted in the boxplots above. Significance levels for the comparisons WT CT vs R132H CT and R132H CT vs R132H GSK864 are indicated as following: Not significant = ns; P > 0.05; * P < 0.05; ** P < 0.01; *** P < 0.001. (N=3 per group). B) Bar plot showing the log2 fold-change of each gene. All genes are statistically different p-value, but not p-adjusted, in R132H Ctrl vs WT Ctrl, but nothing significantly changes upon GSK treatment.*

Thus, prolonged culture of R132H cells shows activation of gene programmes involved in Fatty Acid metabolism and potentially suggests alterations or activity of the SUMOylation and DNA repair pathways. The latter might provide an interesting field to investigate, as recent literature has linked IDH1 mutations to impaired DNA repair mechanisms (Sulkowski et al., 2020).

5.7 Discussion

RNASeq analysis in WT or IDH1 mutant cells treated with GSK864 revealed a small impact of the IDH1 inhibitor on gene expression after 48 hours, despite the successful correction of 2HG levels. Increasing the duration of the treatment for a week induces expression changes on a greater number of genes, affecting primarily RNA transcription, translation and protein synthesis.

Furthermore, R132H control cells show activation of distinct cellular and metabolic pathways when compared to their WT counterpart. Indeed, genes coding for proteins involved in antigen presentation, SUMOylation and DNA repair are enriched in R132H cells, while tyrosine kinase receptor signaling is decreased. Furthermore, increase in the expression levels of genes involved in Folate metabolism (48 hours) and Fatty AcylCoa biosynthesis (one week) suggests activation of biochemical pathways to make up for the altered metabolism in these cells.

Notably, we found a reduced number of differentially expressed genes between WT and IDH1 mutant control cells in the one week dataset compared to the 48 hours one (451 as opposed to 2262). Throughout the duration of the experiment cells were regularly split every 48 hours, so any effect due to hypoxia or nutrient consumption should be excluded. Perhaps the fact that cells in the week-long experiment required splitting, while those in the 48 hour treatment did not is somewhat contributing to the observed decrease in the number of differentially expressed genes between the two genotypes in the one week dataset. As cells were not synchronized for this experiment, another factor potentially introducing noise could be that these results reflect a mixed population of cells, although this is true for both datasets, so it would not explain the reduced number of differentially expressed genes after a week. This discrepancy could also be a consequence of prolonged culturing with DMSO, even at a small volume (10 µl in 10 ml of

cell culture medium). While GSK864 treated samples were also cultured with 10 µl of DMSO, as this is the solvent the drug is diluted in, and the number of differentially expressed genes upon GSK864 treatment is higher after a week-long treatment, maybe in the treated samples the effect of GSK864 itself affects gene expression to a greater extent compared to DMSO. Indeed, if both WT and R132H cells activate the same panel of genes to make up for DMSO-induced stress, the number of differentially expressed genes between the two control cell lines could decrease. Instead, GSK864-treated cells will respond to both DMSO and GSK864 treatment, but because GSK864 has a greater impact on gene expression, the number of genes differentially expressed between the treated and untreated cells is higher after a week-long treatment. To test whether this discrepancy is indeed due to prolonged treatment with DMSO, an RNASeq experiment on WT and R132H U87 cells alone could be performed at the same timepoints to assess differences in gene expression without potentially confounding effects due to DMSO, allowing to define the influence of time in culturing of these cells. Finally, downstream analysis of the 48 hour dataset required the removal of a batch effect, as samples were originally clustering according to the biological repeat. Thus, a possibility for the discrepancy in the number of differentially expressed genes between WT and IDH1 mutant cells is that the batch effect correction is introducing false positives. However, since the one week dataset did not require any batch effect removal, and more generally the two cell lines show a similar growth rate, morphology and only differ in the IDH1 mutation, the expectation would be to see a smaller panel of differentially expressed genes, so it is likely that the one week results are more reliable.

Results interpretation is also affected by criteria of statistical significance. While changes with a p.adjusted <= 0.05 are generally robust, but restricted to a smaller number of genes, a consistent number of genes shows a significant p.value, but a high p.adjusted. In such cases, evaluation of both p.values, coupled with visual representation of the gene's levels in each

biological replicate, in our case in a series of boxplots, should provide enough information to assess, case by case, whether changes in levels of a gene of interest can be considered trustworthy or not.

Longer treatment with GSK864 in R132H cells changed the expression of a greater number of genes compared to the 48 hours treatment, but its impact on the biology of cells is unclear. R132H cells treated with GSK864 did not show any visible change in their phenotype, maintaining similar morphology and comparable rates of cell proliferation at both timepoints tested. Indeed, upon harvesting, cell counts from all experimental conditions (WT Ctrl, R132H Ctrl, WT GSK864, R132H GSK864) were always comparable, reaching around 1.5 million cells each in both the 48 hours and one week experiment. This is somewhat in line with the relatively low number of genes showing significant changes in their expression upon GSK864 treatment at 48 hours, although at one week a greater number of genes are affected by treatment with GSK864 in R132H cells. Perhaps these cell lines proliferate in a similar manner in standard culture conditions, but would react differently to culturing under stress or starvation, especially given the metabolic differences observed in our metabolomics experiment. To investigate this, another RNASeq experiment could be performed after culturing cells in starvation conditions and treating with GSK864 for either 48 hours or 1 week. GO enrichment analysis suggested an increase in the expression of genes involved in translation and protein synthesis, as well as a few metabolic pathways (AcylCoA, Aspartate and Thioesters). This increase in translation and protein synthesis might suggest a role for the mTOR pathway, which can be activated downstream of TRK and PI3K signalling (Janku et al., 2018). Given the decrease in TRK signalling in R132H cells, it is tempting to hypothesize that GSK864 might revert this effect. While an increase in cell proliferation is not a desirable effect in cancer treatment, on the other hand it might sensitize cells to mTOR inhibitors, thus providing a viable therapeutic strategy (Janku et al., 2018). Furthermore, block of differentiation in IDH1 mutant cells was often

reported as a consequence of histone and DNA hypermethylation (Dogan et al., 2017; Lu et al., 2012; Waitkus, Diplas, & Yan, n.d.), and treatment with GSK864 was shown to induce differentiation in IDH1 mutant AML cells (Okoye-Okafor et al., 2015). While we do not see any differentiation pathway changing upon GSK864 treatment, perhaps the activation of transcription and translation is a hint in this direction.

Interestingly, the Fatty AcylCoA metabolism pathway gene set is enriched in R132H control cells, and a small group of genes assigned to the "AcylCoA metabolism" GO is found altered following treatment with GSK864, with many genes decreasing their levels upon drug treatment. Indeed, our metabolomics analysis has found a two-fold increase of AcetylCoA in R132H cells at 48 hours, and these levels were not changed upon GSK864 treatment. Based on these results, we hypothesize that in the one week dataset, this excess in AcetylCoA might lead to activation of genes coding for Acyl-CoA-metabolizing enzymes as an attempt to reduce its levels. Upon GSK864 treatment, expression of many genes from this set mostly decreases (Figure 5.11), perhaps reflecting the fact that GSK864 is successfully lowering levels of AcetylCoA, and since there is no longer need to metabolize it, expression of genes coding for enzymes of this metabolic pathway is reduced. To confirm this, a second metabolomics experiment should be performed after a week-long treatment with GSK864, to re-measure Acetyl-CoA and Malonyl-CoA (the first intermediate in Fatty Acid biosynthesis) levels in R132H and WT cells.

Finally, we assessed changes of gene expression as a way to evaluate potential effects of GSK864 treatment, following its successful correction of the main metabolic defect in IDH1 mutated cells. While little effect was detected after 48 hours, a larger number of genes changed following a prolonged GSK864 treatment. These cells, however, notoriously show an altered epigenome, bearing DNA and histone hypermethylation, and while the epigenome impacts on gene expression, changes on gene expression might not be the most immediate readout for

alteration in the methylation levels of histones or DNA (W A Flavahan et al., 2016; Lu et al., 2012; Modrek, Golub, Khan, Bready, Prado, Bowman, Deng, Zhang, Rocha, Raviram, Lazaris, Stafford, Leroy, et al., 2017; Rohle et al., 2013; Turcan et al., 2018). Thus, we next characterized the epigenetic landscape of WT and R132H IDH1 cells, evaluating potential effects of GSK864 treatment in rescuing the hypermethylation state.

6. Characterization of epigenetic alterations in R132H cells with ChIPSeq

6.1 Aberrant histone methylation in IDH1 mutated cells impairs different cellular processes

A well-studied consequence of IDH1 mutation is the increased production of 2HG. 2HG inhibits both protein and DNA demethylases. Experiment to study the effect of increased 2HG on levels of DNA methylation were halted due to Covid-19 and are not dealt with in this thesis. Here, the effect of increased 2HG on protein, specifically histone, methylation is described. 2HG induces inhibition of Jumonji-C domain histone demethylases (Xu et al., 2011) (KDMs), with a corresponding increase in selective methylation marks, including H3K27me3 and H3K9me3 (Lu et al., 2012). However, whether this increase in histone methylation affects cancer development through the formation of higher order chromatin structures, alterations in gene expression or impairment of other cellular processes is unknown.

Studies in *Drosophila* have described how H3K27me3 distribution seems to divide the genome into H3K27me3-enriched areas, corresponding to prominent TADs domains and delimited by CTCF binding sites, or H3K27me3–depleted areas, whose distribution correlates with TADs boundaries (El-Sharnouby et al., 2017; Van Bortle et al., 2012). While the connection between H3K27me3 and CTCF in maintaining domains is generally accepted, how CTCF exerts its insulator function is unknown. Knockdown of CTCF has been reported to have different outcomes on H3K27me3 distribution: at the genome-wide level, it does not cause spreading of this epigenetic mark into neighbouring domains (Schwartz et al., 2012; Van Bortle et al., 2012). Paradoxically, when considering single genes, “spill-over” of the H3K27me3 chromatin mark into the flanking regions is reported in CTCF knock-downs (Essafi et al., 2011; Soto-Reyes & Recillas-Targa, 2010). The next steps in the field will be fundamental to help clarifying these discordant results, perhaps by focusing on few specific CTCF binding sites to delete with the CRISPR-Cas9 technology, followed by assessment of H3K27me3 levels, rather than aiming at a global CTCF knock-down.

Another histone modification affected by IDH mutations is H3K9me3 (Lu et al., 2012; Sulkowski et al., 2020; Turcan et al., 2018, 2012). Studies have shown how this increase could have repercussions on cell differentiation. Introduction of R172K IDH2 or membrane-permeable 2HG in 3T3-L1 cells causes an increase in H3K9me3 and a block in their differentiation to adipocytes, due to inhibition of the α-Ketoglutarate dependent demethylase KDM4C. Such a mechanism could explain the reason behind the stimulation of differentiation triggered by treatment with IDH1 inhibitors, both in vitro and in patients (Amatangelo et al., 2017; Okoye-Okafor et al., 2015).

Differentiation is not the only cellular process impaired by histone hypermethylation. Work by Sulkowski et al., 2020 has suggested a novel mechanism that links H3K9me3 hypermethylation to an altered DNA repair response. Homology directed repair is a process through which cells repair double strand breaks in the DNA helix, and its most common form is homologous recombination. In normal cells, after the induction of a double strand break, a spike in H3K9me3 at the locus affected guides the recruitment of TIP60 and MRE11, protein sensors which will in turn activate ATM, promoting resection of the damaged filament. This generates 3' end single strand overhangs, promptly bound by the replication protein A (RPA) complex to avoid their degradation. RPA coating is then replaced by RAD51, promoting the search for an homologous sequence to invade. Once found, the broken strand displaces one strand on the homologous template, establishing a displacement loop that allows for DNA synthesis of the missing nucleotides on the damaged strand. Finally, the heteroduplex will resolve or dissolve depending on the pathway activated, with or without a crossover between the chromatids. Upon IDH1 mutation, the general increase of H3K9me3 prevents its initial spike, hindering recruitment of TIP60 and MRE11, thus suppressing DNA repair. The authors show how this could be exploited as a therapeutic strategy, as it sensitizes IDH1 mutant cells to treatment with PARP inhibitors (Sulkowski et al., 2020).

Most research on the impact of IDH1 mutations has focused on the consequences of 2HG-induced DNA methylation, but recent advances in the field have highlighted an equally important role for histone hypermethylation. Given that the correct balance of H3K9me3 and H3K27me3 is key for many cellular processes, such as heterochromatin formation and maintenance, DNA repair and cell differentiation, the search for druggable targets amongst these altered molecular pathways has and will further prove to be necessary to develop better therapeutic strategies.

6.2 Characterization of histone H3K9 methylation in R132H cells with ChIPSeq

To assess levels of histone methylation in our U87 R132H cell line, we performed chromatin immunoprecipitation followed by sequencing (ChIP-Seq) of H3K9me3. This histone modification was chosen after a literature search of the most common methylation marks targeted by α-Ketoglutarate-dependent demethylases (summarised in Table 16). U87 cells express around 21 genes coding for different histone demethylases, most of which trend towards a decreased expression upon IDH1 mutation (Figure 6.1).

Table 16. Summary of main histone modifications targeted by α-Ketogluratate dependent demethylases. (Figueroa et al., 2010; Lu et al., 2012; M. Gagné, Boulay, Topisirovic, Huot, & Mallette, 2017a; Tsukada et al., 2006; Turcan et al., 2012)

Histone Modification	Number of demethylases	Demethylases
H3K4	8	KDM1A, KDM1B, KDM2B, KDM5A, KDM5B, KDM5C, KDM5D, JMJD9
H3K9	13	KDM1A, KDM3A, KDM3B, KDM3C, KDM4A, KDM4B, KDM4C, KDM4D, KDM4E, KDM7A, KDM7B, KDM7C, JMJD10
H3K36	8	KDM2A, KDM2B, KDM4A, 4B, 4C, 4D, KDM8, JMJD9
H1.4K26	4	KDM4A, KDM4B, KDM4C, KDM4D
H3K27	5	KDM6A, KDM6B, KDM6C, KDM7A, KDM7C

H4K20	3	KDM7A, KDM7B, KDM7C

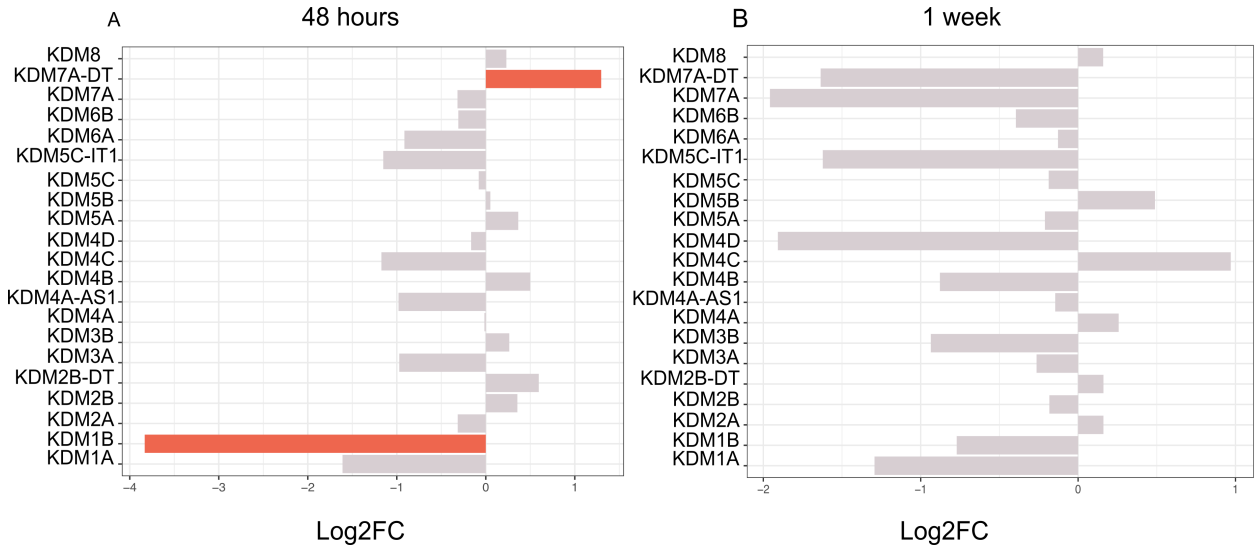


***Figure 6.1 mRNA expression levels of histone demethylases in R132H control cells**. Log2 fold change of KDMs gene expression in the 48 hours (A) or one week (B) dataset. Bar colors correspond to significance levels: grey bars indicate an adjusted p.value >= 0.05 (non-significant), while red bars indicate statistically significant adjusted p.values (KDM7A-DT p.adj: 0.00000052; KDM1B p.adj: 0.001).*

To investigate whether treatment with GSK864 would be sufficient to revert any hypermethylation, given its efficacy in lowering 2HG levels, WT and R132H U87 cells were treated with 0.5 µM of GSK864 for either 48 hours or a week.

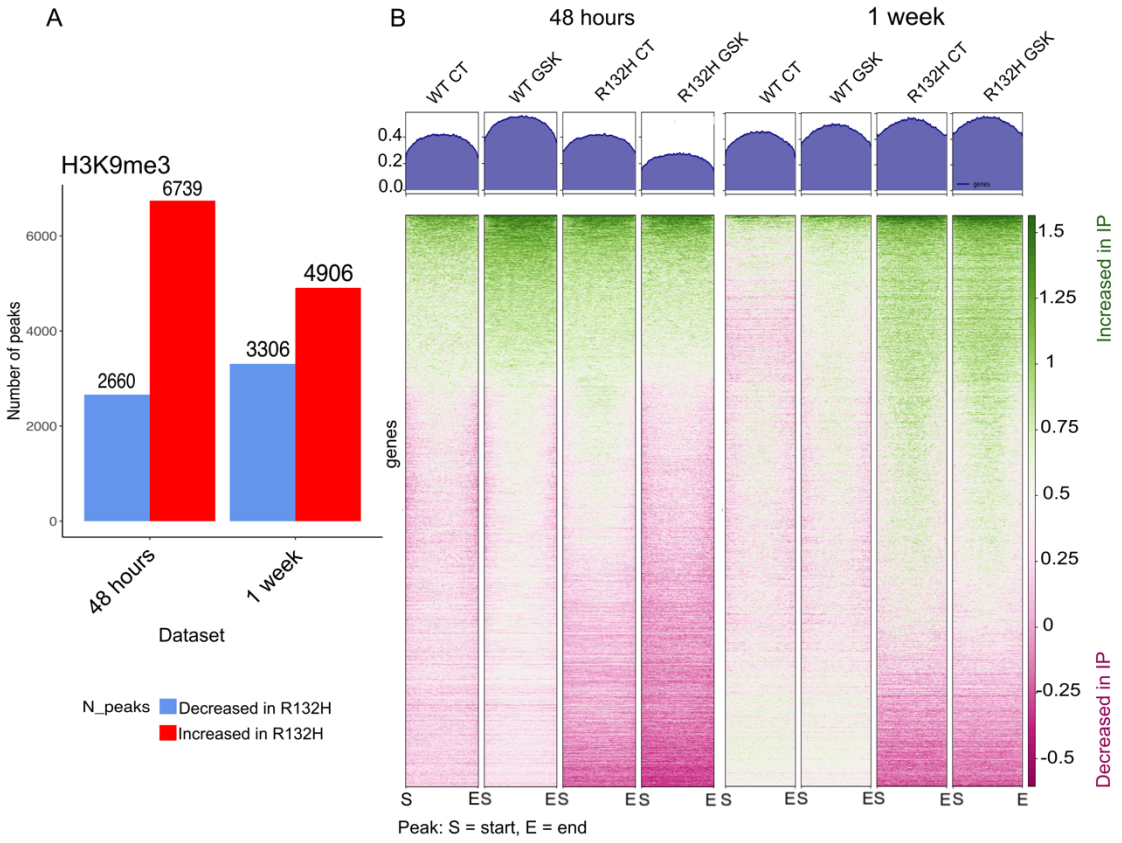


Figure 6.2 Genome-wide increase of H3K9me3 in R132H cells*. A) Bar plot summarizing number of differential H3K9me3 peaks between R132H and WT cells. H3K9me3 broad peaks were called using MACS2, then differential peaks analysis was carried out using HOMER. B) Heatmaps showing the ratio of IP/Input ChIP-Seq reads to highlight H3K9me3 distribution across the genome. Biological replicates reads (N=3 for the 48 hours dataset, N=2 for one week) were merged, then normalized to 1x coverage using bamCoverage (Deeptools). Log2 of the fold change between IP and Input was calculated by running bigwigCompare on the normalized files, then results were visualized in a heatmap using computeMatrix followed by plotHeatmap, all part of Deeptools* (Ramírez, Ryan, Grüning, et al., 2016). *In the heatmap, purple-colored cells indicate loci decreased in the IP sample compared to the Input, whereas green cells show enrichment in the IP. Profiles of the average signal over the peaks are plotted on top of each sample.*

Results indicate an increase in H3K9me3 upon IDH1 mutation (Figure 6.2). In both datasets, the number of differentially enriched peaks largely exceeds the number of peaks significantly decreased: 6739 vs 2660 in the 48 hours dataset, 4906 vs 3306 at one week (Figure 6.2A). In contrast with the general H3K9me3 hypermethylation, a subset of loci shows a decrease in levels of this modification in R132H cells (Figure 6.2B, purple cells). Further analysis of their distribution highlights a striking proportion of H3K9me3 peaks decreased in chromosome X and, to a lesser extent, autosomal chromosomes 6, 13 and 21 in the 48 hours dataset, where the number of peaks decreased is comparable to the number of loci with increased H3K9me3

(Figure 6.3). While a similar pattern is present at the genome-wide level in the one week dataset (Figure 6.2B, right panel), distribution of differentially enriched or decreased peaks shows a greater proportion of peaks increased in almost all chromosomes, except chromosome 18 and chromosome X (Figure 6.3A, right panel).

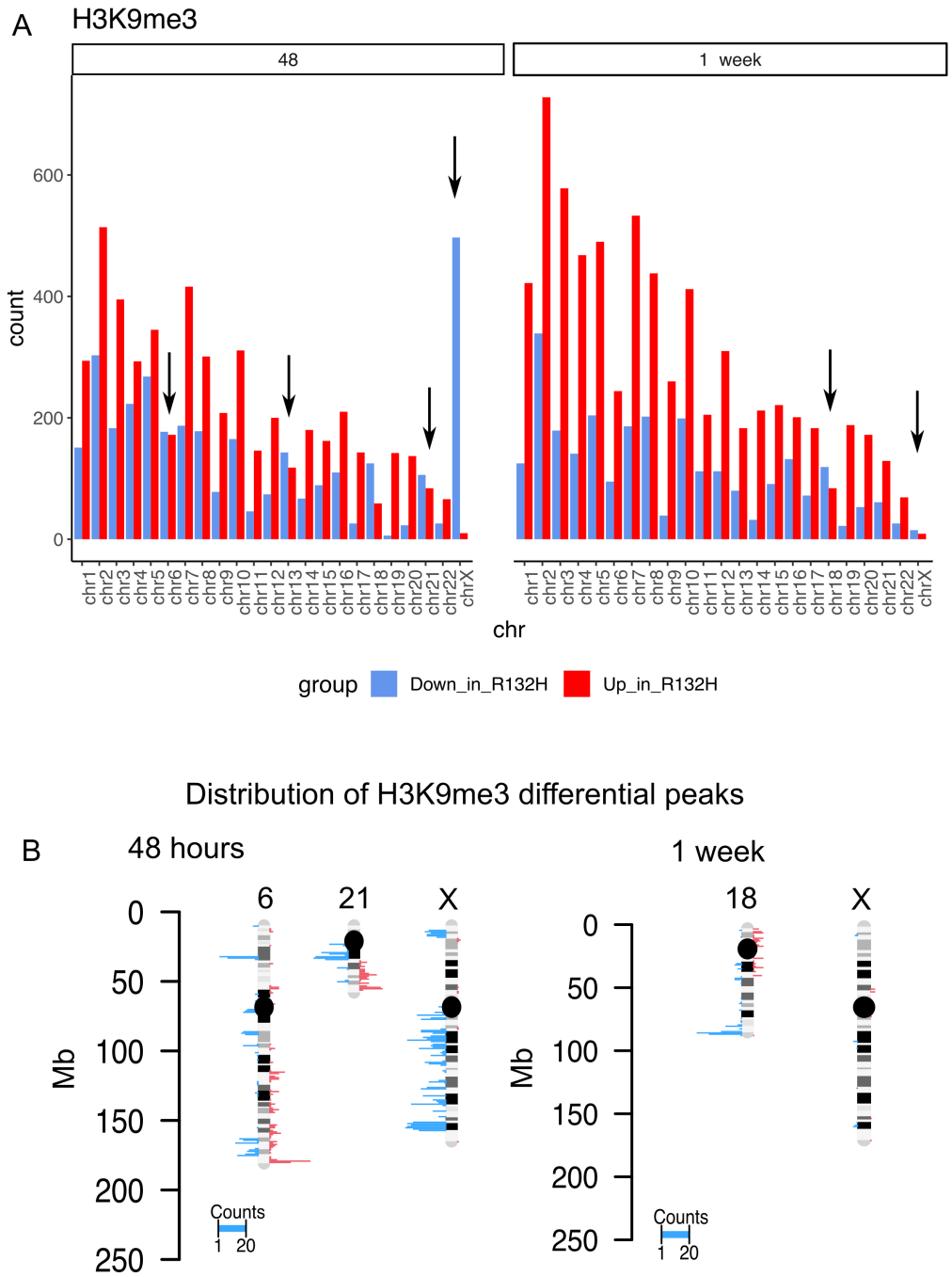


Figure 6.3 Distribution of H3K9me3 differential peaks highlights subset of loci with decreased methylation. *A) Bar plot showing counts of differential peaks increased (red) or decreased (blue) in R132H Ctrl cells compared to WT Ctrl cells in each dataset. B) Focus on chromosomes showing loss of H3K9me3. Chromosomes ideograms were plotted using the chromPlot R package. Red histograms*

correspond to loci bearing increased H3K9me3, while blue ones show decreased methylation (Verdugo & Orostica, 2020).

Visualization of differential peaks distribution along chromosomes hints at a loss of H3K9me3 around heterochromatic loci across chromosome X, whereas in autosomal chromosomes the loss is more localized, sometimes around telomeric (chromosome 6, 48 hours; chromosome 18, one week dataset) or centromeric regions, such as chromosome 21 in the 48 hours dataset (Figure 6.3B). This peculiar loss of H3K9me3 is further discussed in Chapter 9.

Thus, in line with literature findings, H3K9me3 is generally increased in IDH1 R132H cells compared to its WT counterpart, but a smaller subset of loci loses this methylation mark.

6.3 Assessment of CTCF binding in R132H IDH1 mutated cells

CTCF binding on its target loci is methylation-sensitive, and DNA methylation is greatly increased upon IDH1 mutation (Bell & Felsenfeld, 2000; William A Flavahan et al., 2016; Hark et al., 2000; Turcan et al., 2012). As a consequence, CTCF binding in R132H cells might also be impaired. ChIP-Sequencing of CTCF in WT or R132H cells shows a consistent loss of CTCF binding in the latter (Figure 6.4). This effect is more prominent in the 48 hours dataset, where we detect 2692 peaks significantly lost in R132H control cells compared to their WT counterpart, against only 46 peaks identified as significantly enriched (Figure 6.4A). A similar trend is present in the one week dataset: CTCF binding is still greatly reduced in R132H cells, as 1630 peaks are significantly lost, while 519 are differentially enriched. Distribution of differentially enriched peaks for every chromosome highlights a greater loss of CTCF along chromosome 1, perhaps due to its large size (Figure 6.4C). Notably, in both datasets CTCF transcript levels are unchanged between WT and R132H cells (48 hours: Log2FC = -0.016, P.adj = 0.97; one week: Log2FC = 0.2, P.adj = 0.86).

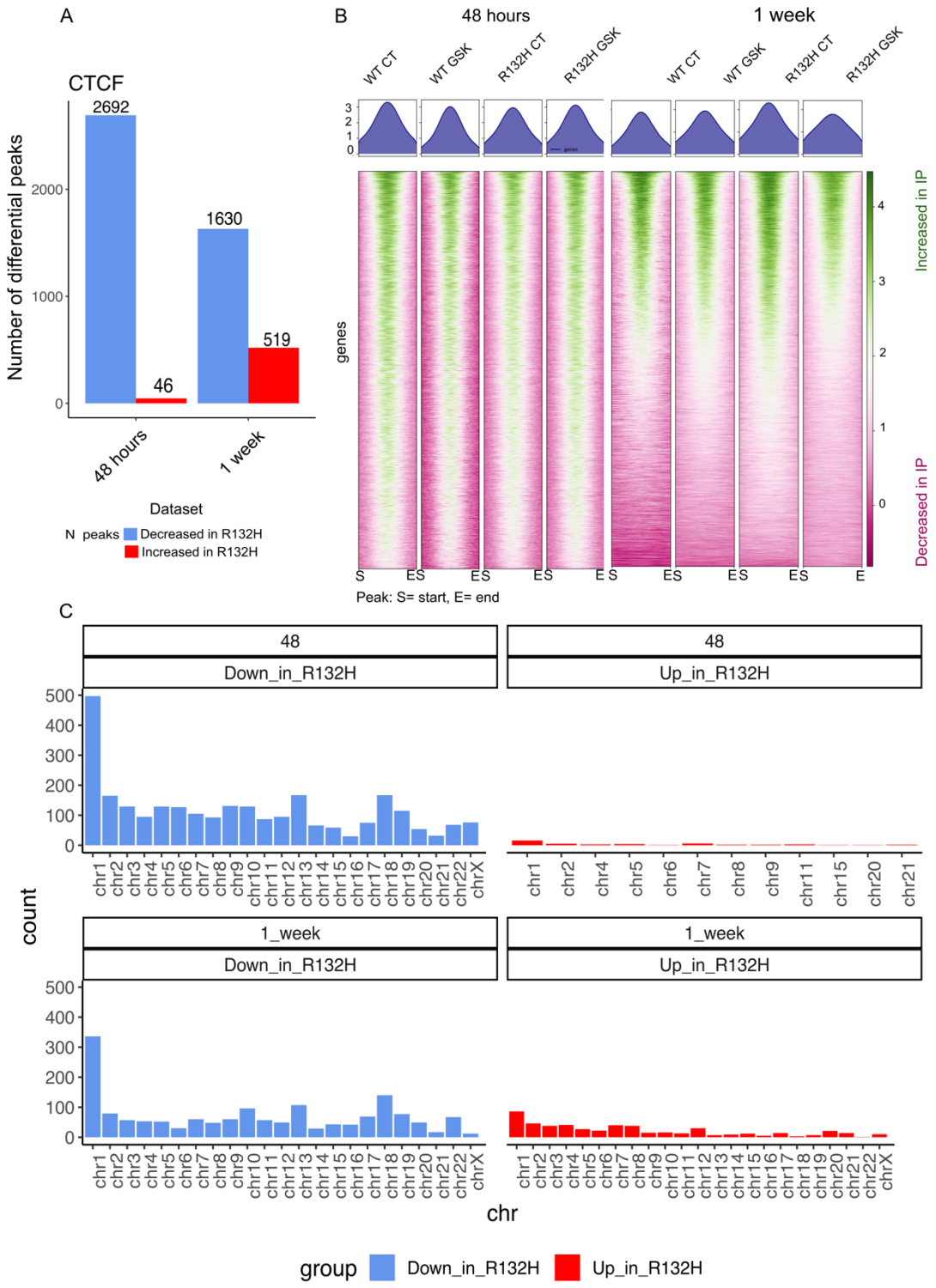


Figure 6.4 Genome-wide loss of CTCF in R132H cells. *A) Bar plot showing the number of differential CTCF peaks enriched or decreased in R132H control cells compared to WT control cells. Narrow peaks were called using MACS2, then statistical analysis was performed using HOMER, as described in the Methods section. B) Heatmap visualization of log2 fold change of IP over Input samples, obtained as described in Figure 6.2. Purple cells represent genomic loci where the Input signal is higher than the*

IP, whereas green cells represent genomic loci enriched in IP. This allows for comparison of CTCF ChIP-Seq signal genome-wide across all four experimental samples, for N=3 (48 hours) or N=2 (one week) biological replicates. C) Bar plot showing chromosomal distribution of CTCF differentially enriched/decreased peaks in both datasets.

Annotation of genomic loci showing significant loss of CTCF binding reveals most peaks are falling around intronic or intergenic areas in both datasets, with only a relatively small number annotated to gene promoters (48 hours: 128 TSS-promoter peaks; one week: 75 TSS-promoter peaks; Figure 6.5). Intersection of the gene promoters showing significant loss with a list of differentially expressed genes from the respective RNASeq experiments revealed little overlap: of 128 promoters with decreased CTCF, only 10 genes are detected in RNASeq in the 48 hours dataset, while 17 genes out of 75 promoters are found at one week.

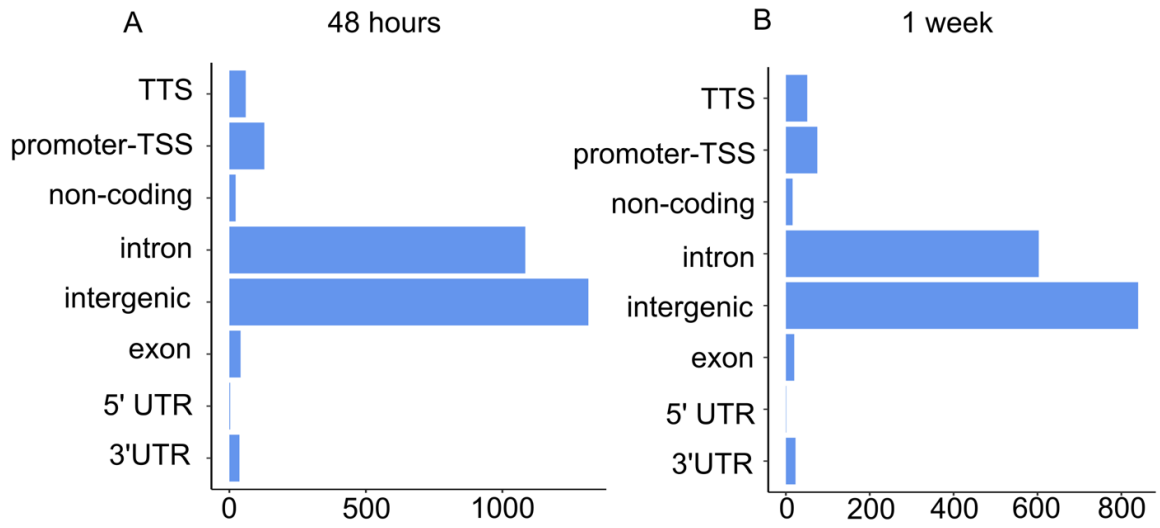


Figure 6.5. Annotation of CTCF loci showing a significant decrease in R132H cells. *Annotation of CTCF differential peaks in the 48 (A) or one week (B) dataset. Annotation was obtained with HOMER's AnnotatePeaks, then only peaks differentially decreased were selected for plotting. X axis indicates the number of peaks annotated to a specific genomic features.*

To investigate potential effects of CTCF loss on gene expression, regions corresponding to CTCF peaks lost in R132H control cells in the 48-hour dataset were given as an input for a Gene Ontology (GO) enrichment analysis using GREAT (http://great.stanford.edu/public/html/index.php). The tool attempts at assigning biological functions, such as GO terms, to non-coding regions of interest based on the annotations of its cis-interacting genes, building on the assumption that coding genes are often well-annotated,

whilst non-coding regions are not (Mclean et al., 2010). Such cis-interacting genes often coincide with those closest, or nearby, to the non-coding region of interest, as described on the tool's homepage. This gene to region-of-interest association occurs in two steps: first, a gene is assigned a "regulatory domain", both a proximal (5 kb upstream and 1 kb downstream of its transcription start site) and distal one (up to 1 Mb from its TSS), then if the genomic region of interest overlaps with these regulatory domains, that particular gene is considered to be associated with the region of interest (see "Association rule settings", Basal plus extension in the tool's homepage). All GREAT analyses in this thesis have been carried out using the default settings ("Basal plus extension").

The analysis returned a list of GO terms associated with lost CTCF peaks, summarized in Figure 6.6C. Notably, the majority of GOs relate to ion homeostasis and depolarization, potentially suggesting the interaction with genes coding for voltage-dependent ion channels. A minority of terms are also related to differentiation, or lack thereof: this would be in line with literature findings, often associating IDH1 mutation with an inhibition of cell differentiation (Lu et al., 2012; Modrek, Golub, Khan, Bready, Prado, Bowman, Deng, Zhang, Rocha, Raviram, Lazaris, Stafford, LeRoy, et al., 2017). Terms such as "negative regulation of osteoblast differentiation", "osteoblast development", "osteoclast differentiation" and "dentate gyrus development" are annotated to 32, 19, 15 and 22 regions, respectively (Figure 6.6C). The abundance of terms related to osteoblast development and differentiation, coupled with the presence of terms related to ion homeostasis and depolarization, could be a further indication that many of the regions where CTCF binding is lost in R132H cells are associated with genes coding for ion channels, perhaps calcium dependent.

Following intersection of CTCF peaks with the panel of differentially expressed genes detected in our RNA-Seq, 413 genes were selected for further analysis (Figure 6.6A). Their expression levels do not show any specific patterns: in spite of their proximity to CTCF binding sites lost,

some genes are greatly decreased, such as MAP2, while others, like ELMO1 (Engulfment and cell motility 1), are significantly increased in R132H cells (Figure 6.6B).

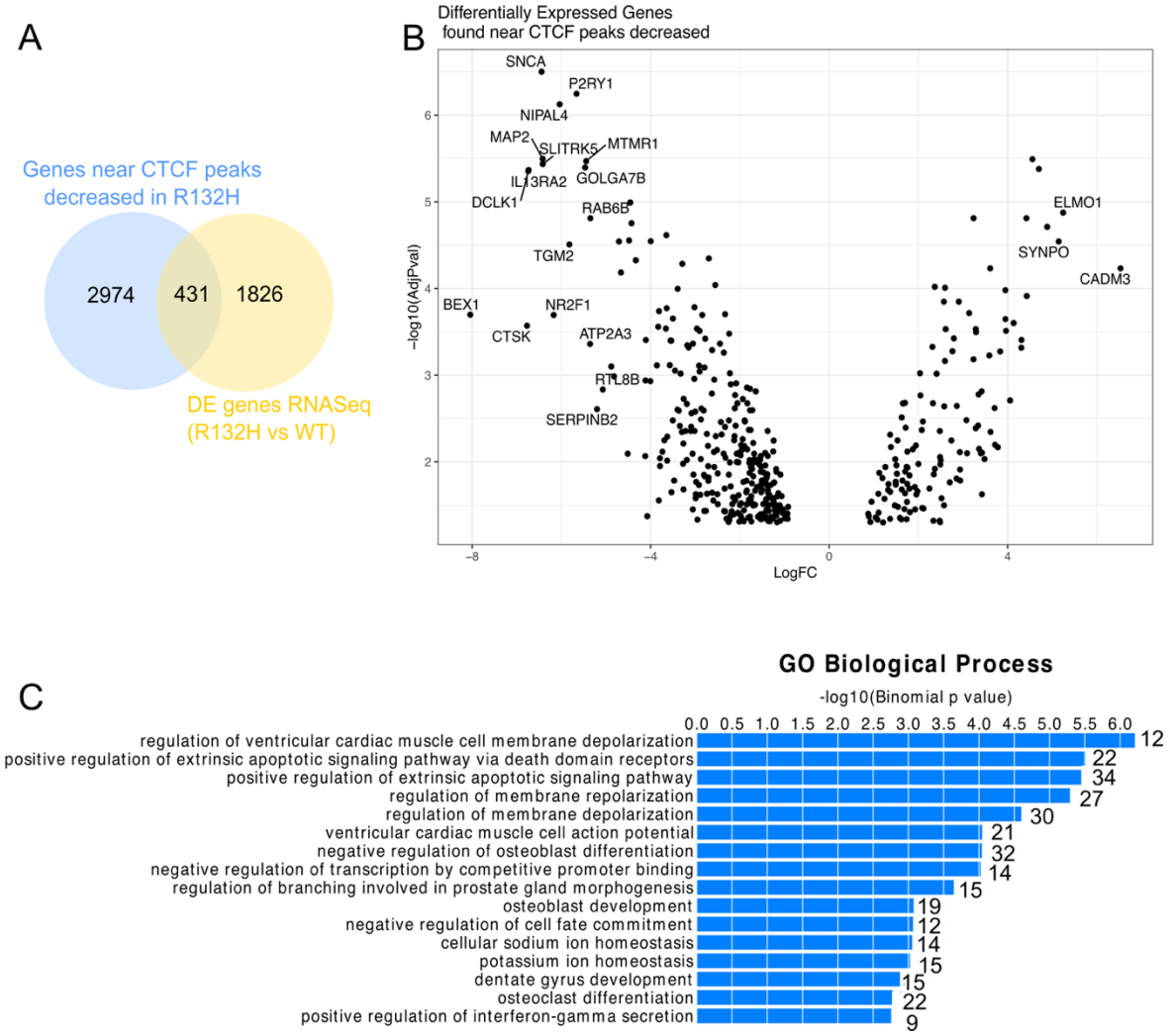


Figure 6.6 No clear link between loss of CTCF and nearby gene expression*. A) Venn diagram showing the intersection between differentially expressed genes detected in RNA-Seq (yellow) and genes annotated nearby CTCF peaks significantly decreased in R132H control cells compared to WT (blue). This annotation was part of the output from GREAT, which performs a gene ontology enrichment analysis of sequences of interest along the genome (in this case, differential CTCF peaks) based on the cis-interacting genes. B) Volcano plot showing the log2 fold change of 431 genes in common between RNA-Seq and GREAT annotation of genes near CTCF peaks lost. C) Gene ontology enrichment results, as calculated by GREAT, for CTCF peaks significantly decreased in R132H cells in the 48-hours dataset. Most significant GO terms were summarized in a bar plot and ordered based on their statistical significance (-log10 of their Binomial p.value).The number of peaks annotated with each term is printed on the top of each bar.*

Closer analysis of two genes with opposite expression levels, the upregulated *ELMO1* (Log2FC = 5.24; P.adj = 0.0000133) and downregulated *MAP2* (Log2FC = - 6.42; P.adj = 0.0000032), in IGV does not highlight any link between CTCF loss and gene expression. In

both genes, the CTCF binding site lost is within the coding region, and while the peak is lost in the 48 hours dataset, the alteration is not detected in the one week dataset (Figure 6.7).

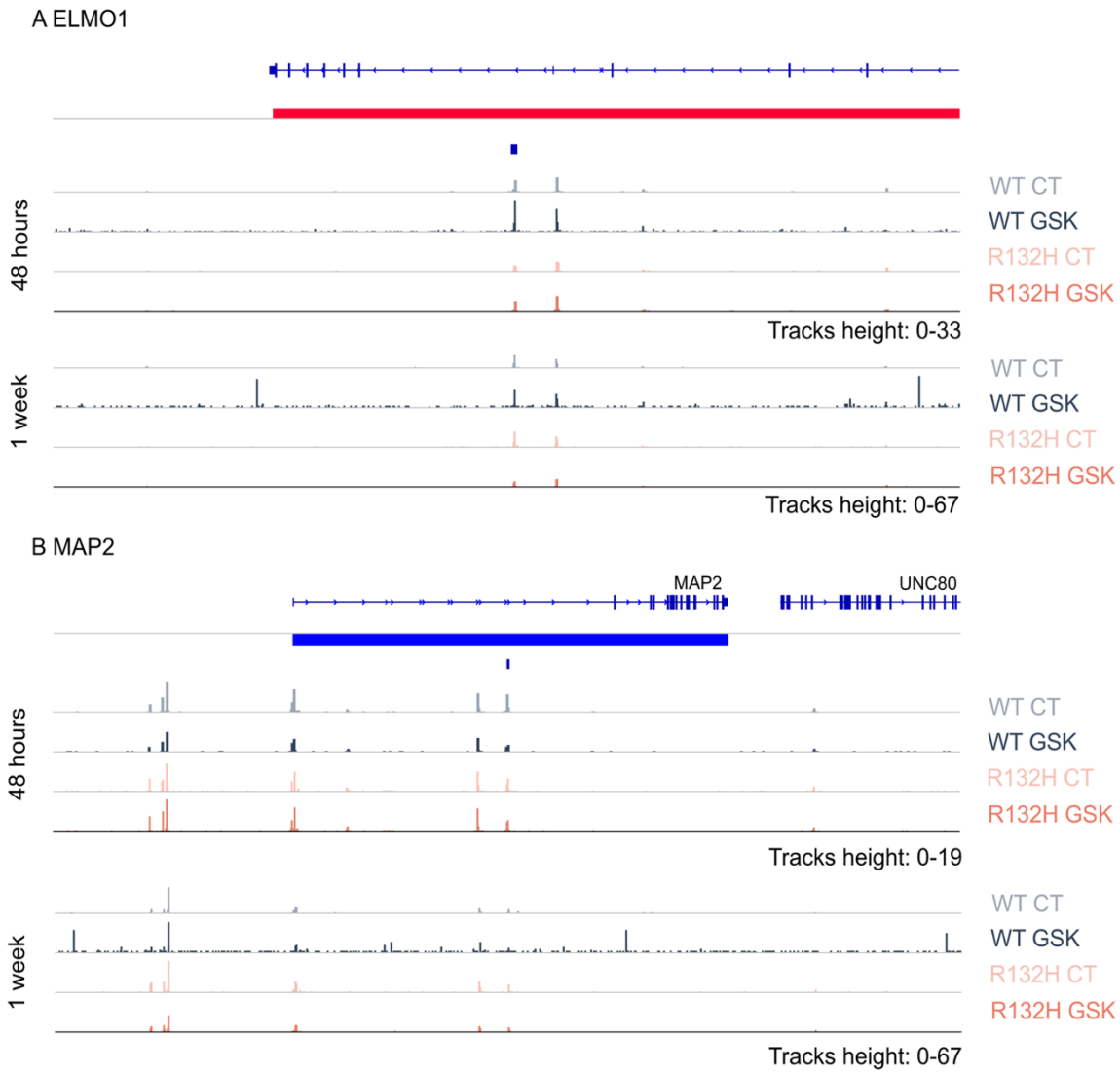


Figure 6.7 IGV tracks of genes with loss of CTCF binding but opposite expression levels. *Tracks show CTCF peaks distribution at 48 hours or one week for ELMO1 (A) or MAP2 (B). The lost CTCF binding site is indicated by the blue square. Expression level of each gene is represented by a red (Log2FC > 0) or blue (Log2FC < 0) rectangle underneath the gene structure. IGV tracks for each dataset were grouped and scaled to the same height, indicated underneath. Bigwig files between two (one week) or three (48 hours) biological replicates were merged and normalized to 1x coverage, as described in the Methods section. ELMO1 peak: Log2FC = -1.59, P.Adj =2.21e-22 ; MAP2 peak: Log2FC =-1.5 , P.adj =2.78e-20.*

To confirm the uncoupling between loss of CTCF binding and changes in gene expression, the same analysis was performed on the one week dataset. GREAT identified 2219 genes annotated near CTCF peaks lost in R132H cells, of which 92 were differentially expressed genes in the RNASeq dataset, and selected for further analysis (Figure 6.8A). Distribution of their log2 fold-change indicates the majority of genes in close proximity to CTCF peaks lost are downregulated

in R132H cells, even though a small group is upregulated (Figure 6.8B), indicating that CTCF binding might not impact, or perhaps not directly, gene expression.

Gene ontology categories associated with loss of CTCF binding in the one week dataset point towards a series of metabolic processes, mainly dealing with phospholipids production: "steroid metabolic process", "steroid biosynthetic process", "regulation of phospholipid metabolic process" are annotated to 61, 32 and 24 peaks, respectively (Figure 6.8C). These categories could also be connected to the "regulation of phosphatidylinositol-3 kinase (PI3K) activity" GO, associated with 20 loci. PI3K is activated downstream of Insulin or other types of Tyrosine kinase receptor signaling, a pathway generally decreased in R132H cells (Chapter 5), so these results could suggest a link between loss of CTCF binding and decrease in genes coding for proteins involved in this signaling pathway.

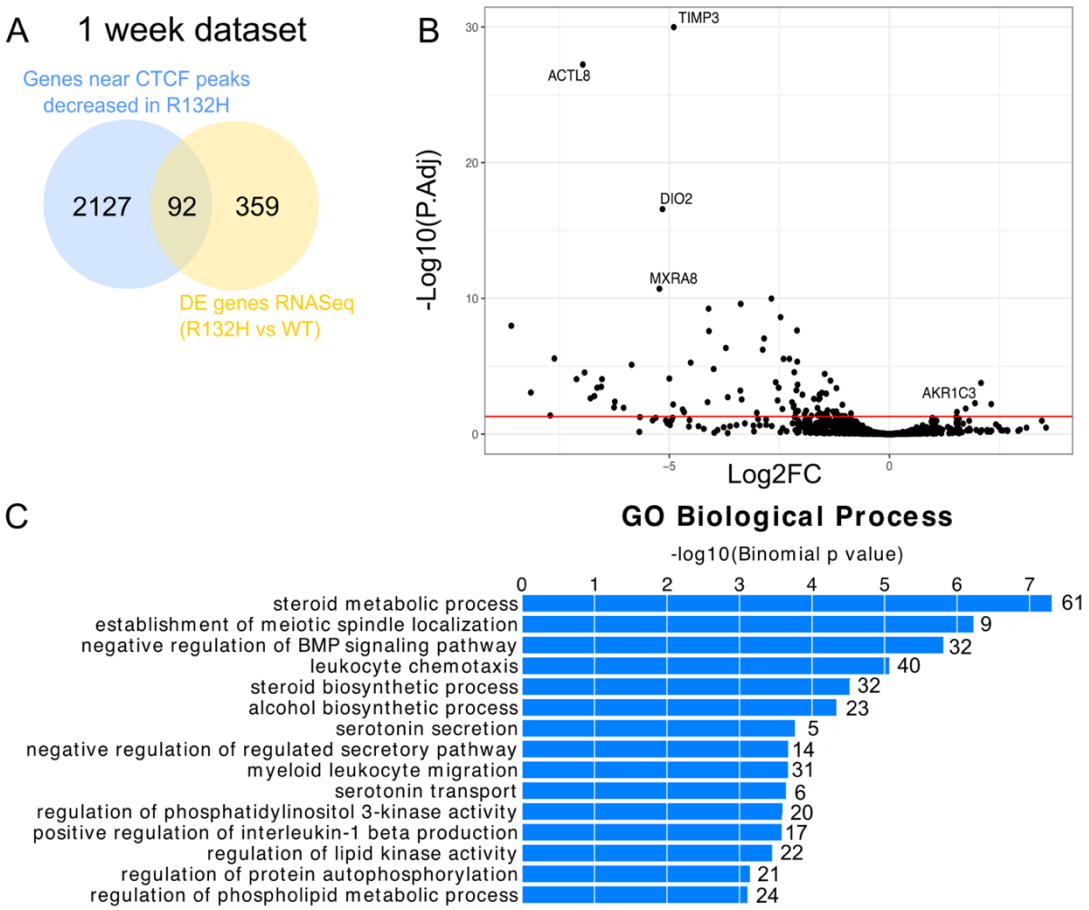


Figure 6.8 Absence of direct link between lost CTCF binding and gene expression is confirmed in one week dataset analysis. *A) Venn diagram showing the intersection between number of differentially expressed genes detected in RNASeq (yellow) and number of genes annotated nearby lost CTCF peaks, as found in analysis with GREAT (blue). B) Volcano plot of the 92 genes in common between groups in*

A). All genes above the red threshold show a significant p.adjusted. Names of the genes showing a greater fold-change or significance level are printed on top of the corresponding data point. C) Bar plot of the most significant gene ontologies associated with genes annotated nearby loci showing loss of CTCF binding. The length of the bar is indicative of their significance level, while the number printed on top of each bar indicates the number of loci annotated with that particular GO.

Two genes showing the most significant changes and opposite expression levels were selected for further visualization in IGV (Figure 6.9). *ACTL8* (Actin-like protein 8) shows decreased expression in R132H cells (Log2FC = -6.96; P.adj = 5e-28) and codes for a component of the cytoskeleton, while the upregulated *AKR1C3* (Aldo-Keto reductase family 1 member C3, Log2FC = 2.07, P.adj = 0.000001) codes for an enzyme converting aldehydes and ketones to alcohols, using NADH or NADPH as cofactor (Freitas et al., 2013; Plourde et al., 2009). In line with previous examples (Figure 6.7), CTCF binding is lost within the coding region of both genes, although this loss is present in both datasets (Figure 6.9A and 6.9B).

Thus, CTCF binding is greatly reduced in R132H cells compared to the WT, but whether this loss is linked to an effect on gene expression remains unclear, as genes close to CTCF peaks lost can show either an increase or a decrease in their expression levels.

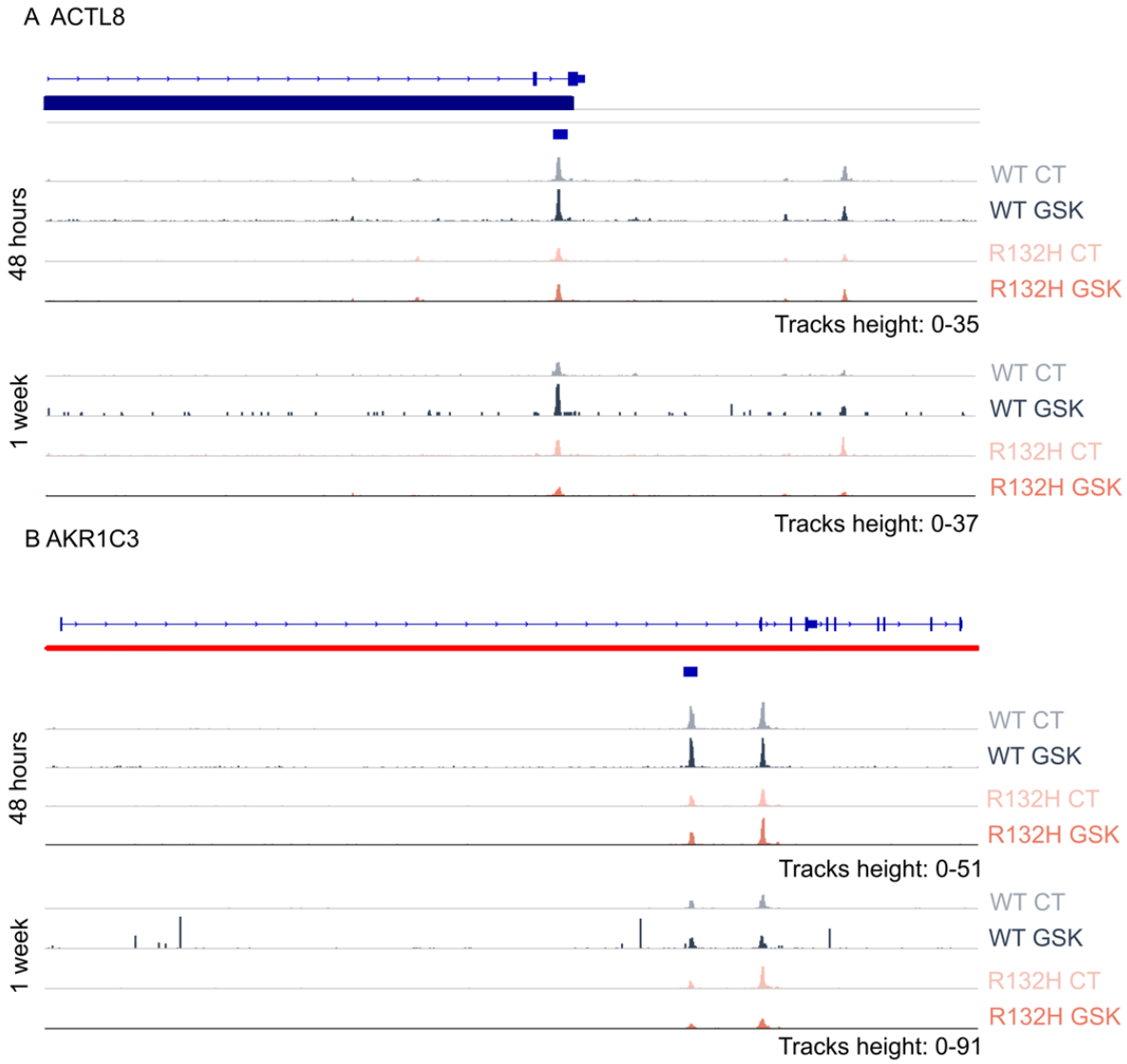


Figure 6.9 IGV tracks of selected genes from GREAT analysis on one week dataset*. A) CTCF distribution of ACTL8 in both ChIPSeq datasets. Peak of interest is indicated by a blue box (Log2FC = -2.02, P.adj = 0.00001). IGV tracks from the same dataset were grouped and scaled to the same height, indicated at the bottom of each panel. B) CTCF distribution of AKR1C3 in 48 hours and one week dataset. As in A), the peak of interest is marked by a blue box (Log2FC = -1.97, P.adj = 0.0002). Bigwig files were obtained as in Figure 6.7 and described in Methods.*

6.4 No effect of GSK864 on CTCF or H3K9me3 levels in R132H cells

GSK864 treatment is successfully decreasing 2HG levels and inducing changes in gene expression of R132H cells, specifically after a longer treatment (Chapters 4 and 5). Given that 2HG inhibits Histone and DNA demethylases, and that CTCF binding is methylation-sensitive, we asked whether treatment with GSK864 was affecting in any way H3K9me3 or CTCF levels (Bell & Felsenfeld, 2000; Hark et al., 2000; Lu et al., 2012).

At the timepoints tested, we detected very little effect of GSK864 on H3K9me3 (Figure 6.10A). Indeed, when comparing differential peaks between R132H control and GSK864-treated cells, only 292 peaks were significantly changing after 48 hours, of which 46 decreasing and 246 increasing upon GSK864 treatment. Fewer changes were registered after one week: only a total of 36 peaks, of which 14 decreasing and 22 increasing in R132H treated with GSK864. Perhaps this could be due to turnover timings of histone methylation: maybe the greatest impact would be seen either soon after drug administration, or after an even longer treatment with the drug.

GSK864 did not restore CTCF binding in R132H cells, although greater changes were detected after a longer treatment with the drug: 335 peaks, of which 172 upregulated, while 163 downregulated, were significantly different between R132H control and treated cells after a week, compared to the six peaks changing after 48 hours (Figure 6.10B). Interestingly, gene expression was also affected to a greater extent after a longer treatment, suggesting that, while the metabolic defect caused by IDH1 mutation is readily reversible after 48 hours, longer timings have to be taken into account to see changes, albeit modest, in gene expression or the epigenome.

Notably, GSK864 greatly alters CTCF binding in WT cells: 12155 and 3957 peaks were significantly decreased upon a treatment of 48 hours or one week, respectively. This effect seems to be restricted to CTCF binding, as gene expression and H3K9me3 levels have remained largely unchanged: no significantly different genes were detected in either datasets (Chapter 5, Tables 9 and 10), and only 68 (48 hours) or 2 (one week) H3K9me3 peaks are differentially enriched (Figure 6.10A). The extremely high number of differential CTCF peaks might indicate a nonspecific effect of GSK864, or that perhaps the immunoprecipitation (IP) experiment did not have a high yield. However, analyses of differential peaks were carried out on merged files from N=3 biological replicates at 48 hours and N=2 at one week, so even if an IP in one of the replicates had a lower yield in the WT GSK sample, this effect should be compensated by merging the replicates together.

Results show that, while correcting the metabolic defect after 48 hours and inducing gene expression changes after a week, GSK864 treatment at the concentration and timepoints tested does not rescue the epigenetic alterations induced by IDH1 mutation.

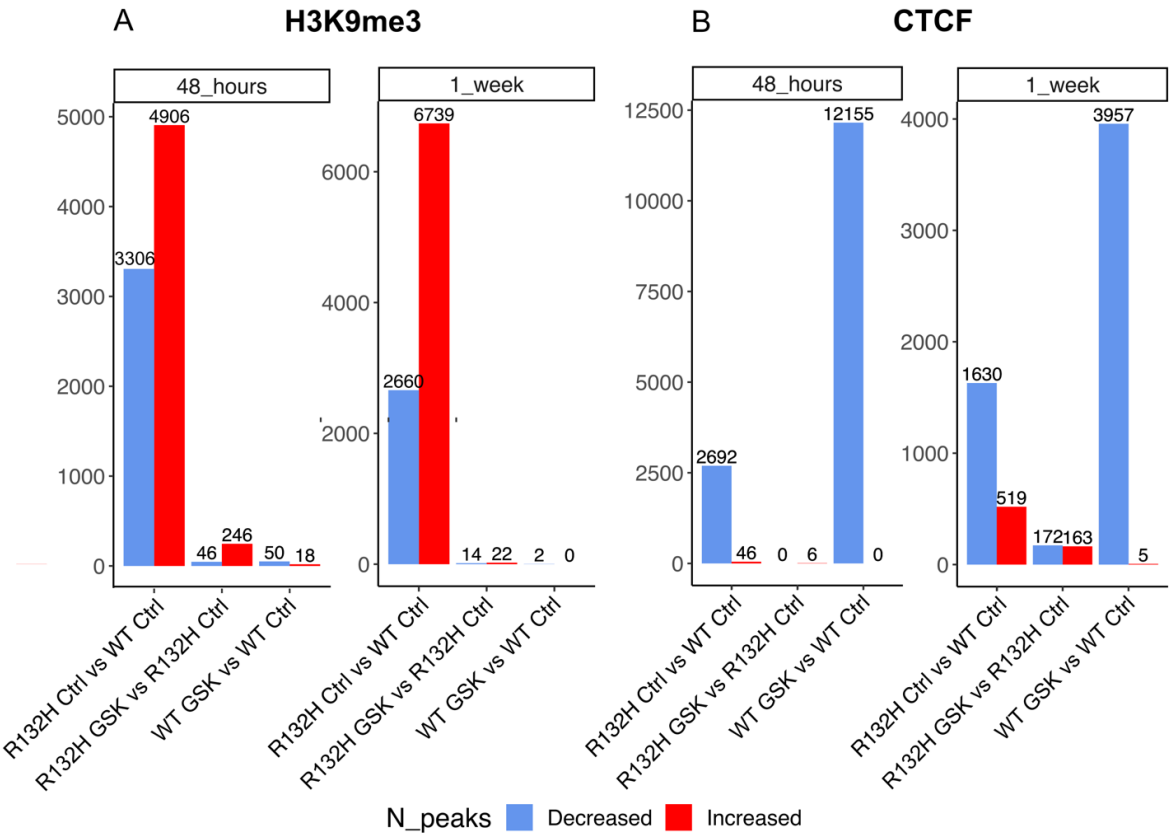


Figure 6.10 . Effect of GSK864 treatment on CTCF or H3K9me3 levels. *Bar plot representing the number of differentially enriched H3K9me3 (A) or CTCF (B) peaks decreased (blue) or increased (red) in the comparisons on the x axis in the two different datasets. Differential peaks were calculated using HOMER's getDifferentialPeaks, as detailed in the Methods section (Heinz et al., 2010)*

6.5 Discussion

Analysis of the epigenetic state of IDH1 mutated cells revealed general changes in histone methylation levels, reflected in a genome-wide increase of H3K9me3 and a widespread decrease of CTCF binding. However, none of these changes were reverted by a short (48 hours) or a longer treatment (one week) with an IDH1 inhibitor.

A striking common feature is the irreversibility of such modifications, as a one-week treatment with an IDH1 inhibitor was not sufficient to revert the phenotype. Similar results have been described before, and authors suggested a temporal factor that must be taken into account: the longer the mutation has been established, the more difficult it is to rescue the epigenetic defect (Turcan et al., 2018). Perhaps, the very nature of our cells, a commercially available CRISPR-edited U87 IDH1 R132H cell line, has had the mutation for so long that its impact is virtually irreversible. This might be a disadvantage when it comes to finding an adequate treatment, as it could mean that a long-term treatment or a higher concentration of drug might be required to see some epigenetic changes, even though 48 hours with 0.5 μM of GSK864 were enough to correct the main metabolic defect, as 2HG levels were greatly reduced (Chapter 4). On the other hand, this cell model might also recapitulate better what happens in a patient affected by glioma, as the IDH1 mutation is considered to be one of the earliest events, thus establishing all these epigenetic changes early on, and causing them to be difficult to rescue (Johnson et al., 2014). Interestingly, the longer treatment induced more changes in terms of gene expression and CTCF binding, even though the number of differential peaks was still modest. Another analysis, combining GREAT GO annotation with RNASeq results and visualization on IGV, could be carried out to test whether this effect is only a function of time or whether there are any commonalities between CTCF changing peaks, especially the loci where CTCF binding is increasing after GSK864 treatment, and changes in gene expression.

GSK864 treatment, while changing gene expression after a week and 2HG levels after 48 hours, does not seem to impact on cell viability or morphology, at least at the timepoints tested. To test whether the relatively modest change detected on the epigenome is due to a temporal component, one might try treating cells for an even longer period, mimicking the therapeutic regimes of chemotherapy, then reassess H3K9me3 and CTCF levels with ChIP-Seq.

Another possibility might be that targeting 2HG production is not an effective way to treat glioma, especially if the epigenetics defects indirectly caused are a lot more difficult to rescue compared to the metabolic one, once they have been established. A detail that needs to be taken into account is the timing of histone demethylases turnover. Several different KDMs are present in the cell, but since their affinity to α-Ketoglutarate varies from protein to protein, the extent of 2HG inhibition will be different (Chowdhury et al., 2011). If the extent of the inhibition is different, then we could hypothesize that some KDMs might need to be ubiquitinated, degraded, and new copies need to be synthesized, while for others perhaps the binding with 2HG is more readily reversible, and lowering 2HG levels might be sufficient for their activity to restart. The time taken for each KDM to revert the inhibition will then impact on the time required to detect changes in the epigenome, but since each KDM targets a specific lysine modification (Table 11), predicting a suitable time for GSK864 treatment becomes complicated. A proteomics experiment could be done to understand what the protein levels of KDMs family members are in R132H cells compared to the WT, before and after GSK864 treatment. While not addressing the problem of a reversible inhibition (eg., the protein is present, but not functional due to 2HG), it would at least show if new demethylases need to be synthesized or not in order to see changes in histone demethylation.

Recent studies have suggested that decreased CTCF binding in IDH1 mutant cells could cause loss of insulation between otherwise separated DNA domains, further contributing to aberrant gene expression (William A Flavahan et al., 2019a, 2016). Validation of this hypothesis in our model will be discussed in Chapter 8, but analysis of differentially expressed genes found in proximity to loci that have lost CTCF binding revealed no clear influence of CTCF on gene expression: some genes showed increased levels, while others were decreasing. However, looking at these two variables without any data on DNA methylation makes results interpretation difficult, as IDH1 mutation is known to induce DNA hypermethylation and

CTCF binding is CpG methylation-sensitive (Bell & Felsenfeld, 2000; Turcan et al., 2018, 2012) (see note 6.B).

Results from H3K9me3 ChIP-Seq, besides confirming a genome-wide histone hypermethylation, have also highlighted a localised loss of this modification in R132H cells, mainly around chromosome X. Interestingly, while on chromosome X this loss is found throughout the chromosome, chromosomes 6, 18 and 21 lose this mark around heterochromatic, centromeric or telomeric regions (Figure 6.3B). Whether this affects the higher architecture of the nucleus or chromosome distribution within it has been further investigated in a separate set of experiments, described in Chapter 9.

Thus, in line with literature findings, IDH1 mutant cells show an altered epigenetic state, bearing genome-wide increased histone methylation (H3K9me3) and loss of CTCF binding. While the metabolic alteration is readily rescued by GSK864 treatment, the correction of such epigenetic alterations has proven to be much harder to revert, suggesting either an uncoupling between these changes on the epigenome, gene expression and metabolism or that a temporal component needs to be taken into account.

6.6 Note on the assessment of genome-wide DNA methylation levels in R132H cells with TAPS

IDH1 mutation indirectly causes DNA hypermethylation due to 2HG-mediated inhibition of α-Ketoglutarate dependent DNA demethylases (Raineri & Mellor, 2018). Furthermore, CTCF binding on its target DNA sites is thought to be CpG-methylation-sensitive (Bell & Felsenfeld, 2000; Hark et al., 2000). To assess methylation levels in our cellular model, we set up a collaboration with Prof. Kriaucionis laboratory to perform TET-assisted pyridine borane sequencing (TAPS) on WT and R132H IDH1 U87 cells following a week-long treatment with GSK864. At the time of writing, due to the Covid19 epidemic this experiment

has been seriously delayed and is currently on queue to be sequenced. Results from this dataset will be key in confirming whether the loss of CTCF detected corresponds to DNA sequences that have been hypermethylated in R132H cells. Furthermore, the high coverage of this sequencing experiment will help in defining the karyotype of these cell lines, defined as hypodiploid by the manufacturer. This information, which will be included in the paper describing results from this project, will help defining whether the unexpected loss of H3K9me3 in certain genomic loci is an artifact due to an altered karyotype or not.

7. Assessment of histone H3K27ac in IDH1 R132H cells using ChIP-Seq

Metabolomic analysis of U87 WT and R132H IDH1 cells has highlighted a two-fold increase of AcetylCoA in the latter (see Chapter 4, Figure 4.8). Furthermore, RNASeq has highlighted changes in gene expression between WT and R132H cells, even though between the 48 hours and one week dataset there is an apparent discrepancy in the number and the type of differentially expressed genes in the two genotypes, which we assume was due to prolonged culturing with DMSO. Given the results from our metabolomics analysis, the role of H3K27ac in lining active enhancers and our interest in mapping enhancers relative to the loci described in Flavahan et al., 2016 (described in detail in Chapter 8), we performed H3K27ac ChIP-Seq.

In this chapter, we assess whether the AcetylCoA increase is associated with higher histone acetylation levels and explore the relationship between H3K27ac-lined enhancers and gene expression in our experimental setup. AcetylCoA, when not metabolized in the TCA cycle, can be transported into the cytosol thanks to its conversion into citrate. Once citrate has reached the cytosol, ATP citrate lyase (ACLY) can revert citrate back into AcetylCoA, which can freely diffuse into the nucleus (Wellen et al., 2009a). Here, the presence of nuclear Pyruvate Dehydrogenase complex (PDC) contributes to de novo nuclear AcetylCoA synthesis. This metabolite is then used by acetyltransferases (HATs) as an Acetyl-donor to acetylate specific lysine residues on histones and other proteins (Nagaraj et al., 2017; Sutendra et al., 2014b; Wellen et al., 2009b).

Analysis of H3K27ac by ChIP-Seq suggests a redistribution of regions enriched with lysine 27 acetylation in R132H cells. Results from our 48 hours dataset indicate a high proportion of differentially enriched peaks (4957), as well as a consistent number of peaks decreased in R132H samples (3309). Longer culturing of cells leads to an even greater number of

significantly increased H3K27ac peaks (6964), compared to 1865 peaks decreased in R132H cells, suggesting that time might be a component influencing the re-distribution or *de novo* deposition of histone acetylation in these cells (Fig. 7.1A). In line with results from the metabolomic profiling, treatment with GSK864 did not impact H3K27ac levels in R132H cells at either timepoint. Indeed, only 10 peaks significantly change between R132H control and treated cells after 48 hours, and slightly more peaks change after a longer treatment (106) (Figure 7.1A). Furthermore, at both timepoints the statistically significant changes are all an increase of H3K27ac, suggesting that GSK864 action is not rescuing acetylation levels.

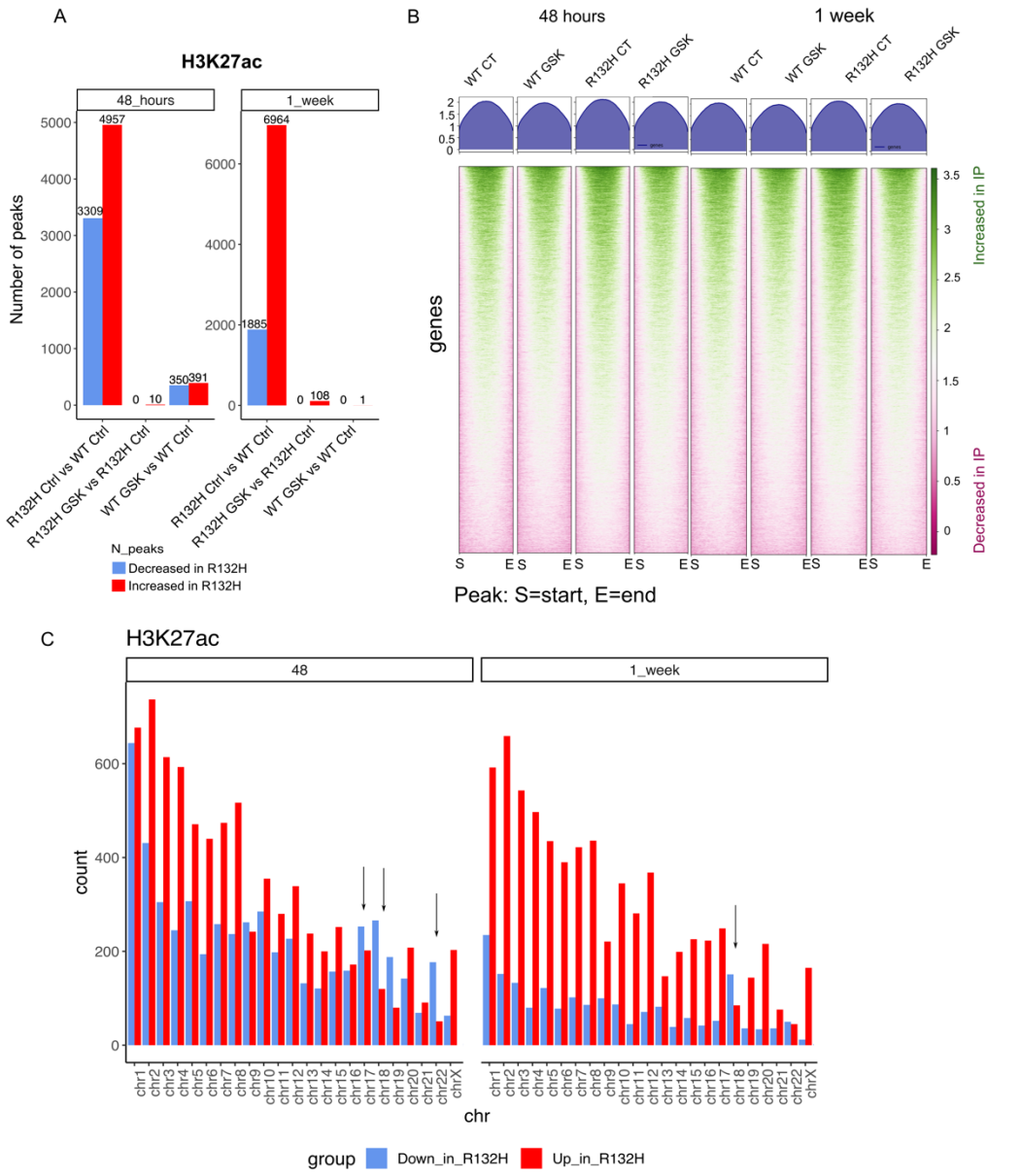

A
H3K27ac
48_hours
1_week
Number of peaks
R132H Ctrl vs WT Ctrl
R132H GSK vs R132H Ctrl
WT GSK vs WT Ctrl
N_peaks
Decreased in R132H
Increased in R132H
B
48 hours
1 week
WT CT
WT GSK
R132H CT
R132H GSK
genes
Increased in IP
Decreased in IP
Peak: S=start, E=end
C
H3K27ac
48
1_week
count
chr
group
Down_in_R132H
Up_in_R132H
D
48 hours
1 week
3' UTR
5' UTR
Exon
Intergenic
Intron
non-coding
Promoter-TSS
TTS
status
decreased_in_R132H
increased_in_R132H
stable

Figure 7.1 Rearrangement of H3K27ac in R132H cells. *A) Bar plot summarizing the number of differential peaks enriched (red) or decreased (blue) in R132H control cells compared to WT. Narrow peaks for each dataset were called using MACS2 on merged N=2 biological repeats, as described in Methods. Statistical significance was assessed using HOMER. B) Heatmap visualization of H3K27ac genome-wide distribution using the log2 fold change of IP values over Input for each experimental sample. Biological replicates reads (N=2) were merged, then normalized to 1x coverage using bamCoverage (Deeptools). Log2 of the fold change between IP and Input was calculated by running bigwigCompare on the normalized files, then results were visualized in a heatmap using computeMatrix followed by plotHeatmap, all part of Deeptools* (Ramírez, Ryan, Grüning, et al., 2016)*. Purple cells represent genomic loci where the Input signal is higher than the IP, whereas green cells represent genomic loci enriched in IP. C) Bar plot describing the number of decreased (blue) or enriched (red) H3K27ac loci along each chromosome, for both datasets. D) Bar plots representing the annotation of differentially enriched peaks in either the 48 hours (left) or one week (right) dataset. Bars for each genomic annotation are colored based on whether the signal coming from the peak was increased (blue), decreased (pink) or stable (black) in R132H cells.*

The distribution of significantly enriched or decreased H3K27ac peaks at 48 hours shows most chromosomes bearing a greater proportion of peaks increased, rather than decreased, with only a few exceptions: chromosome 1 has almost an equal amount of peaks lost and gained, while chromosomes 17, 18 and 22 show more loci with decreased acetylation (Figure 7.1C). Amongst these, only chromosome 18 maintains the same differential peaks distribution in the one week dataset, while along every other chromosome this histone modification is mainly enriched.

Thus, ChIP-Seq analysis suggests a dynamic rearrangement of the H3K27ac mark, with the majority of loci showing an increase in H3K27ac levels, and a smaller subset losing acetylation. Furthermore, treatment with GSK864 did not rescue this redistribution at any of the timepoints tested.

7.1 Potential uncoupling between gene expression and H3K27ac around promoters in R132H cells

Acetylation of Lysine 27 in Histone 3 has often been linked to active gene expression (Creyghton et al., 2010a; Rada-Iglesias et al., 2011). To investigate whether redistribution of H3K27ac in R132H had any effect on gene expression, H3K27ac peaks were called around the transcription start site (TSS) of RefSeq deposited transcripts in both our 48 hours and one week datasets. Differential peak analysis on the 48 hours dataset revealed 3309 peaks significantly decreasing around promoters in R132H control cells, and 4956 increasing (Figure 7.2A).

H3K27 acetylation around promoters at one week is mainly increasing (4872 peaks increasing; 439 decreasing, Figure 7.2B). This would potentially suggest gene expression changes in the two datasets.

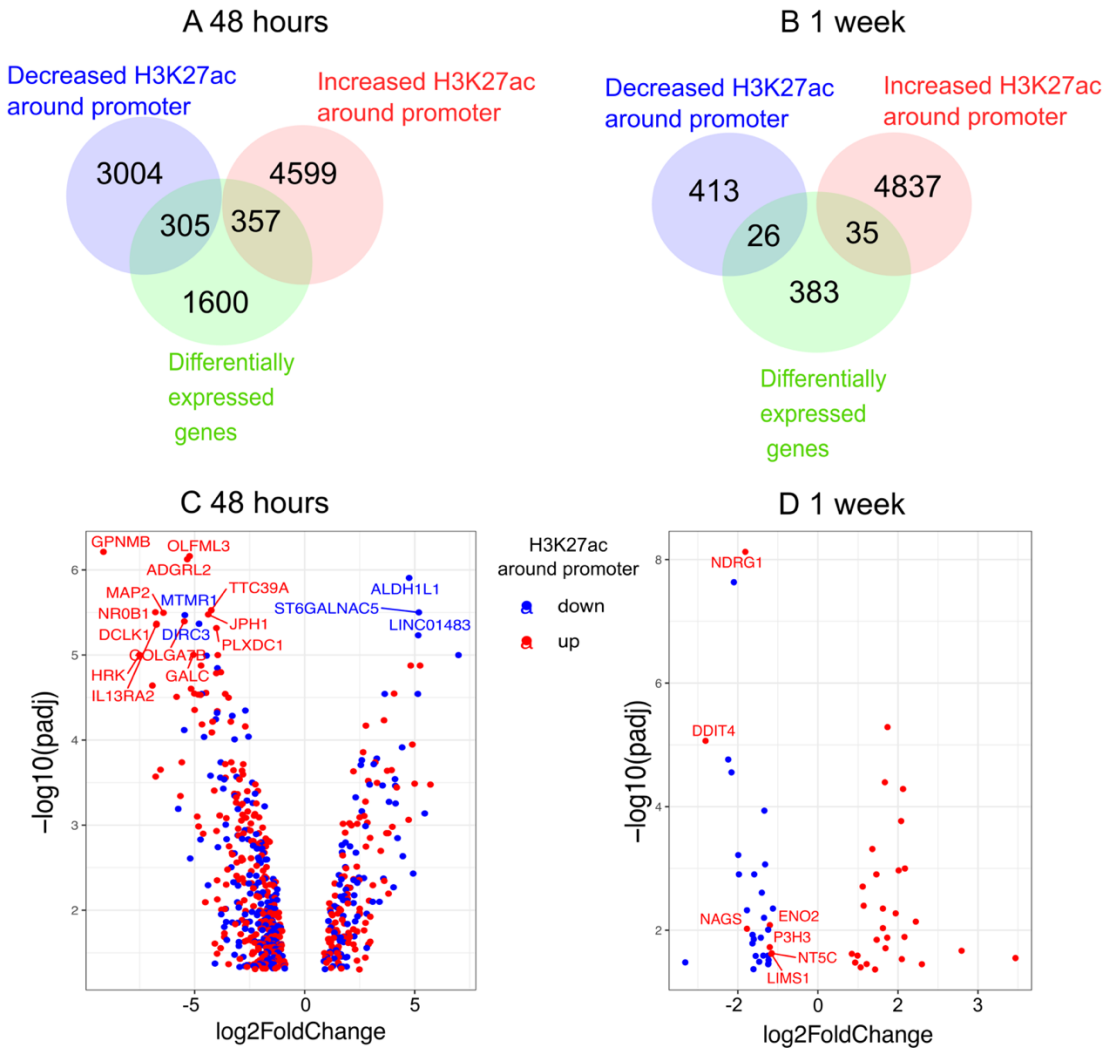


Figure 7.2 Analysis of H3K27ac levels around promoters and corresponding gene expression levels. *A and B) Venn diagram showing the intersection of differentially expressed genes detected in RNA-Seq (green), promoters of genes with decreased H3K27ac (blue), and promoters of genes with increased acetylation (red) in the 48 hours (A) or one week (B) datasets. C and D) Volcano plot showing gene expression (x axis, log2 fold-change) and statistical significance (y axis, -log10 of the adjusted P.value). The color of each data point reflects the status of H3K27ac at their promoters: red if acetylation is increased, blue if acetylation is decreased. Data are shown for the 48 hours (C) and the one week (B) datasets.*

Thus, we intersected the list of promoters whose levels of H3K27ac are significantly changing between R132H and WT cells with the differentially expressed genes detected in RNASeq, then focused on the differentially expressed genes whose promoters showed significant changes in

H3K27ac levels. The expectation would be to observe an increase of gene expression linked to promoters with increased H3K27ac, while a decrease in acetylation would correspond to a decrease in gene expression.

Unexpectedly, results show an apparent uncoupling of H3K27ac around promoters and gene expression, especially in the 48 hours dataset (Figure 7.2C and D). Indeed, genes whose levels are significantly higher in R132H cells, such as *ALDH1L1*, show a significant decrease in H3K27 acetylation around their promoter. On the contrary, downregulated genes such as *MAP2*, show an increase of K27ac on its promoter. While this effect is predominant in the 48 hours dataset, fewer examples are present in the one week dataset: all genes showing overexpression, as well as a few downregulated genes, are also showing increased levels of H3K27ac on their promoter. *ALDH1L1* and *ENO2* were selected for visualization on IGV. *ENO2* codes for a neuron-specific enzyme, phosphopyruvate hydratase, involved in glycolysis and promoting the conversion of 2-phospho-glycerate to phosphoenolpyruvate (Olsen et al., 2008). Upon IDH1 mutation, its expression is significantly decreased (Log2FC = -1.2; P.adj =0.008), while H3K27ac on its promoter is increased (Log2FC = 0.69; P.adj = 7.4e-20) (Figure 7.3A). Inspection of its locus on IGV shows that the acetylation increase on its promoter present at both timepoints, and even higher at 48 hours, even though at the earlier time point the change in gene expression is not statistically significant. *ALDH1L1* codes for a cytosolic enzyme involved in folate metabolism, 10-formyltetrahydrofolate dehydrogenase (Krupenko, Holmes, Tsybovsky, & Krupenko, 2015). This gene is significantly upregulated in R132H cells (Log2FC = 4.74; P.adj = 1.24e-06), but its promoter shows a significant decrease in H3K27ac (Log2FC = -1.25; P.adj= 7.32e-119). Indeed, acetylation levels around its TSS are consistently decreased in R132H cells at both timepoints, while its expression is significantly upregulated (Figure 7.3B). This gene could be of interest because of its involvement in Folate metabolism, a pathway linked to SAM biosynthesis, which represents the main source for methyl groups. Given the higher methylation levels in R132H cells, perhaps the lower acetylation around its

promoter could be a way to try to reduce its expression, even though its expression levels remain high.

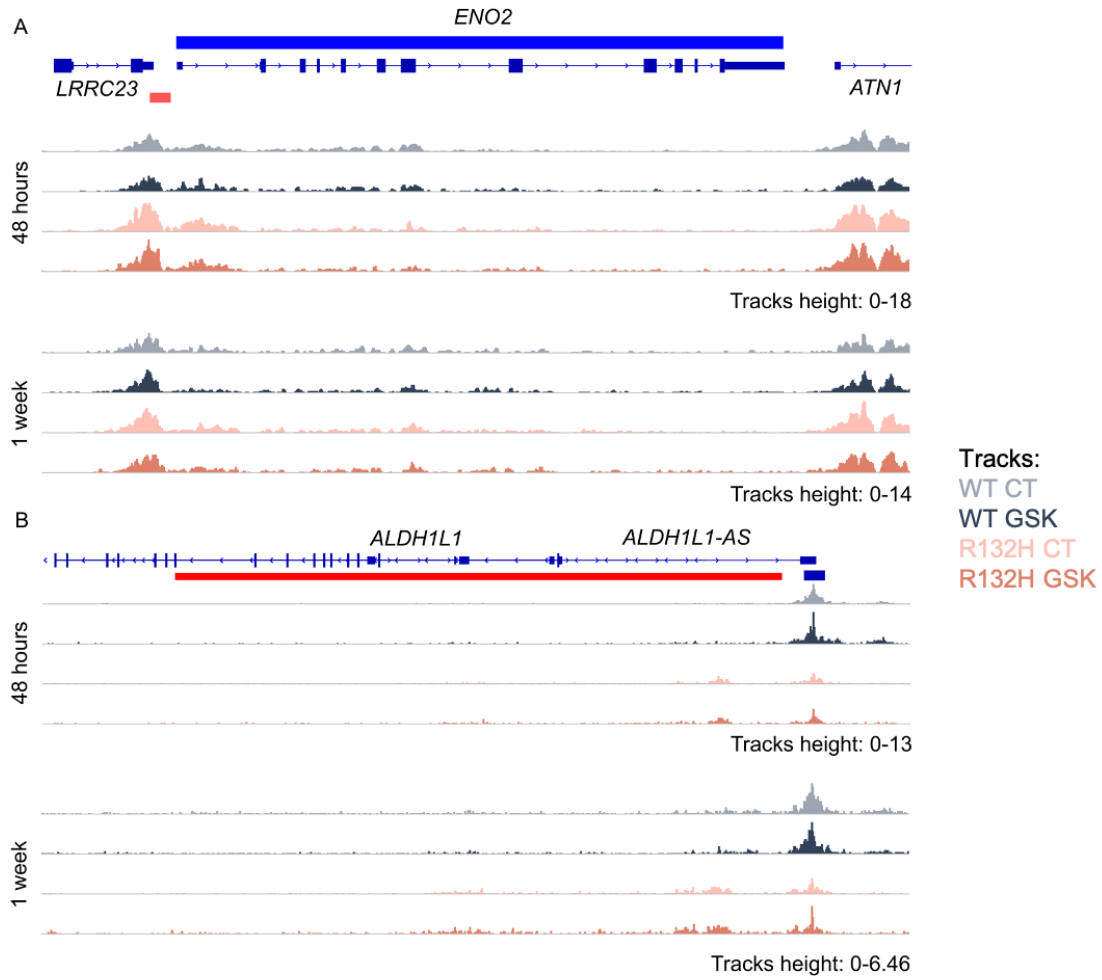


Figure 7.3 IGV tracks showing the ENO2 and ALDH1L1 loci. *A) IGV representation of the ENO2 locus, showing its RefSeq gene diagram, followed by 1x coverage normalized tracks for the 48 hours and one week datasets. A blue box corresponding to the expression level of the gene represents its downregulation, while the red box indicates the position of the significantly increased H3K27ac promoter peak. Tracks within the same dataset have been grouped and scaled to show the same height, indicated below. B) IGV visualization of the ALDH1L1 locus. Tracks are displayed as in 7.3A, but a red box here represent the gene's overexpression, while the blue box indicates the position of the significantly decreased H3K27ac promoter peak.*

Thus, analysis of promoter acetylation levels suggest an uncoupling between H3K27ac and gene expression, in contrast with what could be expected from literature.

7.2 Potential uncoupling between gene expression and H3K27ac around enhancers in R132H cells

Acetylation of Lysine 27 in Histone 3 has often been considered a histone modification to mark active enhancers (Creyghton et al., 2010a; Rada-Iglesias et al., 2011). To investigate whether redistribution of H3K27ac in R132H had any effect on enhancers and their targets' expression, an annotation of FANTOM5 transcribed human enhancers was intersected with a list of loci with differential binding of H3K27ac in both our datasets (M. Dalby, Rennie, & Andersson, 2017). Results indicated an almost identical number of differential peaks increased (1367) or decreased (1377) around enhancers in R132H cells in the 48-hour dataset (Fig. 7.4). A list of chromosome coordinates of these enhancer peaks was given as an input to GREAT, an online tool that predicts the function of loci of interest based on their nearest genes (Mclean et al., 2010). GO enrichment analysis with GREAT highlighted terms related to fatty acid metabolism ("positive regulation of fatty acid metabolic process"), hypoxia and extracellular matrix organization enriched in loci whose H3K27ac levels were increased around enhancers, while terms related to cell adhesion and immune response were more prominent around enhancers where H3K27ac was lost (Figure 7.4C).

To understand whether an increase or decrease of H3K27ac around an enhancer would affect gene expression, the list of genes associated by GREAT to the H3K27ac enhancers peaks was intersected with genes detected as differentially expressed in our RNA-Seq (Figure 7.4A). Interestingly, 305 genes (257 plus 48) are found near multiple enhancer peaks, some of which show increased H3K27ac, whilst others have lost H3K27ac (Figure 7.4, intersection of light red and light blue circles, or red, blue and green circles). Of these, 48 genes (the intersection between light red, blue and green circles) are differentially expressed and detected in RNASeq. However, because they are near enhancers that show either increased or decreased H3K27ac levels, it is difficult to understand which enhancer is influencing their expression. Furthermore, GREAT does not predict enhancers' targets, but rather looks for its closest genes, while

enhancers might influence genes located quite far from their own sequence (Mifsud et al., 2015). A total of 418 (217 near enhancer peaks with increased H3K27ac, 201 near enhancer peaks decreased) were selected for further analysis. The expectation would be to see an increase in expression of 217 genes with increased acetylation on the enhancer, and a decrease in levels of 201 genes near enhancers with decreased acetylation. These genes were plotted to visualize whether the change in expression was in line with the change in H3K27ac (Figure 7.4B).

Results suggest a potential uncoupling between H3K27ac and gene expression: genes near enhancers that have lost H3K27ac (coloured in light blue) can show both an increase in gene expression (*SYNPO*, *SUSD2*) or a decrease (*IL36B*, *TGM2*). The same is true for enhancers whose H3K27ac levels have increased (red): some of their nearby genes have decreased expression levels (*MAP2*, *GALC*), whilst others have shown an increase.

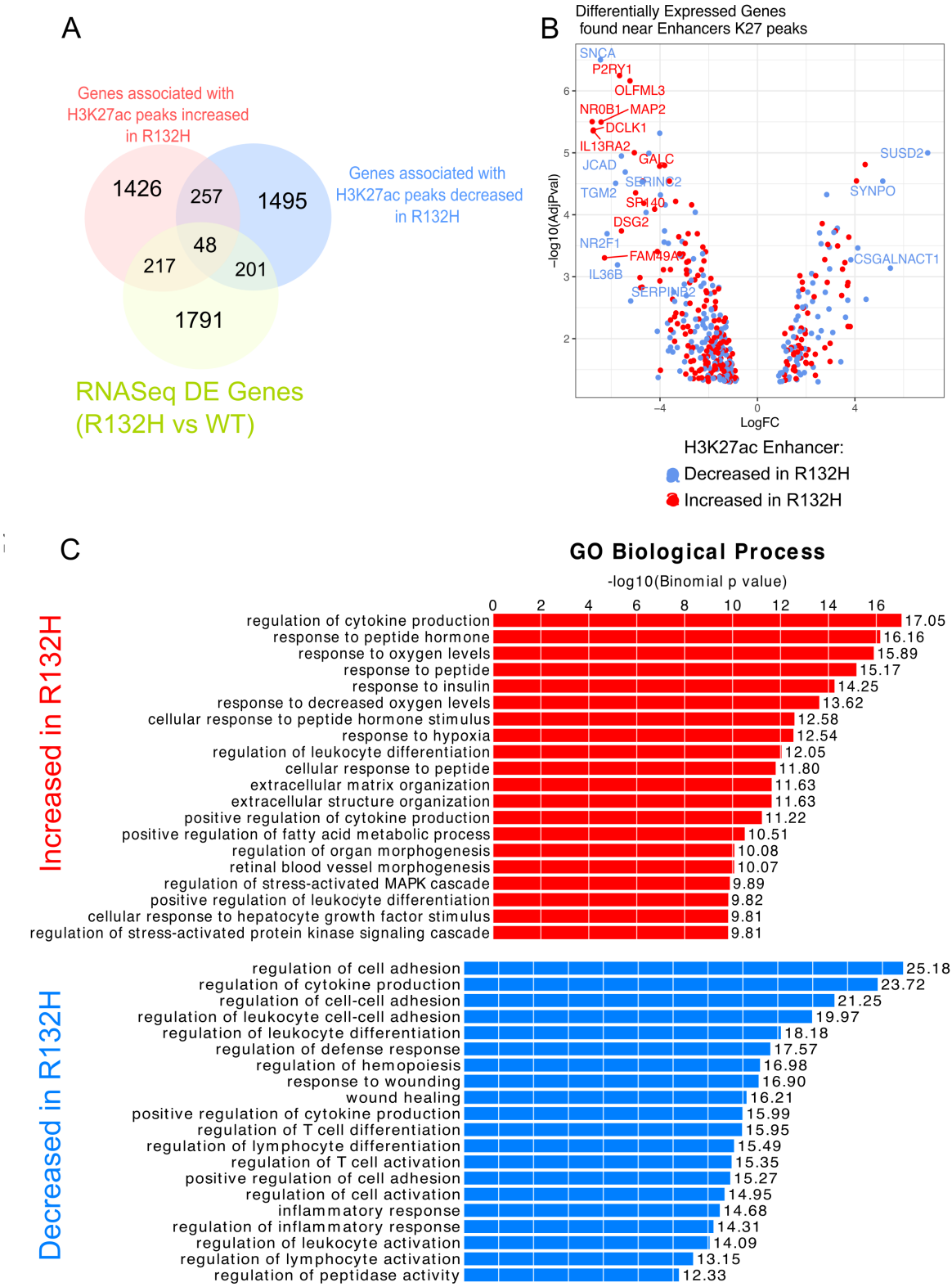


Figure 7.4 Analysis of H3K27ac changes across enhancers in the 48 hours reveals potential uncoupling with gene expression. *GREAT online tool was used for Gene Ontology (GO) enrichment analysis of H3K27ac enhancers peaks increased or decreased in R132H cells. To do this, GREAT takes as an input a list of loci corresponding to genomic regions of interest (in this case, H3K27ac differential peaks around enhancers), then finds the nearest genes and based on those genes performs the GO enrichment. As output, besides the GO results it is possible to retrieve the list of nearby genes upon*

which the analysis was made. A) Venn diagram showing the intersection of differentially expressed genes detected in RNA-Seq (green), genes near enhancers with decreased H3K27ac (light-blue), and genes near enhancers with increased acetylation (light-red). B) Volcano plot showing the log2 fold change of gene expression for 418 genes, 201 near enhancers that showed loss of H3K27ac (light -blue dots) and 217 near enhancers which gained acetylation (red dots). Genes showing a fold change greater than ± 4 have been annotated on the plot . C) Bar plot of the most significant GO annotation in each set, as obtained after analysis with GREAT.

Two examples were selected for further visualization in the IGV genome browser. Synaptopodin (*SYNPO*) is a protein associated with the actin cytoskeleton, often present at post-synaptic densities and dendritic spines (Mundel et al., 1997). In our dataset, *SYNPO* expression is increasing (Log2FC = 5.14, P.adj = 2.87e-05), even though its closest enhancer is losing acetylation (Log2FC = -1.72, P.adj = 7.24e-244). Indeed, IGV tracks show that the enhancer peak is located towards the 3' end of the gene (blue rectangle, Figure 7.5A), and while the peak is slightly decreased in R132H control compared to the WT at 48 hours, the difference is not present in the one week dataset (Figure 7.5A). MAP2, microtubule-associated protein 2, is a structural protein involved in microtubule assembly, and its levels mirror those of Synaptopodin: its gene expression is decreased (Log2FC = -6.41, P.adj = 3.2e-06) in spite of the presence of an enhancer with increased H3K27ac levels (Log2FC = 2.08; P.Adj < 0.0001) (Kalcheva et al., 1995). Indeed, a string of H3K27ac peaks are enriched around MAP2 promoter, with one of them being annotated as enhancer (Figure 7.5B, highlighted in yellow).

Thus, IDH1 mutation induces a rearrangement of the H3K27ac mark across the genome, including acetylation changes on enhancers. However, changes in lysine 27 acetylation levels on enhancers do not always correspond to the expected changes in gene expression in the 48 hours dataset, suggesting a potential uncoupling between the two.

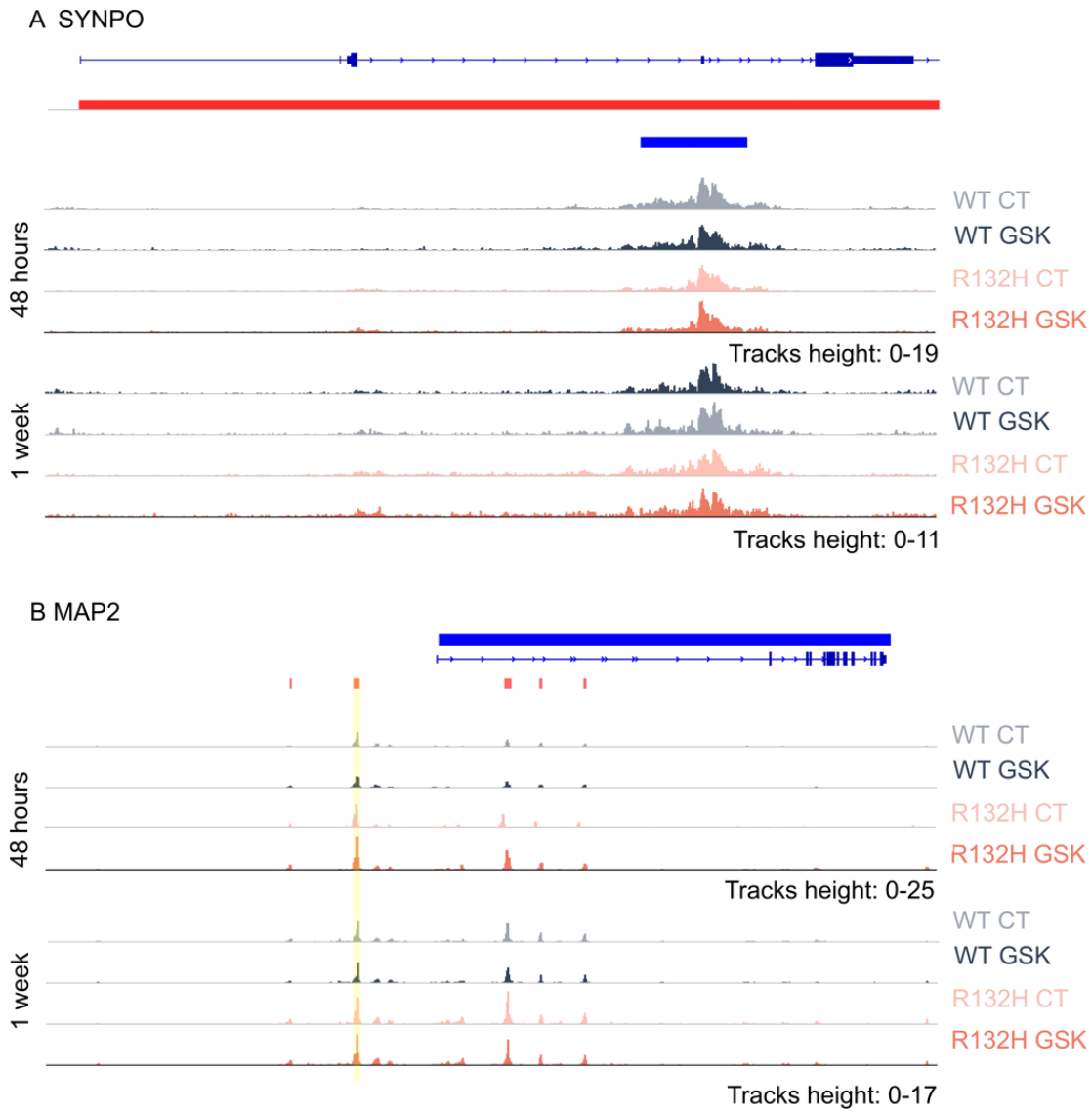


Figure 7.5 IGV tracks of SYNPO and MAP2. *A) SYNPO locus showing RefGene profile of SYNPO, gene expression level (red rectangle, indicating overexpression in R132H), enhancer peak location (blue rectangle), and panel of H3K27ac in all experimental samples. Tracks in each dataset have been grouped and scaled to the same height, indicated below. Bam files between 2 biological replicates were merged and normalized into 1x coverage bigwigs (see Methods). B) MAP2 locus representation. A blue box on top of the RefGene profile indicates the gene's downregulation in R132H cells. Red squares indicate H3K27ac differential peaks between R132H control and WT control cells, with the enhancer peak being highlighted in yellow. Tracks are shown for both 48 hours and one week datasets, and have been grouped and scaled to the common height, printed below. Bigwigs were obtained as described in A).*

7.3 Analysis of acetylation levels on enhancers at one week confirms poor association of enhancer acetylation and nearby gene expression

The same analysis, repeated on the one week dataset, gave similar results (Fig. 7.6A). A total of 1928 differentially enriched H3K27ac peaks were classified as enhancers by FANTOM5 annotation, while 697 enhancer loci had significantly lost H3K27ac in R132H cells. Gene ontology analysis finds similar terms between the two sets of loci at 48 hours and one week: terms related to oxygen levels, cell migration and collagen biosynthesis are associated to enhancers showing an increase of H3K27ac in R132H cells, while terms related to cell adhesion and the activation of immune response are linked to enhancer loci whose H3K27ac levels have decreased in cells bearing the IDH1 mutation (Fig 7.6C). Retrieval of the list of genes near these loci and subsequent intersection with differentially expressed genes from RNA-Seq finds 214 genes linked with enhancers that have gained H3K27ac and 122 near enhancers with decreased acetylation (Figure 7.6A). Expression of those 214 genes would be expected to increase, since the enhancer nearby has increased levels of H3K27ac, while that of the 122 genes near enhancers that are losing acetylation would be expected to decrease.

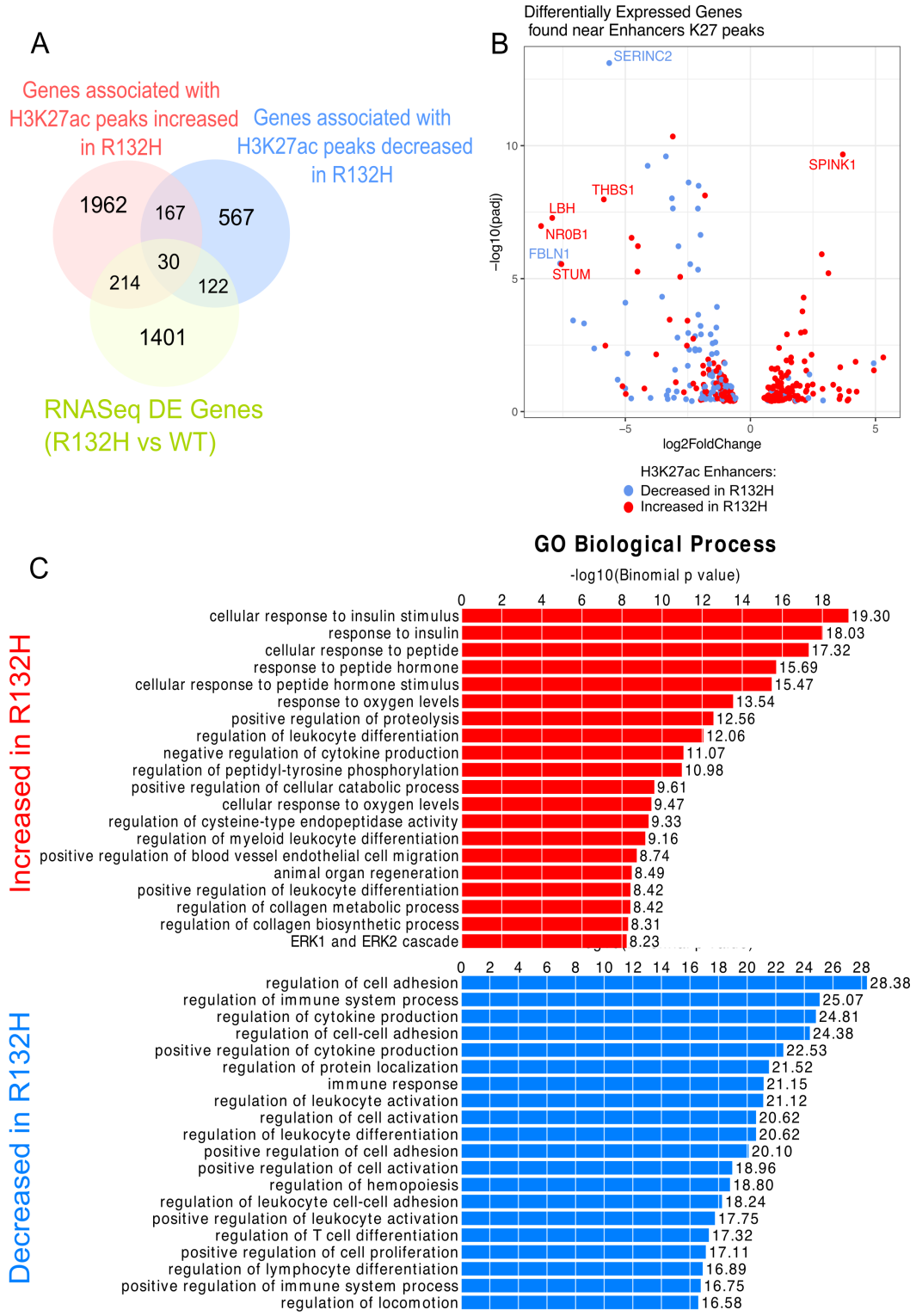


Figure 7.6 . Analysis of H3K27ac changes across enhancers in the one week dataset. *GREAT online tool was used for Gene Ontology (GO) enrichment analysis, as described in Figure 7.4. A) Venn diagram showing the intersection between differentially expressed genes detected in RNA-Seq (green), genes near enhancers with decreased acetylation (light-blue) and genes near enhancers with increased H3K27ac (light-red). B) Volcano plot showing log2 fold change of expression levels for 214 differentially expressed genes near enhancers showing increased H3K27ac (red dots) and 122 genes near enhancers with decreased acetylation (colored in blue). C) Significant GO terms associated with*

each set of loci, as calculated by GREAT. Statistical significance, indicated as -log10(P.adjusted) is printed at the top of each bar.

In line with the 48 hour dataset, we find that gene expression levels vary independently of enhancer acetylation, at least in this dataset (Figure 7.7B). While most genes showing an increase in gene expression are now linked to enhancers that are gaining acetylation, there is still a high number of genes whose expression is decreasing, while acetylation levels on the nearby enhancer are increasing (Figure 7.7B). One such example is *NR0B1*, coding for DAX1, a transcription repressor involved in endocrine tissue development and function (Iyer, Zhang, & McCabe, 2006). Expression levels of this gene are decreased in R132H control cells (Log2FC = -8.36; P.Adj = 1.05e-07) while acetylation around its close enhancer are increased (Log2FC = 1.76, P.Adj = 4.86e-46). Indeed, the enhancer peaks is located in the 3' end of the gene, and its increase is maintained in the one week dataset (Figure 7.7A).

The expected situation is detected at the *SPINK1* locus (Figure 7.7B). This gene codes for a serine protease inhibitor, its expression is upregulated in R132H cells (Log2FC = 3.69; P.adj = 2.14e-10) and its closest enhancer shows increased acetylation (Log2FC = 1.88; P.adj = 4.46e-73) (Witt et al., 2000). Two differentially enriched peaks are present around the SPINK1 locus, with the closest one being the annotated enhancer. Given this peak's proximity to the 5' of the JAKMIP2 gene, perhaps the actual target of this enhancer might be JAKMIP2, rather than SPINK1. Indeed, GREAT associates both genes to this enhancer, but since JAKMIP2 is not differentially expressed in the RNASeq, we cannot confirm any changes to its expression levels.



Figure 7.7 . IGV tracks of the NR0B1 and SPINK1 loci. *A) Description of NR0B1 locus showing the RefSeq diagram, together with its expression (Blue box, indicating downregulation) and K27ac levels (red box, indicating increase). Profile of H3K27ac distribution across all experimental groups in the two datasets are shown below. Tracks from the same dataset were grouped and scaled to the same height, listed below. Bigwigs were normalized to 1x coverage (see Methods). B) SPINK1 locus, organizes as described in A). Notably, two genes have been linked to the K27ac enhancer (highlighted in yellow), although only SPINK1 was overexpressed in RNASeq (red rectangle).*

Thus, H3K27ac levels are changing around a subset of enhancers in R132H cells. However, GREAT associates a locus of interest with its closest genes, but enhancers do not necessarily regulate their nearby genes, as a target can be found several kb away in the linear DNA sequence, and brought closer via higher order chromatin structures (Mifsud et al., 2015; Modrek, Golub, Khan, Bready, Prado, Bowman, Deng, Zhang, Rocha, Raviram, Lazaris, Stafford, Leroy, et al.,

2017). Hence, a second analysis, using a different enhancer annotation (EnhancerAtlas, which predicts a target gene for every annotated enhancer) has been carried out to improve robustness.

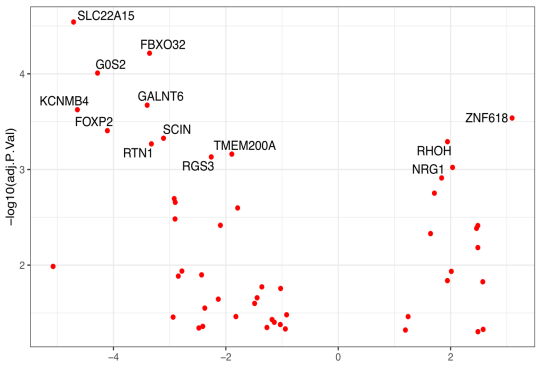
7.4 EnhancerAtlas annotation confirms uncoupling between gene expression and enhancer acetylation

Acetylation of Lysine 27 in Histone 3 around promoters highlighted an uncoupling between promoter acetylation and gene expression. To investigate whether redistribution of H3K27ac in R132H had any effect on enhancers, and to evaluate effects on their targets' expression, an annotation of human enhancers was intersected with a list of loci with differential binding of H3K27ac in both our datasets (M. Dalby et al., 2017). In this case, EnhancerAtlas2.0 (http://www.enhanceratlas.org/) was used to predict target genes for genomic regions of interest in U87 (T. Gao & Qian, 2020).

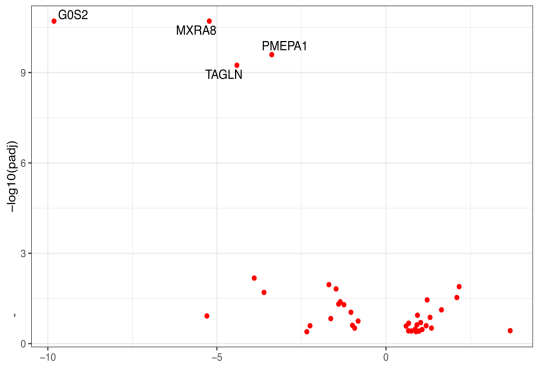
Results found 658 enhancer peaks presenting increased H3K27ac at 48 hours and 707 in the one week dataset. Nevertheless, the majority of the predicted target genes shows decreased gene expression compared to the WT, while only a smaller number of genes are increasing (Figure 7.4A and B). Three genes were selected for further inspection, based on their expression changes and significance levels: *RHOH* (Log2FC = 1.94, P.Adj = 0.0005), *FOXP2* (Log2FC = -4.1, P.Adj = 0.0003) and *TAGLN* (Log2FC = -4.4, P.Adj = 5.7e-10). All the genes selected show increased enhancer acetylation at both 48 hours and one week, yet two out of three are downregulated compared to the WT (Figure 7.4C, D and E). A possible explanation might be that these genes are still being actively expressed, thus their enhancer is marked with H3K27ac, even if their expression is lower compared to the WT. While explaining the apparent contradiction between a decrease in gene expression and presence of acetylation, it does not

fully justify the increase in K27ac on these enhancers compared to the WT counterpart (*RHOH* enhancer: Log2FC = 2.14, P.Adj = 4.1e-69; *FOXP2* enhancer: Log2FC = 2.23, P.Adj < 0.0001 ; *TAGLN* enhancer: Log2FC = 2.18, P.Adj =2.9e-61). It could also be that loci that were already acetylated are gaining even more acetylation as a way for the cell to dispose of excess AcetylCoA, but this increased acetylation on their enhancer is not what drives the changes in expression levels, which are following another cue.

A Differentially Expressed Genes near K27ac Enhancer peaks (48 hours)

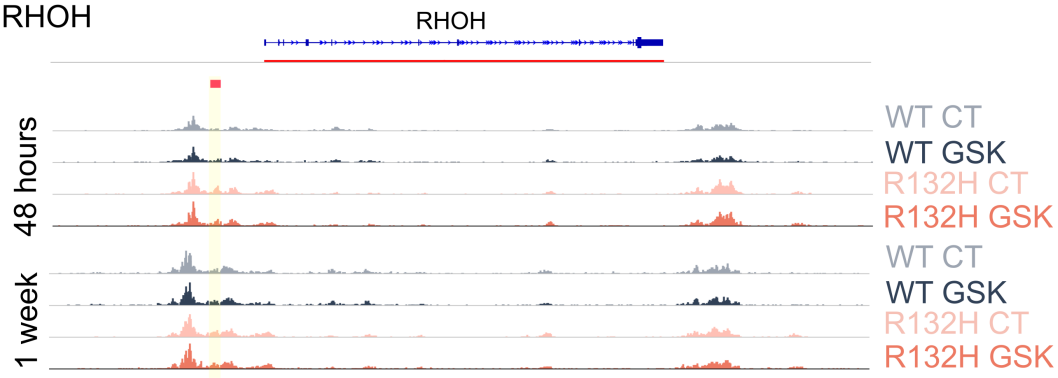


B Differentially Expressed Genes near K27ac Enhancer peaks (1 week)



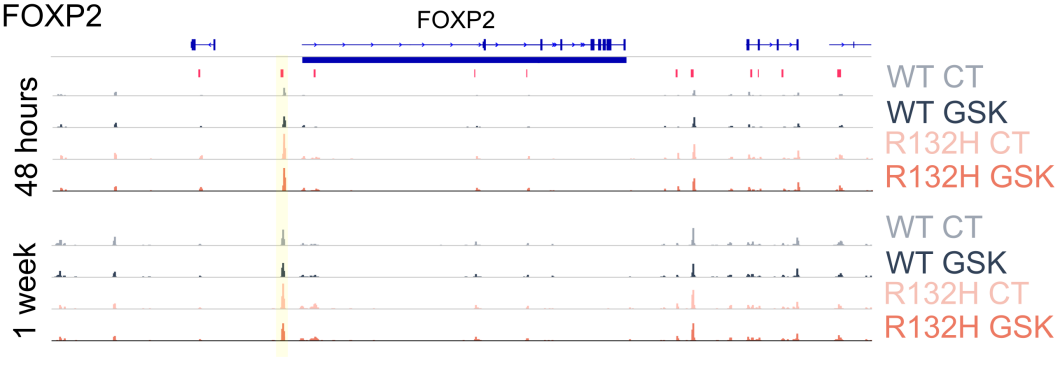
H3K27ac Enhancers:
Decreased in R132H
Increased in R132H

C RHOH



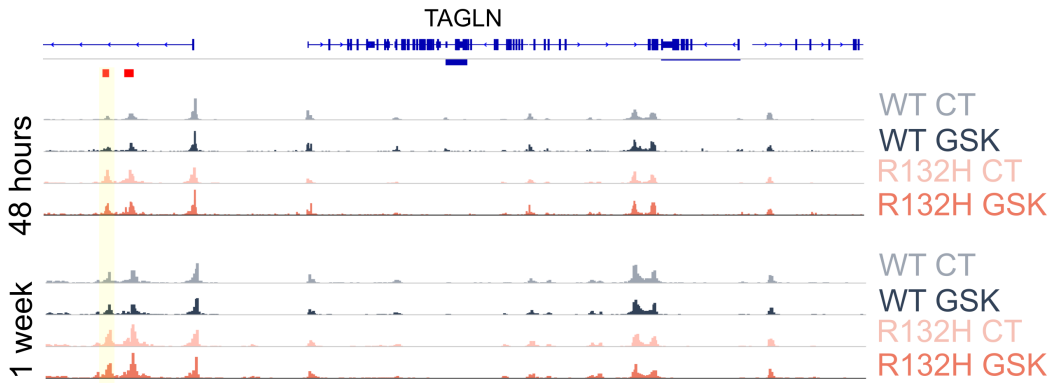
Tracks height: 0-30 (48h); 0-23 (1 week)

D FOXP2



Tracks height: 0-37 (48h); 0-29 (1 week)

E TAGLN



Tracks height: 0-12 (48h); 0-6.03 (1 week)

Figure 7.8 Annotation with EnhancerAtlas confirms uncoupling between enhancer acetylation and target gene expression*. A) Volcano plot showing the Log2FC (x axis) and -Log10 of the adjusted P.value of significantly expressed genes targeted by enhancers with increased K27ac in the 48 hours dataset. Dots color represent the acetylation change on the enhancer: red = H3K27ac is increased, blue is decreased. All enhancers were found to show increased H3K27ac. B) Volcano plot describing the levels and significance of differentially expressed genes in the one week dataset. Each gene has been predicted to be targeted by an enhancer showing increased H3K27ac. C,D and E) IGV tracks showing the genomic locus of three representative genes: RHOH (C), FOXP2 (D) and TAGLN (E). H3K27ac peaks are represented by red boxes, and the Enhancer in all three panels is highlighted in yellow. Tracks for each dataset were grouped and scaled to the same height, indicated below. Expression levels for all three genes are indicated with a red (RHOH, overexpression) or blue box (FOXP2, TAGLN, decrease in expression) beneath their RefSeq diagram. Bigwigs were obtained by merging bam files from two biological replicates, then normalized to 1x coverage.*

Thus, while EnhancerAtlas analysis gave overall a more confident pairing of enhancer sequences to their gene targets, we still detect a discrepancy between acetylation and gene expression changes in R132H cells.

7.5 Discussion

Assessment of histone H3K27ac in R132H IDH1 cells revealed a genome-wide rearrangement of this histone modification, with many loci showing an increase. In line with findings from our metabolomics experiment, treatment with GSK864 did not rescue any of these changes at the timepoints tested. Furthermore, focused analysis on enhancers suggested a potential uncoupling between increased enhancer acetylation and target gene expression.

Findings in our metabolomics analysis have informed our H3K27ac ChIP-Seq, as R132H cells show a two-fold increase of AcetylCoA compared to the WT counterpart. While most differential peaks are indeed showing H3K27ac enrichment, there is a consistent portion of loci that shows loss of this histone modification in IDH1 mutated cells. This might still be indicative of a dynamic rearrangement of H3K27ac. Further experiments would need to address whether these acetylation changes are indeed due to an Acetyl-CoA surplus. A way to test this could be to pharmacologically inhibit Acetyl-CoA or its production and perform another ChIP-Seq to assess any changes on its distribution. Another important experiment would be to use metabolomics to measure Acetyl-CoA levels in different cellular compartments, to check that the excess of this metabolite is indeed spilling into the nucleus, thus explaining the change in

acetylation levels. Finally, histones are not the only proteins subjected to this post-translational modification (Choudhary et al., 2009). To fully appreciate how much this increase in AcetylCoA levels is impacting on regular cellular functions, a proteomics study assessing the acetylome of WT or R132H U87 cells should be considered.

Analysis of H3K27ac distribution around enhancers has confirmed acetylation changes at several loci, however, these were not always coupled with similar variations in the expression of nearby genes. This could be due to the fact that the closest genes to an enhancer are not necessarily its targets, and to prove an enhancer-target interaction would require a 3C (chromatin conformation capture) experiment.

Many of the perceived anomalies (based on current view of gene regulation) in this dataset are due to gene expression decreasing in spite of its promoter or (predicted) enhancer showing increased acetylation levels. Assuming that the bioinformatics predictions are correct, another factor to be considered is also the extent of the expression change: if a gene in R132H cells shows decreased levels compared to the WT, it does not mean it is not being transcribed altogether, but rather that its transcript levels are *lower* when compared to the WT, perhaps due to a change in RNA turnover. Active expression, even if at lower levels, would explain the presence of enhancer acetylation, but perhaps not the increment of this modification compared to the WT. Thus, a useful distinction could be to further subdivide these differentially expressed genes into two categories: those showing a "binary" change, e.g. an expression level of 5 in the WT to 0 in the IDH1 mutant, and those showing a "relative" change, e.g. an expression of 5 in the WT to 4 or 3 in the IDH1 mutant. This could be easily achieved by filtering genes based on their log2 Fold-Change and absolute counts into the two groups, then by comparing acetylation levels of the respective enhancers to their gene expression to further confirm or disprove the uncoupling between enhancer acetylation and gene expression.

While more experiments would be needed to confirm the presence of a clear uncoupling between enhancer acetylation and target gene expression, we do observe that changes in acetylation levels around enhancers and promoters are not necessarily in line with gene expression. This could suggest that gene expression changes are concomitant to, but not immediately instructed by, epigenetic changes.

8. Deconstructing the models proposing loss of insulation of chromatin domains in R132H cells[1]

8.1 Adding a new piece to the puzzle: Does the IDH1 mutation influence the formation of chromatin domains?

As some gain of function IDH1 mutations lead to increased production of 2HG, which in turn is proposed to inhibit the activity of both protein and DNA demethylases, the net effect is likely to be increased levels of methylation on proteins and DNA. DNA methylation can inhibit the binding of transcription factors such as the insulator binding CCCTC-binding factor (CTCF) (Bell & Felsenfeld, 2000; Hark et al., 2000). As functional insulators are proposed to influence the formation of long-range chromatin interactions, the question is whether cells expressing an IDH1 gain of function mutation have altered long-range chromatin interactions. Flavahan et al. (2016) proposed that by increasing DNA methylation levels, IDH1 mutations would alter CTCF binding leading to loss of insulator function, changes in long-range chromatin interactions and changes in patterns in gene expression. However, work in *Drosophila melanogaster*, has uncovered no link between higher order structure in chromatin and gene expression, although this needs to be confirmed in mammalian systems (Ghavi-Helm et al., 2019). In addition to this, there are other limitations to this study (Flavahan et al., 2016) and others (Flavahan et al., 2019, Modrek et al., 2017) that link changes in DNA methylation levels to altered chromatin conformation and changes in gene expression in disease. These are explored and tested experimentally in this chapter.

The genome is organised into self-interacting genomic regions, called topologically associated domains (TADs) (Bickmore & Van Steensel, 2013). Proteins like CTCF are believed to act as insulators (Hanssen et al., 2017; Nakahashi et al., 2013), separating TADs from one another by binding to sequence-specific sites on the DNA (Dixon et al., 2012). This is proposed to

effectively insulate a gene on one domain from activation by an exogenous enhancer on a neighbouring domain (Flavahan et al., 2016, Modrek et al., 2017). TADs can often contain sub-TADs: nested, smaller domains, characterized by a weaker insulating properties and potentially more likely to be cell type-specific (Beagan & Phillips-Cremins, 2020) (Figure 8.1).

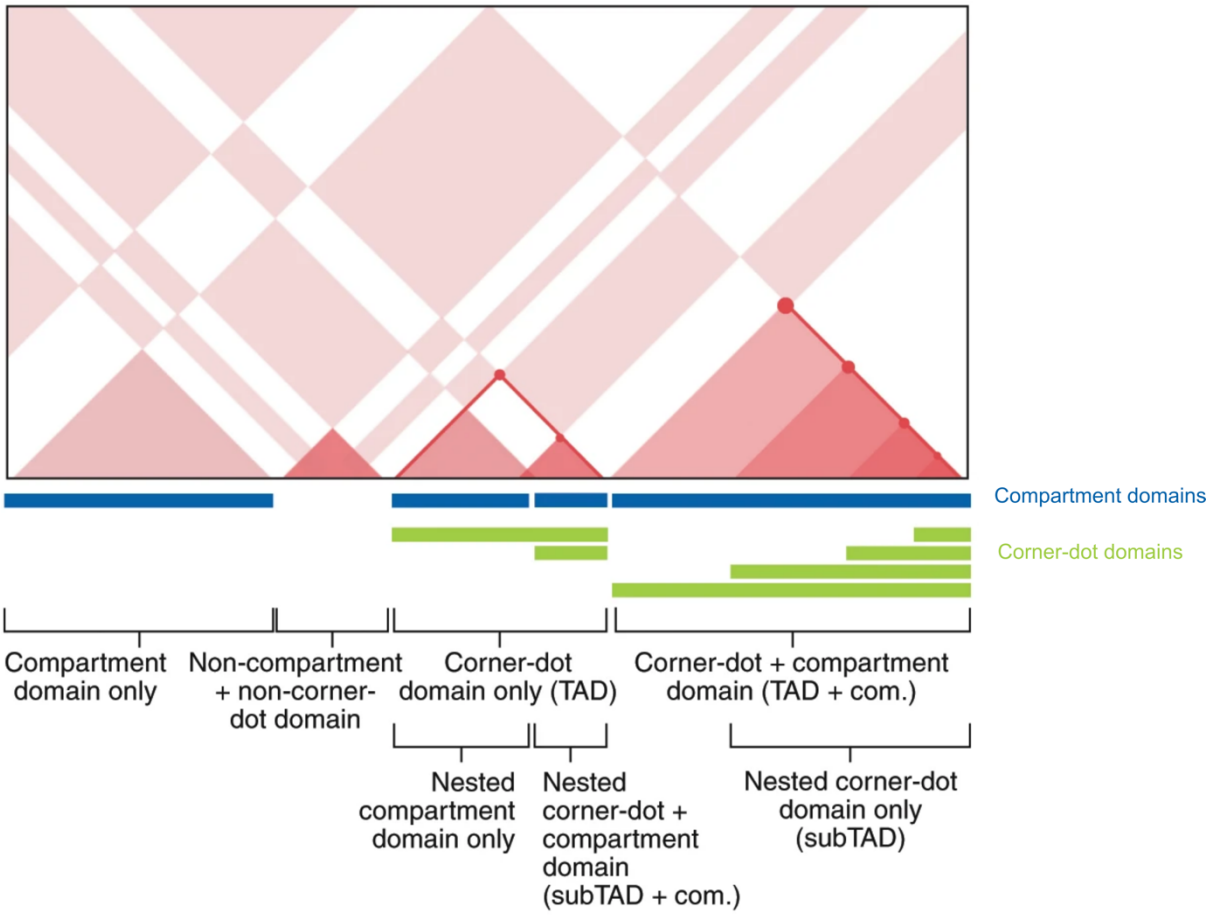


Figure 8.1. Classification of higher order chromatin structures. *Higher order chromatin structures are classified according to two main characteristics: their compartimentalization and the presence of a corner-dot. The corner-dot represents a physical contact between the bases of the equilateral triangle of which the corner-dot is the apex. These can represent TADs or sub-TADs, if nested within a larger TAD (green bars). Compartments domains are chromatin domains whose borders align with A/B compartments and are demarcated by stretches of similar chromatin marks or bound proteins, and while being distinct from TADs, the two can often overlap (blue)* (Beagan & Phillips-Cremins, 2020). *The presence of TADs,sub-TADs and compartment domains can give rise to different combinations of higher order structures, adding consistent complexity to genome topology. Image adapted from Beagan and Phillips-Cremins, 2020.*

By studying the 3D DNA structure in cells or patient-derived glioma samples bearing the most common *IDH1* mutation, R132H, Flavahan et al. (2016) find support for this model. They propose that high DNA methylation levels prevent binding of CTCF to the DNA, thus

destroying existing chromatin domains and promoting the formation of new TADs within the chromatin, contributing to the dysregulation of an already compromised gene expression. Indeed, they propose that loss of a domain boundary between a constitutive enhancer upstream of the *FIP1L1* gene and the gene encoding PDGF Receptor Alpha (PDGFRA) induces its aberrant expression (W A Flavahan et al., 2016) (Figure 8.1).

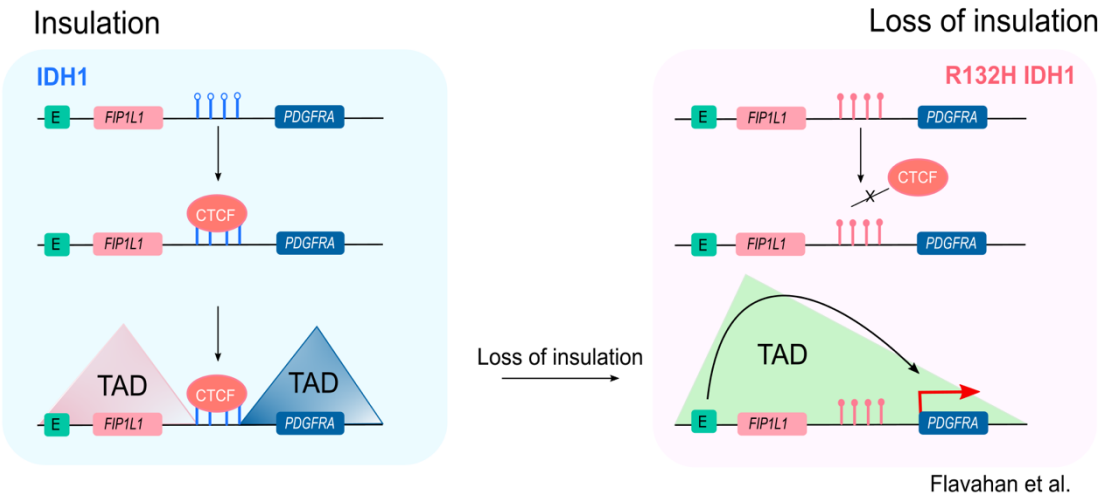


Figure 8.2. Changes in insulation of TADs as a consequence of IDH1 mutation. *According to the model suggested by Flavahan et al., (2016) in a WT IDH1 setting, CTCF is able to bind to particular target sites along the genome, in certain contexts acting as a functional insulator, creating DNA domains that separate neighbouring regions. At CTCF binding sites containing the CG dinucleotide, changes in levels of DNA methylation influence its DNA binding. Specifically, the production of 2HG inhibits TETs, leading to hypermethylation (red, filled lollipops) of CTCF-binding sites, reducing binding (right panel). In the context of PDGFRA, the formation of domains sequesters an enhancer upstream the FIP1L1 gene into a separate domain (left panel). Upon IDH1 mutation, however, the overall increase in methylation levels due production of 2HG induces loss of CTCF binding to its target sites, leading to loss of insulation between TADs. In this example, destruction of a boundary induces a rearrangement that brings PDGFRA in proximity of an enhancer found upstream the FIP1L1 gene, thus inducing its deregulation (right panel).* (Raineri & Mellor, 2018)

Although in the 2016 paper, Flavahan et al focus mainly on the *PDGFRA l*ocus, a total of 203 putative boundaries were annotated, following an extensive bioinformatic analysis. To do this, the authors first map CTCF binding by ChIP-Seq in 11 primary glioma tumours, 6 WT and 5 mutated in IDH1, as well as four cell lines, two wild-type (GSC4 and GSC6, primary glioblastomas) and two mutated in IDH1 (BT142, grade III oligoastrocytoma and GSC119, secondary glioblastoma). They find a substantial fraction of CTCF binding sites are lost in cells expressing the IDH1 mutation, although there is no mention of whether the level of CTCF

protein or transcript is changed in this context. They then use publicly available whole genome bisulfite sequencing data from the cancer genome atlas (TCGA) to assess the DNA methylation level on selected loci that showed loss of CTCF binding, and find that these showed increased methylation upon IDH1 mutation (mean % DNA methylation: approximately 60% in IDH1 mutant, around 10% in WT). Next, they collate a set of publicly available HiC maps into constitutive domain boundaries, using datasets derived from several different cell types (IMR90 human lung fibroblasts, GM12878 lymphoblastoid cells, K562 bone marrow chronic myelogenous leukaemia cells and NHEK normal epidermal keratinocytes), based on the assumption that boundaries remain stable across cell types. They assess gene expression levels in 357 normal brain tissue samples, using genotype-tissue expression consortium (GTEx) datasets, and observe that gene expression normally correlates well between genes positioned in the same domain, but that correlation falls when considering genes found in separate, neighbouring domains (Ardlie et al., 2015). Then they ask whether this relationship between gene expression and domain boundaries is maintained in tumours bearing the IDH1 mutation. They use RNASeq datasets from 230 IDH1 mutant and 56 WT lower-grade glioma samples (TCGA), and scan the genome for pairs of genes whose expression correlates more strongly in IDH1 mutated samples. In contrast with results from normal brain tissue samples, they observe that in IDH1 mutated samples, gene expression between genes residing in two different TADs correlates better compared to genes residing within the same domain. Based on this result, the authors hypothesize that IDH1 mutation might induce a disruption in TAD boundaries. Finally, they annotate 203 lost boundaries based on the presence of gene pairs that satisfied three criteria: crossing a CTCF binding site, *regardless of whether its binding is decreased or not upon IDH1 mutation*; crossing a HiC-defined boundary; correlating more strongly with cross-boundary genes compared to genes within the same boundary. According to the authors, these 203 loci also showed higher DNA methylation and lower CTCF binding, although it is not stated whether this data was taken into consideration when looking for the gene pairs of interest,

or whether DNA methylation and the extent of CTCF binding at the resulting 203 loci was assessed only afterwards.

A more recent study from the same group has shown other examples of such a proposed mechanism in a different cellular model. Using a similar bioinformatic analysis, but focusing on gastrointestinal stromal tumours (GIST) with a deficiency in succinate dehydrogenase activity, a metabolic defect inducing DNA hypermethylation, the group finds that loss of a TAD boundary between a super-enhancer close to the *ANO1* gene and an upstream locus coding for the FGF3 and FGF4 tyrosine kinase receptors induces the aberrant interaction of their promoters with the super-enhancer. A similar phenomenon is described around a locus coding for the well-known GIST oncogene KIT. Here, a super-enhancer is generally enclosed in a 100 kb domain marked by CTCF binding at its boundaries. In SDH-deficient GISTs, loss of CTCF binding due to DNA hypermethylation induces an increase in interactions, not only between the enhancer and sequences beyond the insulator, but also with KIT gene itself. While these loci of interest were found with a similar bioinformatics analysis, compared to their 2016 paper the results are more convincing, as the authors have generated all the sequencing data, ChIP and chromosome conformation experiments in their own cellular model. However, their cellular model does not contain the IDH1 gain of function mutation, but another metabolic defect with similar consequences, and some inconsistencies still remain. For example, while RNA sequencing shows *FGF3, FGF4* and *KIT* are overexpressed in SDH-deficient GISTs compared to their SDH-intact counterpart, the exact impact of these new interactions on *FGF3, FGF4* and *KIT* expression is unclear. This is because the authors perform quantitative Real-Time PCRs (qPCR) on their levels only following CRISPR-Cas9 deletion of the annotated boundary, but not in regular SDH-deficient GIST cells. However, upon insulator deletion *KIT* expression increases by 50%, a relatively modest change if compared to the 6-fold or 35-fold induction of *FGF3* and *FGF4*, respectively (William A Flavahan et al., 2019b).

The establishment of a system to model lower-grade glioma genetic lesions in human neural stem cells (HNSC), by combining the R132H *IDH1* mutation together with silencing of both *p53* and *ATRX*, has also linked higher order structures and gene expression at the *SOX2* locus, using a combination of RNA-Seq, DNA methylation arrays, as well as localised characterization of the locus of interest by targeted bisulfite sequencing, CTCF ChIP-qPCR and 4C sequencing using *SOX2* as bait (Modrek, Golub, Khan, Bready, Prado, Bowman, Deng, Zhang, Rocha, Raviram, Lazaris, Stafford, LeRoy, et al., 2017). They propose that reduced CTCF binding around the *SOX2* gene is associated with loss of a DNA domain or loop which normally positions the *SOX2* promoter in close proximity to an enhancer, approx. 0.5 to 1 Mb downstream from *SOX2*. Thus, altered CTCF binding is associated with downregulation of *SOX2* expression, blocking differentiation, in contrast to the increase expression of *PDGFRA* in the Flavahan study.

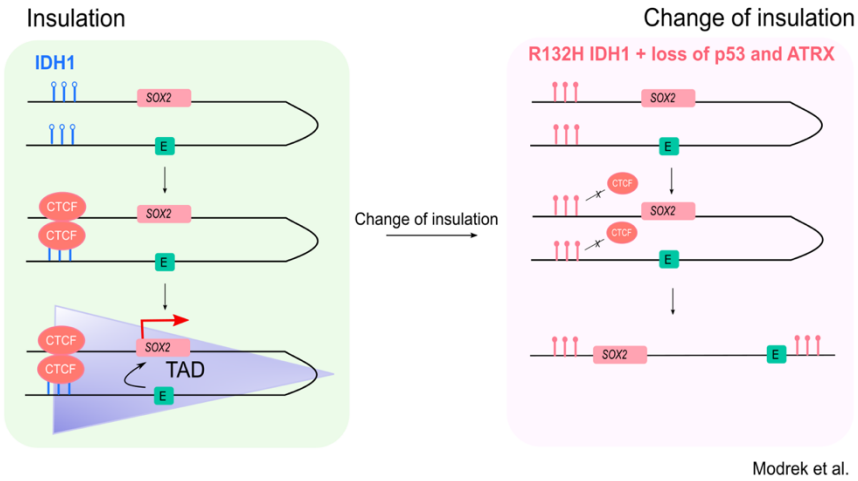


Figure 8.3. SOX2 loses contact with its enhancer following introduction of three genetic lesions (three-hit model): a gain of function IDH1 mutation combined with loss of ATRX and p53. *The example described in Modrek et al. focuses on the SOX2 locus. Here, in the WT IDH1 context, CTCF binding induces the formation of a loop that brings the SOX2 gene and its promoter in close proximity to a downstream enhancer that would be otherwise out of reach (left panel). This interaction favors SOX2 expression. Upon a three-hit mechanism that includes mutation of IDH1 and loss of both p53 and ATRX, reduced binding of CTCF at the SOX2 locus impairs the formation of the loop, thus causing a change, rather than a loss, of insulation. In this new setting, the downstream enhancer is too far away to interact with the gene, whose expression is now downregulated (right panel).* (Raineri & Mellor, 2018)

Despite findings with similarities to the model proposed by Flavahan et al., the Modrek et al. paper offers some interesting points for discussion. When looking at the methylation levels around the *SOX2* promoter, there were no striking differences between their three-hit cells and the empty-vector controls. Only when taking into account a much larger region around *SOX2* (1.2 Mb) they were able to identify specific areas up- and downstream of the gene that indeed showed increased methylation levels. When comparing these areas to the CTCF ChIP-seq data from Flavahan et al., they identified five potential CTCF binding sites that could be influenced by increased DNA methylation. The authors face what will be the ultimate challenge for future research in the field: how to correctly map the domain boundaries by merging the CTCF binding data with the DNA methylation and the chromatin conformation information and to show whether one or all five of these CTCF binding sites contribute to a chromatin conformation that facilitates promoter:enhancer interactions. Indeed, this will require the combination of a solid

mapping of CTCF binding sites across the genome, a reliable description of the chromatin domains in both *IDH1* wild-type and mutated cells, a thorough annotation of the H3K27ac mark to define enhancer sequences, an accurate portrayal of DNA methylation landscape and finally a method to validate the findings.

8.2 From local to a genome-wide mechanism: challenges ahead

The Modrek and Flavahan papers reflect the challenges in these sort of studies. For example, matching different -omics into a single picture of the epigenetic state of the cell will prove to be difficult. When starting from publicly available data, the main issue will be choosing the appropriate datasets. Data from the different –omics might not be available in the same cell line, or, at times, even the same cell type. Furthermore, sequences might have been analysed or normalised according to different methods, and raw data is not always disclosed. All these might seem small details, but they add up introducing biases in the analysis, making it difficult for the scientist to draw clear conclusions.

A second option is carrying out the all the -omics experiments in one system in a more controlled environment. This would ensure consistency of the cellular model, and a better control over the technical biases that might be introduced in the experimental procedure. However, this would require designing a proper cellular model. Many papers have generated their own system by stably transfecting an empty vector, wild-type *IDH1* or R132H *IDH1* into glioma cell lines. While this might have been a good solution to initially study the alterations induced by the *IDH1* R132H mutation, if the focus has now shifted to higher order chromatin structure then perhaps introducing a gene via transfection might cause some alterations to the DNA loops by itself. A possibility could be to selectively introduce the *IDH1* R132H mutation in a wild-type *IDH1* cell line using the CRISPR-Cas9 system.

Obtaining the data will only solve part of the problem, as potential difficulties will lie in correctly mapping the TADs and understanding which CTCF binding sites are responsible for

the disruption or the formation of new contacts between gene promoters and enhancers. Multiomics is an approach to data analysis that aims at integrating, rather than comparing, results from different –omics experiments, in an effort to model complex phenotypes. Despite this being a task that presents its own challenges, it could be the most appropriate way to move forward. The final challenge will be to define whether *IDH1* mutations affect particularly sensitive loci containing potential oncogenes such as *PDGFRA* and *SOX2*, or whether this is a genome-wide mechanism.

Taken together, the work from Flavahan et al. and Modrek et al. have re-inforced the link between changes to higher-order structures in chromatin and changes in gene expression in disease. However, they have some limitations, as discussed. Here, using the U87 glioblastoma derived cell line, with or without the gain of function IDH1 R132H mutation, and the drug GSK864 which inhibits the production of 2HG, key aspects of the link between metabolic changes, the chromatin and gene expression are explored with the aim of testing the overarching hypothesis in the Flavahan and Modrek papers.

8.3 Deconstructing the model based on Flavahan (2016)

The current model is defined by three main steps. First, mutant *IDH1* produces 2HG from α-Ketoglutarate. Second, 2HG inhibits histone and DNA demethylases, causing an increase in methylation levels. Finally, methylation of CTCF binding sites in the DNA inhibits CTCF binding inducing rearrangement of TADs and this leads to altered gene expression.

The second step in this model presents complications when considering its impact on gene expression, as the specific effects of increased global hypermethylation are context-dependent. For example, increased promoter DNA methylation (potentially repressive) may have different consequences from increased insulator methylation. Flavahan et al. report a five-fold increase of *PDGFRA* expression in *R132H IDH1* glioma cells, which they claim is due to the loss of insulation leading to the new interaction between a strong enhancer upstream of the *FIP1L1*

gene and the *PDGFRA* promoter. However, given that mutation of *IDH1* is sufficient to induce a CpG island methylator phenotype (G-CIMP) (Turcan et al., 2012) and that the UCSC Genome Browser describes the presence of a CpG island within the *PDGFRA* promoter, an expectation is that in the *R132H IDH1* context, this promoter becomes hypermethylated, which is generally linked to transcriptional inactivation (De Smet, Lurquin, Lethé, Martelange, & Boon, 1999). There are two possible explanations for this apparent contradiction. First, the proximity of a stronger enhancer is able to overcome what would be an otherwise silencing effect on gene expression. Second, the promoter escapes DNA hypermethylation due to H3K4me3, a histone modification associated with active genes (Santos-Rosa et al., 2002). Given its mutual exclusivity with DNA methylation (Weber et al., 2007), the presence of this mark could protect DNA sequences from being methylated by hindering binding of Dnmt3L, a protein thought to help the de novo methyltransferase Dnmt3A/B (Hata, Okano, Lei, & Li, 2002; Ooi et al., 2007). Interestingly, H3K4me3 is one of the few methylation marks that does not seem to increase upon *IDH1* mutation (Lu et al., 2012). Potentially, it could mean that those DNA sequences associated with H3K4me3 remain unmethylated, and transcriptionally active, in spite of the global increase in methylation.

The third step in this model, that methylation on CTCF binding sites inhibits CTCF binding to induce domain rearrangements, is the most challenging to validate. The first indications that 2HG-driven hypermethylation induces a rearrangement of chromatin domains was reported in Flavahan et al. (2016). The authors exploit a combination of computational methods, chromatin immunoprecipitation followed by sequencing (ChIP-Seq) to assess where on the DNA CTCF interacts, and publicly available data reporting changes to degrees of DNA methylation in glioma cells with or without the *IDH1* mutation (Noushmehr et al., 2010). Results suggest that tumours bearing the mutated *IDH1* lose CTCF binding and show increased DNA methylation. Then, the authors proceed to analyze higher order chromatin structures, using chromatin capture sequencing techniques which are based on the crosslinking of cells with formaldehyde to link

together chromatin segments in close spatial proximity. Chromatin is then digested with restriction enzymes and ligated in DNA hybrid molecules containing parts of the two sequences of DNA that were interacting with each other. In the classical 3C protocol, a real time PCR is performed using specific primers to amplify a locus of interest. In more advanced methods, such as Hi-C, gaps in the DNA sequences are filled in with biotinylated nucleotides after restriction enzyme digestion. Hybrid sequences are then pulled down and used to prepare libraries, enabling whole-genome analysis of the interactions, as opposed to one locus at a time (Dekker, Marti-Renom, & Mirny, 2013).

Flavahan et al. couple publicly available data from Hi-C, used to assess chromatin domains genome-wide in different cell types, such as human lung fibroblasts, lymphoblastoid cells or keratinocytes (Rao et al., 2014), and RNA sequencing experiments from patient-derived low grade glioma samples (Brennan et al., 2013; R. G. Verhaak et al., 2010) to compute the correlation of gene expression between genes contained in the same chromatin domain and those belonging to neighbouring domains. When comparing gene expression from gliomas with or without the *IDH* mutation, they find that in mutated gliomas, genes tend to correlate better with genes from a close, but separate, domain than with genes within their same domain. Finally, they scan these loci of interest for genes whose expression is higher in *IDH1* mutant gliomas, obtaining a final list of genes. As a result of this key computational experiment, the authors select the locus on chromosome 4 between *FIP1L1* and *PDGFRA* as an example of disrupted insulation. To validate their hypothesis, the authors use a series of capturing chromosome conformation (3C) experiments around the *FIP1L1-PDGFRA* locus to show the altered domain structure.

The way they formulated their hypothesis using a largely computational method raises an interesting problem. While the publicly available RNA sequencing datasets used derive from glioma samples, the Hi-C experiments had been carried out in a series of very different cell

lines: IMR90 human lung fibroblasts, GM12878 lymphoblastoid cells, K562 bone marrow chronic myelogenous leukaemia cells and NHEK normal epidermal keratinocytes. Publicly available datasets are a valuable resource to test out a hypothesis, especially given the complexity of carrying out a Hi-C experiment, and it could be argued that if the same conformations exist in different cell types they are likely to be widely conserved. On the other hand, is it valid to use 3C data from different cell types to make specific assumptions about glioma cells, when they might have a completely different arrangement of TADs? To address this, the authors follow their computational analysis with local 3C experiments to assess the interactions at the *FIP1L1-PDGFRA* locus in a series of primary glioma cells and a panel of glioblastoma cell lines. To confirm that loss of insulation, and the subsequent *PDGFRA* overexpression, is indeed due to loss of CTCF binding, Flavahan et al. genetically edited out the putative CTCF binding site between *FIP1L1* and *PDGFRA* using the CRISPR-Cas9 system. However, instead of measuring the actual formation of a novel interaction through a 3C experiment on such CRISPR-modified cells, their readout is based on *PDGFRA* expression levels obtained by qPCR, and the presence of PDGFRα on the cell membrane, measured by FACS. Thus, they do not actually show that the CTCF and methylation-dependent function of the insulator alters higher order chromatin loops.

This section aims at understanding whether loss of CTCF binding could indeed affect TADs insulation; whether this effect is genome-wide or happening only at specific loci, and whether it can be rescued through treatment with an IDH1 inhibitor. To investigate this, we set out to test each step of the model with a combination of sequencing techniques, all performed in the same cellular model: the human glioblastoma cell line U87, WT or bearing the R132H IDH1 mutation. We assessed CTCF binding through chromatin immunoprecipitation followed by sequencing (ChIPSeq) of CTCF, alterations of gene expression with 3'end mRNA Sequencing (RNASeq) and will assess DNA methylation levels through TET-assisted pyridine borane sequencing (TAPS, in collaboration with Prof. S.Kriaucionis group, see note 6.B).

8.4 Assessing the overlap of CTCF binding sites with lost boundaries from Flavahan et al. 2016

According to the model proposed by Flavahan et al., methylation of CTCF target DNA sequences inhibits its binding in the IDH1 mutant. Results from CTCF ChIP showed a general decrease in CTCF binding in R132H control cells: 2692 and 1630 peaks were lost in the 48 hours and one week dataset, respectively (Chapter 6). However, this decrease in binding did not seem to be due to a decrease in CTCF itself, as its transcript level remain unchanged between WT and R132H cells (48 hours: Log2FC = -0.016, P.adj = 0.97; one week: Log2FC = 0.2, P.adj = 0.86). We next asked whether these CTCF binding sites lost are overlapping with lost boundaries annotated in Flavahan et al. 2016.

To assess CTCF binding at these sites, levels of CTCF IP over Input were calculated, then plotted on a heatmap (Figure 8.3), where green corresponds to sequences where CTCF is enriched in the IP, thus bound, while purple indicates DNA sequences where the IP signal is lower or equal to the background, meaning that CTCF binding is absent. Results show that at most Flavahan-annotated boundaries there is no CTCF binding detected in any of the samples, including WT (Figure 8.3A). Furthermore, the few sequences that show CTCF binding (green) do not seem to change between R132H and WT control cells, potentially suggesting that this subset of boundaries would not be lost upon IDH1 mutation.

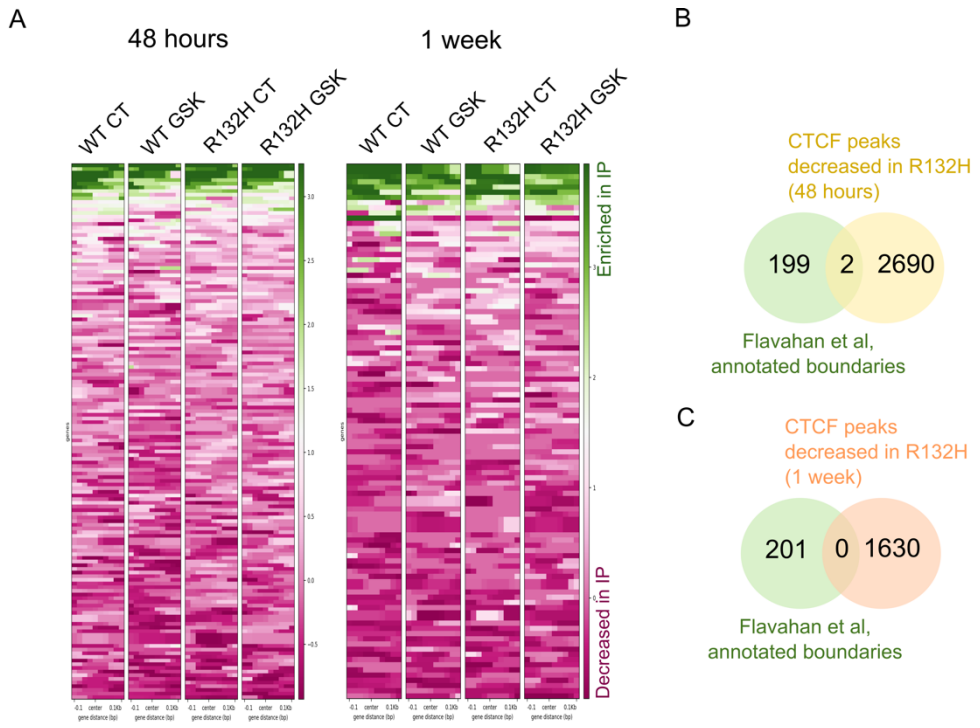


Figure 8.4. CTCF binding around Flavahan-annotated boundaries. *A) Heatmap representing the levels of CTCF IP enrichment along 201 Flavahan-annotated boundaries. Biological replicates reads (N=3 for the 48 hours dataset, N=2 for one week) were merged, then normalized to 1x coverage using bamCoverage (Deeptools). Log2 of the fold change between IP and Input was calculated by running bigwigCompare on the normalized files, then results were visualized in a heatmap using computeMatrix followed by plotHeatmap, all part of Deeptools* (Ramírez, Ryan, Grüning, et al., 2016). *In the heatmap, purple-colored cells indicate loci decreased in the IP sample compared to the Input, whereas green cells show enrichment in the IP. B) Venn diagram showing the intersection between Flavahan-annotated boundaries and differentially decreased CTCF Peaks detected in ChIPSeq at 48 hours. C) Venn diagram showing the intersection between Flavahan-annotated boundaries and differentially decreased CTCF Peaks detected in ChIPSeq 1 week. Differential peaks were calculated with HOMER getDifferentialPeaks using default options.*

To evaluate whether any of these loci were differentially bound by CTCF, genomic coordinates of Flavahan-annotated loci were intersected with the list of differential peaks from both CTCF datasets. Out of 203 annotated boundaries, only two CTCF sites were differentially decreased in the 48 hours dataset, and none at one week (Figure 8.3B and C). The two common CTCF binding sites were selected for further inspection in the IGV genome browser (Figure 8.4 and 8.5). Notably, in both cases the overall ChIP signal, upon 1x coverage normalization, is much higher in the WT samples, resulting in a string of significantly decreased CTCF peaks. This, combined with the fact that most of the gene expression detected is decreased could suggest alterations in the entire loci.

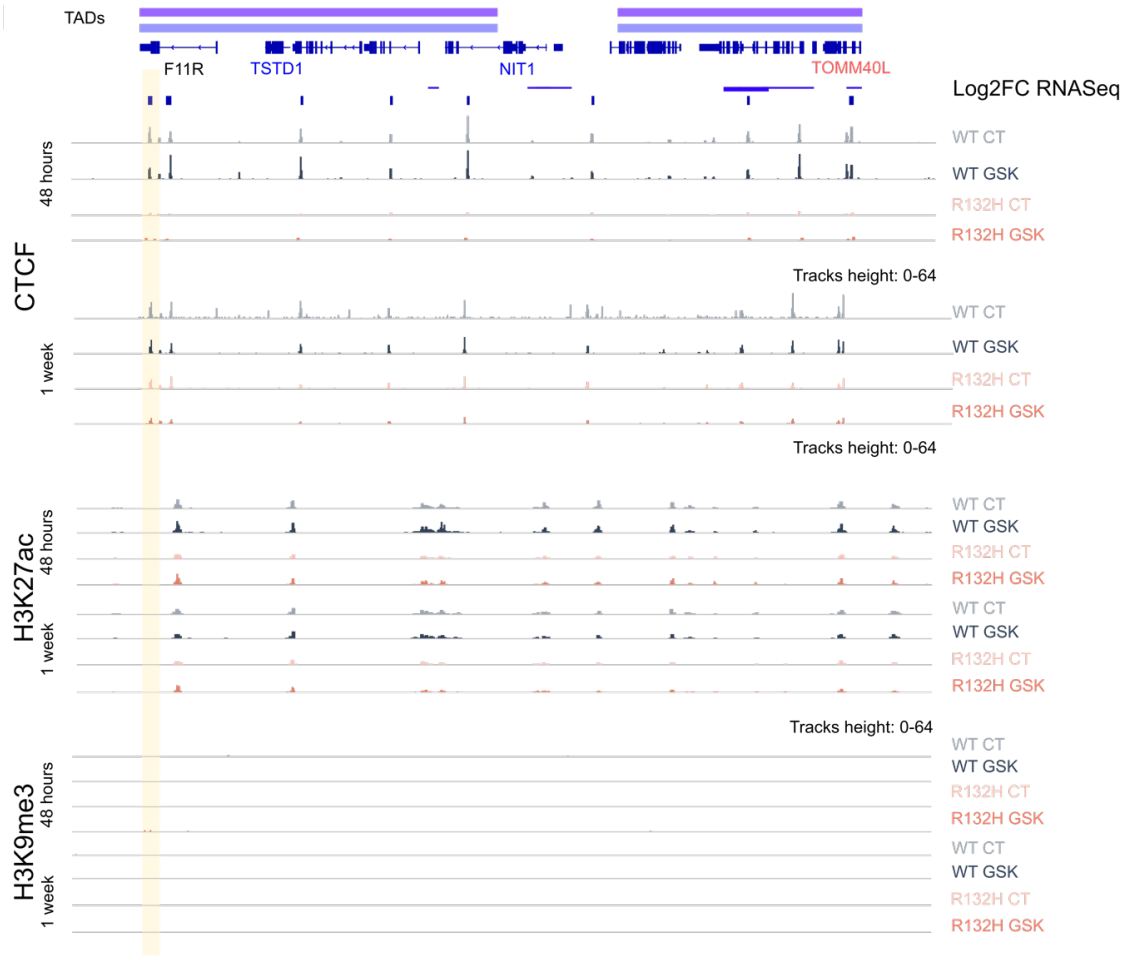


Figure 8.5. Distribution of CTCF around the F11R locus*. Genome browser tracks representing the F11R locus, shown in the following order: two astrocytes TADs annotation, RefSeq genes, Log2FC histograms of differentially expressed genes from RNASeq (48 hours), blue boxes indicating differentially decreased CTCF peaks, normalized CTCF ChIP at 48 hours and at one week, H3K27ac and H3K9me3 distribution. Names of genes losing correlation with F11R are shown in blue, while those gaining correlation are red. A yellow box indicates the CTCF differentially decreased peak overlapping with a boundary from the Flavahan annotation. Tracks have been normalized to 1x coverage, grouped and scaled in IGV to the maximum height indicated below.*

The CTCF site in Figure 8.4 is found near the *F11R* gene (Log2 Fold-change = -1.43, P.value = 1.16e-113). Upon IDH1 mutation, gene expression of *F11R* is thought to lose correlation with the neighbouring genes *NIT1* and *TSTD1*, while the latter gains correlation with *TOMM40L*. According to the Flavahan hypothesis, these genes in a WT setting would be found in two neighbouring TADs, whose boundaries would be marked by CTCF binding. While no data is available regarding TAD boundaries in U87 cells, two publicly available datasets from astrocytes of the cerebellum and astrocytes of the spinal cord were loaded on IGV (purple tracks, Figure 8.3A and B). According to both TADs annotations, the boundary would be downstream

of the site annotated in the Flavahan study (highlighted in yellow). There appears to be no correlation with TAD boundaries and CTCF peaks, as several CTCF peaks significantly decreased in R132H are scattered throughout both TADs. Results from RNASeq are not providing much insight, as we do not detect *F11R* or *TSTD1*, although genes such as *NIT1*, *ADAMTS4*, *NDUFS2* and *TOMM40L* are all downregulated in the IDH1 mutant (Table 17). In line with the downregulation of such genes, H3K27ac around their promoters also seems to be generally decreased in R132H cells.

Table 17. Gene expression and significance levels of genes around the F11R locus. *Log2FC values and respective P.values, as adjusted by Benjamini-Hochberg, of differentially expressed genes around the F11R locus in the 48 hours RNASeq dataset. Differential expression was calculated as explained in Chapter 3 and Methods.*

Gene	Log2FC	P.Adjusted
NIT1	-1.41	0.04
ADAMTS4	-2.79	0.0008
NDUFS2	-1.01	0.02
TOMM40L	-1.21	0.03

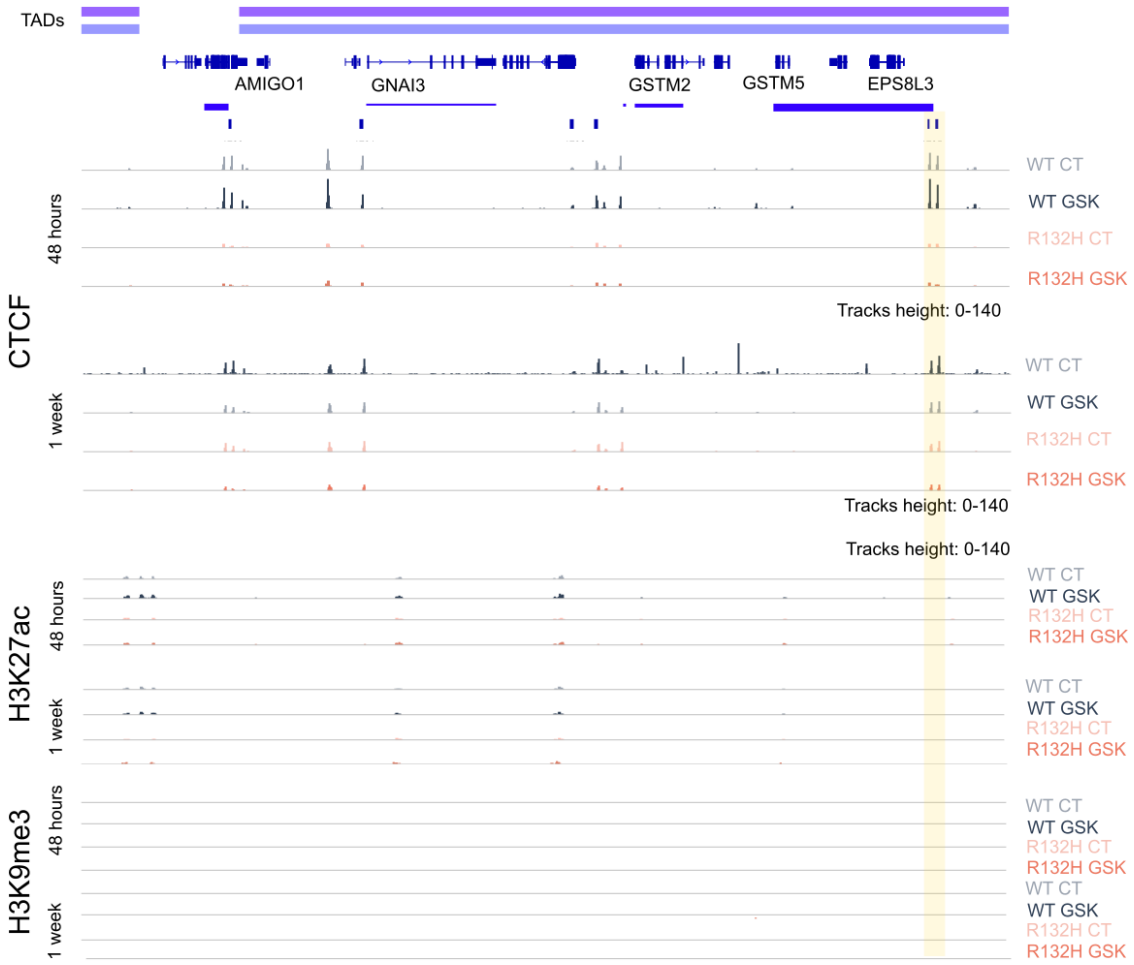


Figure 8.6 CTCF distribution around the GSTM locus. *IGV tracks for CTCF (top), H3K27ac (middle) or H3K9me3. Tracks are displayed as in Figure 6.4. Bigwig files were normalized to 1x coverage, then grouped and scaled to the same height (range: 0-140). A yellow box highlights the CTCF peak annotated in Flavahan et al., 2016.*

The second locus where we observe an overlap is around the *GSTM* gene locus where a CTCF binding peak is lost in the IDH1 mutant (Log2 Fold-change = -0.68, P.value = 3.66e-35, Figure 8.5 highlighted in yellow). Here, upon IDH1 mutation, *GSTM5* gene expression loses correlation with the neighbouring *GSTM2*, while gaining correlation with the upstream *AMIGO1*. As with the example with *F11R*, the Flavahan annotation of the boundary around the GSTM locus is at odds with the TADs defined in the astrocytes. Furthermore, according to the astrocyte annotation *GSTM5*, *GSTM2* and *AMIGO1* fall into the same TAD. Interestingly, gene expression, CTCF binding and H3K27ac distribution around this locus are downregulated.

Indeed, *ATXN7L2*, *GNAI3*, *GSTM4*, *GSTM5* and *EPS8L2* levels are all decreased in R132H cells (Table 18).

Table 18. Gene expression and significance levels of genes from the GSTM locus*. Log2FC and respective adjusted p-values of differentially expressed genes around the GSTM locus in the 48 hours RNASeq dataset. Differential expression was calculated as explained in Chapter 3 and Methods.*

Gene	Log2FC	P.Adjusted
ATXN7L2	-4.005	0.002
GNAI3	-1.03	0.03
GSTM4	-2.07	0.008
GSTM5	-4.3	4e-06
EPS8L2	-1.5	0.02

Analysis of the two loci of interest has revealed a discrepancy between Flavahan annotated boundaries and astrocytes TADs annotation. While ChIPSeq results have been mapped to the hg38 genome assembly, Flavahan boundaries and astrocytes TADs annotations are based on the hg19 human genome assembly. To make them comparable with ChIPSeq results, all have been converted into hg38 coordinates using liftOver (https://genome.ucsc.edu/cgi-bin/hgLiftOver), so a possibility could be that the conversion from hg19 to hg38 has caused a shift in the coordinates of either the TADs annotation or Flavahan boundaries, or both. However, given that both were initially built with the same genome assembly, the expectation would be that this discrepancy would be present even in their original annotation. Another detail to consider is that TADs annotations, Flavahan's boundaries and CTCF ChIPSeq have all been generated using different cell lines: astrocytes, patient-derived glioblastoma cells and U87 cell line. While TADs are thought to be stable across different cell types, it may be that there is a small shift depending on the cellular model under study (Barutcu, Maass, Lewandowski, Weiner, & Rinn, 2018; Dixon et al., 2012).

Results show there is poor overlap between Flavahan-annotated boundaries and CTCF binding sites lost in R132H cells. Furthermore, in the two loci that do overlap, the boundary annotated

by Flavahan and the one marked by two publicly available TADs annotations do not coincide, at least in our cellular model.

8.5 *PDGFRA expression is not a suitable readout for loss of insulation between TADs in R132H IDH1 U87 cells*

A key point in Flavahan's model is the increased overexpression of PDGFRA, resulting from the loss of insulation with an upstream enhancer. Indeed, PDGFRA levels in IDH1 mutated gliomas are shown to be up to five fold higher than their WT counterpart (W A Flavahan et al., 2016). To assess whether a similar phenomenon could be happening in our experimental set up, we evaluated expression levels of PDGFRA and other genes around the 201 lost boundaries annotated in the paper. Results from RNASeq show thirteen genes in common with their annotation, seven of which are differentially expressed in R132H cells. Amongst these, four out of seven show decreased levels in R132H cells (Figure 8.6A).

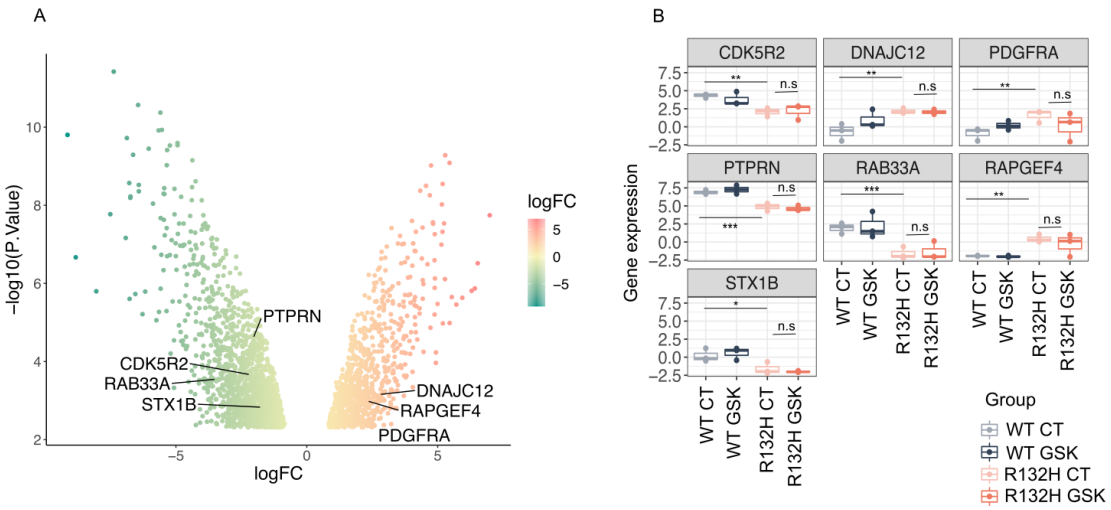


Figure 8.7. Volcano plot of differentially expressed genes shows commonalities with the Flavahan dataset. *Differentially expressed genes between WT and R132H cells were calculated through a linear model (~ Genotype + Genotype*Treatment) using EdgeR, after TMM normalization, batch effect removal and voom transformation of input gene counts. A volcano plot shows the log2 Fold Change of each gene (x axis) coupled with -log10 of its P.Value (y axis). Text labels indicate genes in common with the Flavahan dataset. B) Boxplots showing levels of gene expression for N=3 biological replicates, after TMM normalization and voom transformation. Significance levels for the comparisons WT CT vs R132H CT and R132H CT vs R132H GSK864 are indicated as following: Not significant = ns; P > 0.05; * P < 0.05; ** P < 0.01; *** P < 0.001. (N=3 per group).*

While we find an increase of PDGFRA expression in IDH1 mutant cells and its reduction upon GSK864 treatment, only the former is statistically significant (R132H control vs WT control:

Log2FC = 2.47, P.adj = 0.04; R132H GSK vs R132H control: Log2FC = -2.5, P.adj = 0.5). However, quantitative Real Time PCR (qPCR) validation of PDGFRA levels using the same primers from their 2016 paper did not confirm these results, as the gene shows the same relative expression level in R132H and WT control cells (Figure 8.7A, left panel).

Table 19. Raw Ct values of PDGFRA and DNAJC12 expression as detected in qPCR.

In a qPCR experiment, the Ct (cycle threshold) value corresponds to the number of PCR cycles required for the fluorescence in a sample to exceed the background threshold. Thus, the more highly expressed the gene is, the lower the Ct value will be, as fewer rounds of PCR amplification will be required to have enough molecules to produce the fluorescence needed to exceed the threshold. Conversely, the lower the starting amount of a certain sample, the more rounds of PCR will be needed to produce sufficient fluorescence, the higher the Ct value will be.

Biological Repeat	Sample	PDGFRA Ct	DNAJC12 Ct
N1	WT CT	25.73	36.49
N1	WT CT	25.68	37.07
N1	WT CT	25.99	39.95
N1	WT GSK864	25.51	38.51
N1	WT GSK864	25.42	37.12
N1	WT GSK864	25.66	38.44
N1	R132H CT	24.95	36.17
N1	R132H CT	24.34	39.63
N1	R132H CT	24.4	36.33
N1	R132H GSK864	24.9	41.41
N1	R132H GSK864	25.92	30.09
N1	R132H GSK864	26.02	35.64
N2	WT CT	24.69	36.47
N2	WT CT	24.82	34.16
N2	WT CT	24.95	30.81
N2	WT GSK864	24.59	31.33
N2	WT GSK864	25	30.29
N2	WT GSK864	24.35	24.73
N2	R132H CT	24.61	25.72
N2	R132H CT	24.98	25.51
N2	R132H CT	24.14	24.75
N2	R132H GSK864	25.01	24.65
N2	R132H GSK864	24.99	24.78
N2	R132H GSK864	24.67	23.85
N3	WT CT	29.32	31.34
N3	WT CT	28.79	37.79

N3	WT CT	--	29.68
N3	WT GSK864	30.57	31.3
N3	WT GSK864	30.15	39.59
N3	WT GSK864	30.54	35.94
N3	R132H CT	28.22	29.02
N3	R132H CT	28.1	30.74
N3	R132H CT	28.31	29.21
N3	R132H GSK864	28.91	29.93
N3	R132H GSK864	28.47	30.77
N3	R132H GSK864	29.26	29.05
N4	WT CT	29.28	
N4	WT CT	28.42	
N4	WT CT	30.21	
N4	WT GSK864	29.36	
N4	WT GSK864	30.56	
N4	WT GSK864	30.49	
N4	R132H CT	32.65	
N4	R132H CT	31.42	
N4	R132H CT	30.2	
N4	R132H GSK864	29.85	
N4	R132H GSK864	31.85	
N4	R132H GSK864	--	

Thus, we see a discrepancy between *PDGFRA* levels detected in RNASeq and results from qPCR validation. *PDGFRA* absolute levels in U87 cells are quite low, as its Ct values are above 29 in many of the samples (Table 19), suggesting low baseline expression levels in this cell line, which could explain the reason of this incongruence. Nevertheless, where detected, the increase in *PDGFRA* expression is not as high as the one reported in Flavahan et al., 2016, where *PDGFRA* is five-fold enriched upon IDH1 mutation. Interestingly, a specific subclass of glioblastomas, called proneural, is characterized by *PDGFRA* overexpression following point mutation or its amplification, but such alterations are rare in tumours with *IDH1* mutation (R. G. W. Verhaak et al., 2010). Thus, it may be that the patient-derived *IDH1* mutant cells analysed in the Flavahan paper bear some abnormalities at the *PDGFRA* locus, whilst our cell line does not. Indeed, while the RNASeq results from the paper were obtained from publicly available cancer genome atlas (TCGA) datasets and did not show any alteration around PDGFRA, their

qPCRs were carried out in the BT142 cell line, which bears an homozygous IDH1 mutation, while in patients and U87 R132H cells, the mutation is heterozygous. Moreover, it is not known whether PDGFRA is WT or not.

Similar results were found for *DNAJC12*, another gene annotated near a lost boundary in the paper, coding for a chaperone involved in complex assembly and protein folding, highly expressed in the brain (Choi, Djebbar, Fournier, & Labrie, 2014). Indeed, RNASeq detects an increase of *DNAJC12* levels in R132H controls compared to the WT (Log2FC = 2.84, P.adj = 0.01), but treatment with GSK864 does not significantly rescue its levels (Log2FC = -1.7, P.adj = 0.6) (Figure 8.7B). Validation using qPCR showed a relative increase in its expression in R132H, albeit not significant, and confirmed no effect of GSK864 treatment. Much like *PDGFRA*, *DNAJC12* baseline expression seems to be low in U87 cells, as its Ct ranges between 30 and 37, depending on the experimental sample (Table 19).

Thus, discrepancies between RNASeq results and qPCR validation make it difficult to choose either *PDGFRA* or *DNAJC12* as an appropriate readout for loss of insulation following decreased CTCF binding. While *PDGFRA* qPCR levels did not mirror the changes observed in RNASeq, *DNAJC12*, despite the variability across replicates, did show the same increase in R132H cells in both experiments.

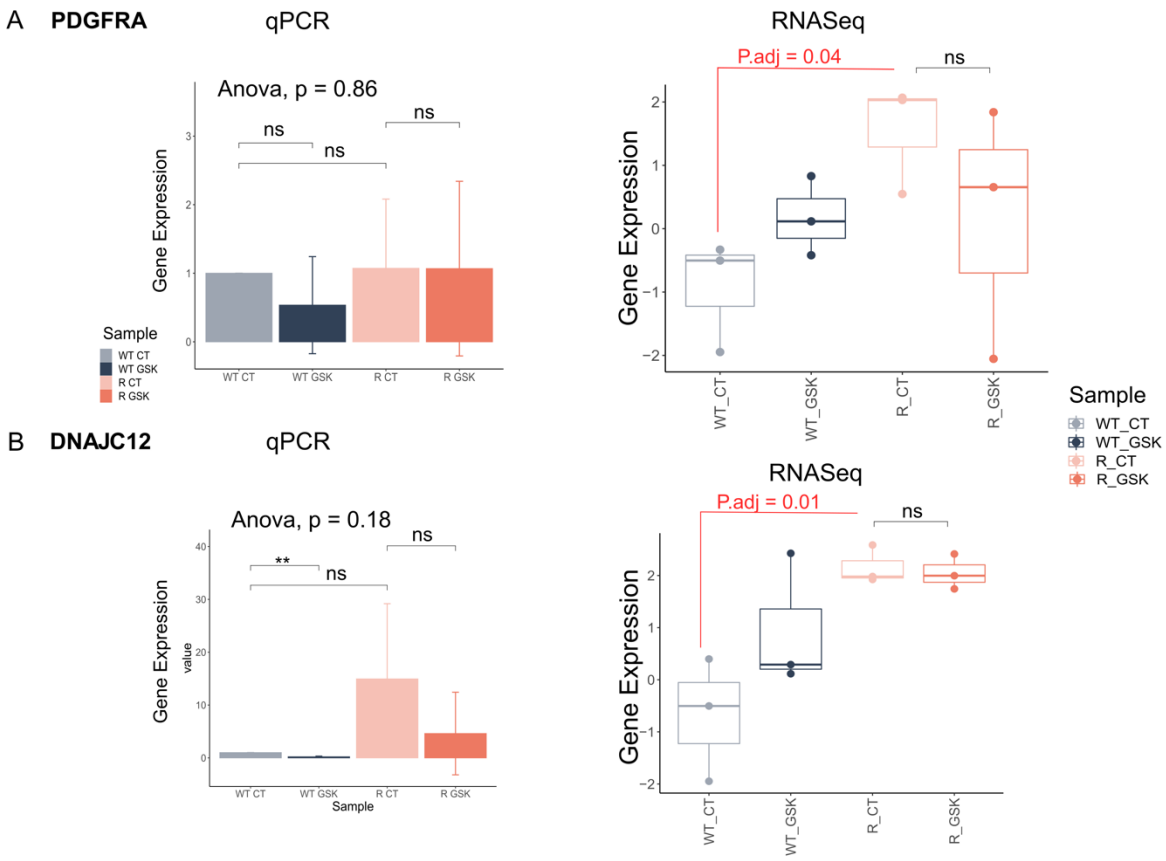


Figure 8.8. Validation of PDGFRA and DNAJC12 expression. *A) Left: Expression levels of PDGFRA (top, N=4 biological replicates) in quantitative PCR (qPCR) experiments. Expression levels were normalized over ACTB and calculated according to the ΔΔCt method. Relative quantification results are shown using WT CT as reference. Right: expression levels of PDGFRA as detected in 3'end mRNA Sequencing (N=3 biological replicates). B) Left: qPCR relative quantification of DNAJC12 (N=3 biological replicates). Statistical significance was assessed both by ANOVA (top of the plot) and by T-Test (comparisons shown). Significance levels: ns = non significant, * = P.Value <= 0.05; ** = P.Value <= 0.01; *** = P.Value <= 0.001. Data are expressed as mean ± standard deviation. Right: RNASeq detected expression of DNAJC12 in N=3 biological replicates. Gene expression levels are shown after TMM normalization, batch effect removal and voom transformation.*

8.6 *1 week dataset confirms DNAJC12 upregulation in R132H cells and little to no change in PDGFRA levels*

In the RNASeq at one week, we detect a decrease in PDGFRA levels in IDH1 U87 mutant cells (Log2FC= -1.3, P.Adj= 1), confirmed by qPCR results (Figure 8.8C, middle panel). In line with what already seen in the 48 hours dataset, GSK864 has no effect on this gene's expression, even after a prolonged treatment (Log2FC R132H GSK864 vs control = 0.17, P.adj = 1).

DNAJC12 levels are still increased between WT and R132H cells (Log2FC= 3.94, P.Adj= 0.17, P.value = 0.01), but are not significantly affected by the week-long treatment with GSK864 (Log2FC R132H GSK864 vs control = -0.32, P.adj = 1), as confirmed by qPCR results (Figure 8.8C, left panel).

Furthermore, a total of six genes were detected in common with the Flavahan annotation, of which three are downregulated in R132H cells (Figure 8.8A and B). GSK864 treatment only increases levels of PTPRN (Log2FC R132H vs WT control = -1.07, P.Adj = 0.05; Log2FC R132H GSK864 vs Control = 0.82, P.Adj = 0.13, P.value = 0.01), which codes for a protein tyrosine phosphatase receptor involved in cell growth and differentiation. qPCR validation of its gene expression confirmed a significant decrease in R132H control cells and no impact of GSK864 treatment (Figure 8.8C, right panel).

Notably, the genes in common between RNASeq and Flavahan's annotation in the 48 hours dataset differ from those detected at one week. This effect is a consequence of the more general observation that RNASeq at 48 hours showed a greater number of differentially expressed genes between WT and R132H controls compared to the same dataset at one week: 2262 and 475 genes, respectively. Potential reasons behind this phenomenon have been discussed more in detail in Chapter 5, but a possibility might be that prolonged treatment with DMSO interferes with the transcriptome of the cells.

Thus, RNASeq results do not align with those shown in Flavahan et al. at either timepoints tested. *PDGFRA* levels are not significantly increasing and are often inconsistent, thus suggesting that *PDGFRA*, at least in U87 cells, is not the appropriate gene to use as readout for loss of insulation between TADs. Slightly more promising results have been shown for *DNAJC12* and *PTPRN*, although whether these genes' expression is suitable to use as a proxy for loss of insulation will require a more in-depth look at their loci. Whether these results mean that the mechanism proposed in the paper is not actually happening in these cells will be proved

only by investigating more thoroughly the epigenetics defects upon which it is based: CTCF binding and DNA methylation.

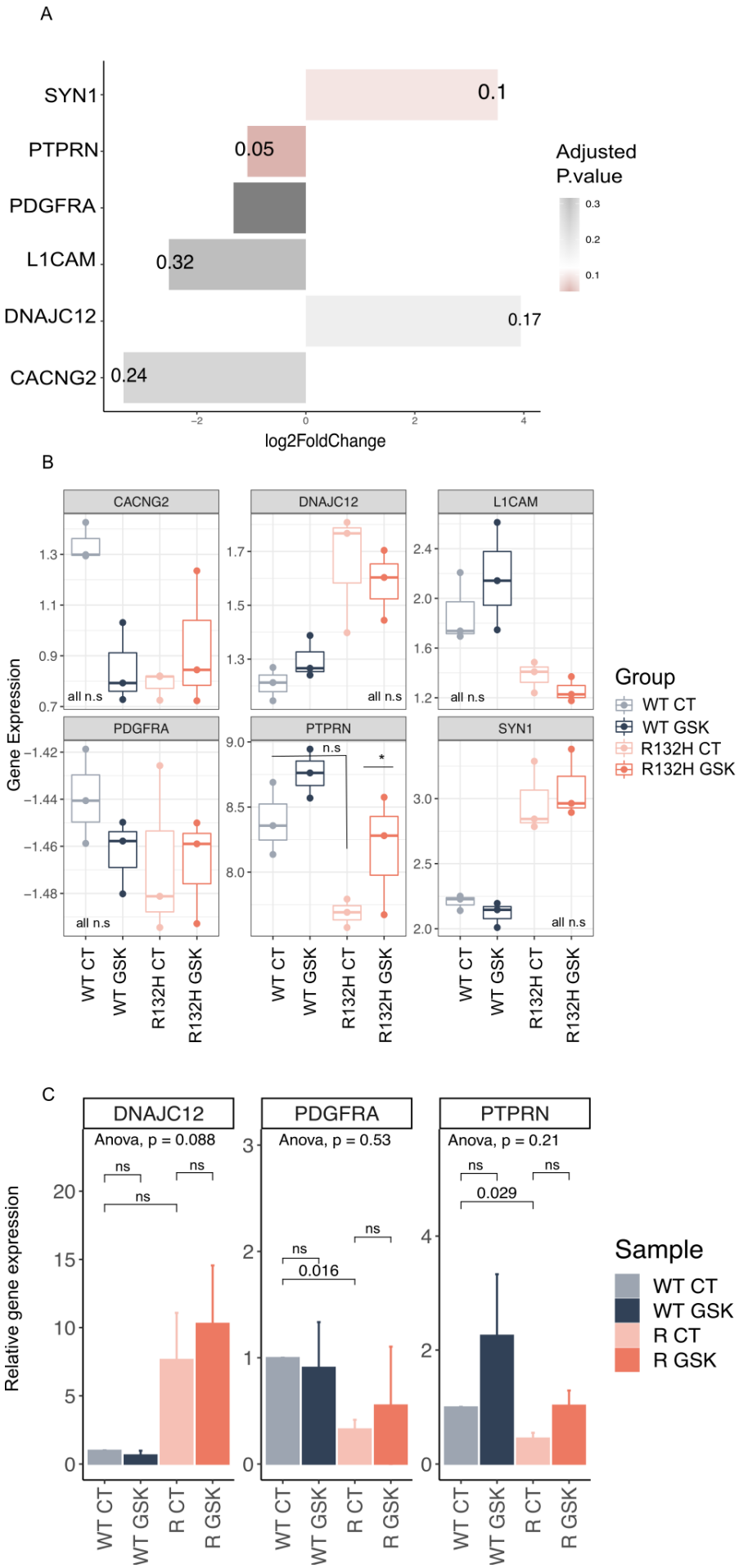
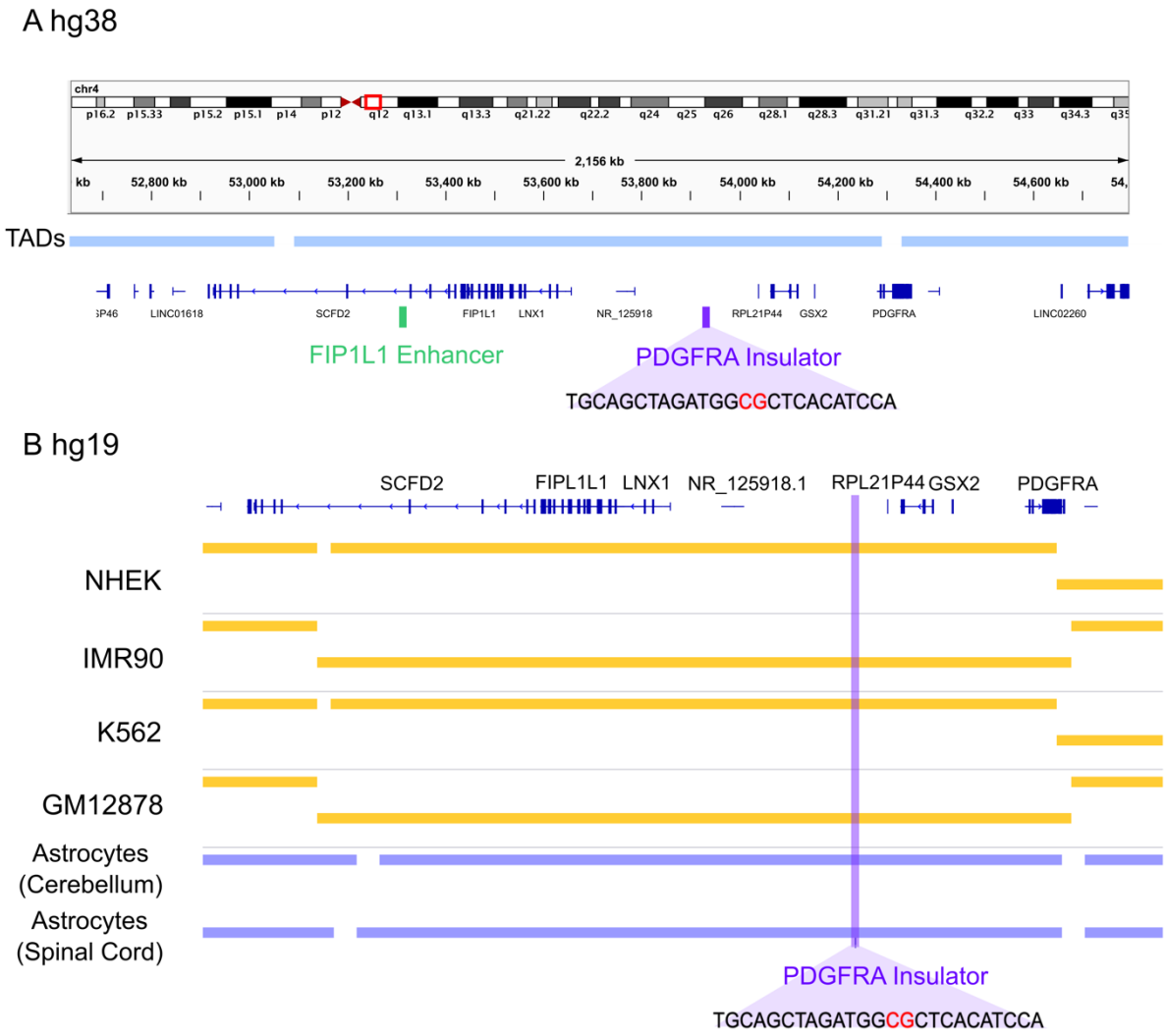

A
SYN1
PTPRN
PDGFRA
L1CAM
DNAJC12
CACNG2
0.1
0.05
0.32
0.17
0.24
Adjusted P.value
0.3
0.2
0.1
log2FoldChange
B
CACNG2
DNAJC12
L1CAM
PDGFRA
PTPRN
SYN1
all n.s
n.s
*
Gene Expression
WT CT
WT GSK
R132H CT
R132H GSK
Group
C
DNAJC12
PDGFRA
PTPRN
Anova, p = 0.088
Anova, p = 0.53
Anova, p = 0.21
ns
0.016
0.029
Relative gene expression
WT CT
WT GSK
R CT
R GSK
Sample

Figure 8.9. Gene expression levels of Flavahan related genes in 1 week dataset. *A) Bar plots showing log2Fold change between R132H and WT control cells and adjusted P.values for Flavahan related genes detected in the one week RNASeq dataset. Color scale mirrors significance levels, with grey being non-significant and red being significant. B) Boxplots representing normalized log10 expression levels of six genes in common with the Flavahan dataset. Out of six, only PDGFRA is not differentially expressed. Differential expression was calculated with DESeq2, using the linear model ~ Genotype + Genotype*Treatment (Chapter 3). Significance levels for the comparisons WT CT vs R132H CT and R132H CT vs R132H GSK864 are indicated as following: Not significant = ns; P > 0.05; * P < 0.05; ** P < 0.01; *** P < 0.001. (N=3 per group). C) qPCR validation for three genes of interest. Expression levels were normalized over ACTB and calculated according to the ΔΔCt method for N=3 biological replicates. Relative quantification results are shown using WT CT as reference. Statistical significance was calculated with the R/Bioconductor package ggpubr, using either ANOVA or a Student's T-test with the following comparisons : R132H control vs WT control, WT control vs WT GSK864, R132H Control vs R132H GSK864. Data are expressed as mean ± S.E.M.*

8.7 No change of CTCF binding around PDGFRA insulator

According to Flavahan et al., PDGFRA overexpression upon IDH1 mutation is due to loss of a boundary between PDGFRA and an upstream enhancer located within the SCFD2 gene. Genomic coordinates of both the insulator and the upstream enhancer were converted from hg19 to hg38, resulting in the positions described in Figure 8.9. Notably, the Insulator described in Flavahan et al., 2016 does not fall at the boundary between two TADs, but rather in the middle of a single one (Figure 8.9A). In order to exclude potential problems in the transfer of genomic coordinates from hg19 to hg38, a representation of the TADs annotations used in the 2016 paper, under the hg19 genome build, is provided in Figure 8.9B. Indeed, even in the datasets used by Flavahan et al. to annotate TAD boundaries, the Insulator coordinates fall within, rather than between, TADs (Figure 8.9B, yellow tracks). This could mean that the Insulator is lining the boundaries of a sub-TAD, rather than a bigger TAD, as in their paper, the corresponding HiC map shows a contact between the Insulator and the enhancer.



Locus	Coordinates (hg19)	Coordinates (hg38)
FIP1L1 Enhancer	Chr4: 54,181,104	Chr4:53,314,937-53,314,958
PDGFRA Insulator	Chr4: 54,796,343	Chr4:53,930,176-53,930,201

Figure 8.10. Description of PDGFRA locus. *A) PDGFRA locus annotation according to the hg38 genome build. IGV tracks showing chromosome 4 ideogram, cerebellum Astrocytes TADs distribution (light blue) and RefSeq genes around PDGFRA. TADs track data for U87 cells is not currently deposited in a database, so two publicly available tracks from the ENCODE dataset were selected: ENCSR011GNI, from astrocytes of the cerebellum, and ENCSR974LXZ (shown), from astrocytes of the spinal cord (Dunham et al., 2012). BED files were converted from hg19 to hg38 with LiftOver (https://genome.ucsc.edu/cgi-bin/hgLiftOver). These samples were selected because U87 cells are a glioblastoma cell line, which is a cancer originating from astrocytes, and as TADs are considered to be stable across different cell types, the assumption was that these tracks would be the closest representative of TADs distribution in our own cell lines (Barutcu et al., 2018; Dixon et al., 2012). Positioning of the FIP1L1 enhancer (green) and PDGFRA insulator (purple) is indicated beneath the Refseq gene track. Sequence of the CTCF binding site in the Insulator is provided below, and methylation-sensitive CGs are highlighted in red. B) Top: PDGFRA locus annotation according to the original hg19 genome build, in the following order: RefSeq gene track illustrates the position of genes around the PDGFRA locus, with a purple box indicating the position of the Insulator; TADs tracks of NHEK, IMR90, K562, GM12878, as used in Flavahan et al., for the mapping of TAD boundaries (all from publicly available data* (Lieberman-Aiden, Van Berkum, et al., 2009; Rao et al.,

2014)*; Astrocytes TADs distribution from (Dunham et al., 2012). Bottom: Genomic coordinates of FIP1L1 Enhancer and PDGFRA Insulator in hg19 and hg38.*

To better understand TADs distribution around this locus, a contact map from the IMR90 dataset was analysed (Figure 8.10, Lieberman-Aiden et al., 2009). Here, three corner dots, representing a contact between the sequences at the bases of the triangle, are of particular interest. The PDGFRA Insulator seems to be indeed located at a boundary, but of a sub-TAD rather than a TAD (Figure 8.10, circle 2), as it is contained within a much bigger domain, spanning from upstream *SCFD2* to *PDGFRA* (chr4:53.9 Mb – 55.1 Mb, Figure 8.10, circle 1). Furthermore, the presence of a corner dot domain (Figure 8.10, circle 3) is seemingly indicating that a contact between the FIP1L1 enhancer (green) and PDGFRA might already exist in a WT setting, rather than being formed upon IDH1 mutation. Thus, analysis of publicly available data supports the presence of the Insulator at a sub-TAD boundary, explaining the discrepancy between TADs annotation and boundaries described in Flavahan (2016) at the PDGFRA locus.

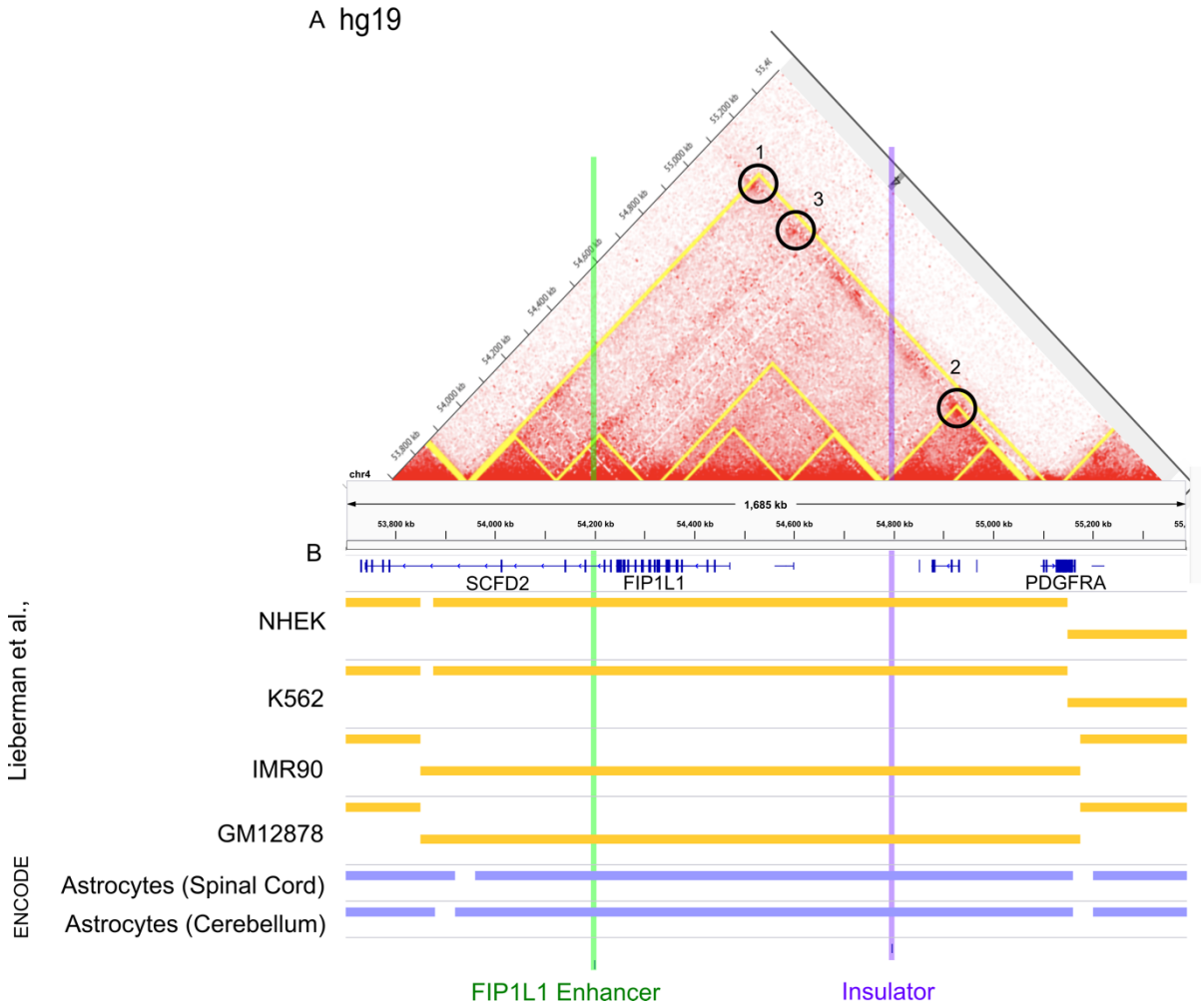


Figure 8.11. TADs distribution around the PDGFRA locus (hg19). *A) HiC contact map of IMR90 cells from Lieberman-Aiden et al., 2009. This is part of the publicly available data used in Flavahan et al., 2016 to map TAD boundaries. The map was visualized using a web instance of Juicebox,* (J. T. Robinson et al., 2018)*. Yellow lines enclose domains according to the IMR90 combined domains annotation from Rao et al., 2014. Numbered circles enclose corner dots domain of interest. Corner dots represent a contact between the sequences at the base of the equilateral triangle of which the corner dot is the peak. B) PDGFRA locus annotation according to the original hg19 genome build, as in Figure 8.9: RefSeq gene track, with a purple line indicating the position of the Insulator and a green one representing the upstream FIP1L1 enhancer; TADs tracks of NHEK, IMR90, K562, GM12878, as used in Flavahan et al., for the mapping of TAD boundaries (all from publicly available data* (Lieberman-Aiden, Van Berkum, et al., 2009; Rao et al., 2014)*, and two TADs tracks from astrocytes from ENCODE* (Dunham et al., 2012)*.*

The PDGFRA insulator (Chr4:53930176-53930201) corresponds to the CTCF binding site lost upon IDH1 mutation in the Flavahan analysis. Analysis of our CTCF ChIPSeq revealed the presence of a decreased CTCF peak, albeit not statistically significant (Figure 8.11B, WT CT (grey) compared to R132H CT (light red)). Further examination of the dataset shows four statistically significant CTCF peaks lost in the IDH1 mutant, two of which fall between

PDGFRA and the upstream FIP1L1 enhancer (Figure 8.11A, stars). Of note, R132H CTCF ChIPSeq tracks in Figure 8.11A look greatly decreased, but only four peaks were statistically significant. This may be due to the filtering threshold applied during ChIPSeq analysis, because only peaks with a number of tag counts higher than 32 across all samples were selected for differential enrichment analysis, as a way to ensure robust results (as per ChIPSeq analysis protocol in Turcan et al., 2018).

Table 20. Enrichment levels of CTCF peaks around the PDGFRA Insulator.

Levels of CTCF binding loss in R132H vs WT control samples are expressed in fold-change, with the respective Adjusted P-values, as corrected by the Benjamini-Hochberg method. As the software only calculates fold-change and adjusted p-value for significantly different peaks, and none of these were significantly changing upon GSK864 treatment, levels of CTCF binding between R132H GSK and control cells are expressed as Tag counts.

Peak number	R132H vs WT control		R132H GSK vs R132H control	
	Fold Change	P.adj	Tag Counts R132H GSK	Tag Counts R132H Ctrl
Peak 11101	-1.76	7.02e-33	76.34	88.76
Peak 11105	-1.5	8.04e-22	72.84	81.09
Peak 11111	-1.5	1.21e-23	69.22	66.97
Peak 11122	-1.5	6.23e-28	87.35	95.82

A 48-hour treatment with GSK864 does not affect CTCF binding at the Insulator in WT cells, and despite showing a trend toward increasing CTCF levels in R132H cells, this increment is not statistically significant (Tag Counts: WT CT= 124.1, WT GSK864 = 82.17 , R132H CT = 87.22 , R132H GSK864 = 97.34; Figure 8.11B). None of the significant CTCF peaks lost in R132H are rescued by GSK864 treatment (Table 20).

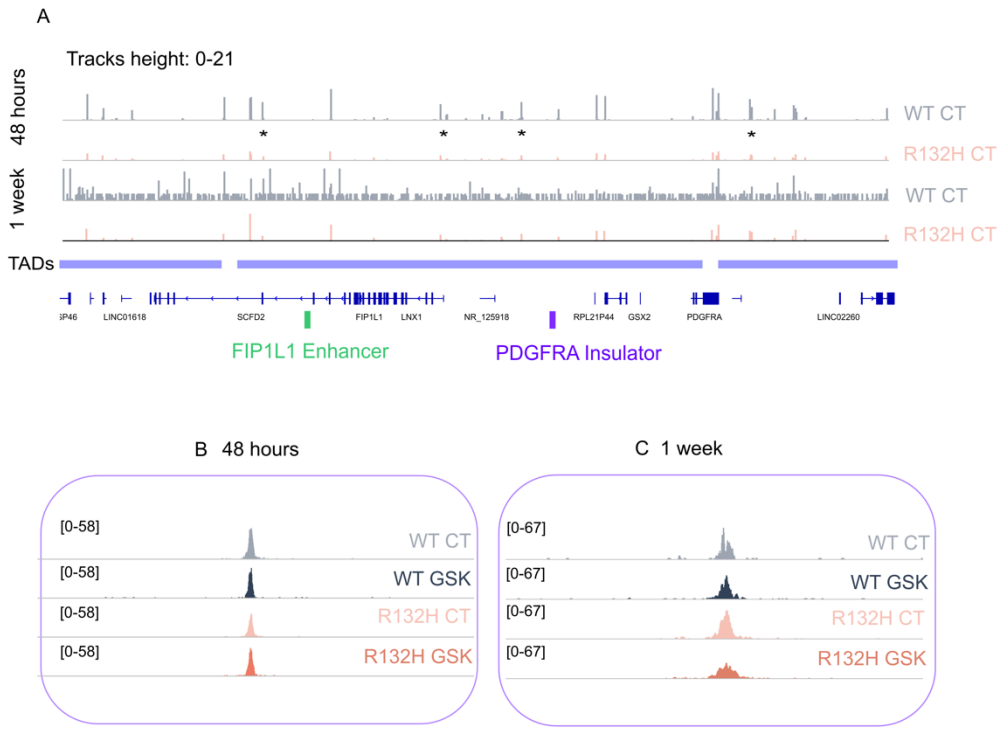


Figure 8.12. CTCF ChIPSeq shows no statistically significant loss of CTCF around PDGFRA Insulator. *A) Panel showing IGV Tracks from 48 hours (top) and one week (bottom) CTCF ChIP around the whole PDGFRA locus. Stars are drawn over statistically significant peaks. Only WT and R132H control tracks are shown. B-C) Zoom in on the PDGFRA insulator CTCF peak, across all conditions, in the 48-hour (B) or one week (C) datasets. IGV Tracks data range is shown in brackets. Bigwig tracks were generated by merging all biological replicates (48 hours: N=3; one week: N=2), then normalizing to 1x coverage using bamCoverage from Deeptools (*(Ramírez, Ryan, Bj¨, et al., 2016)*, further details in the Methods section).*

Data from the 1 week treatment dataset shows similar levels of CTCF between WT and IDH1 mutant around the insulator in Ctrl samples (WT Ctrl Tag counts = 147.05; R132H Ctrl Tag counts: 168.52, n.s., Figure 8.11C), and no significant effect of a prolonged treatment with GSK864 (R132H Ctrl Tag counts: 168.52, R132H GSK Tag Counts = 184.06, n.s.). Unlike its 48 hour counterpart, no significantly enriched peaks were found along the PDGFRA locus, despite general levels being decreased.

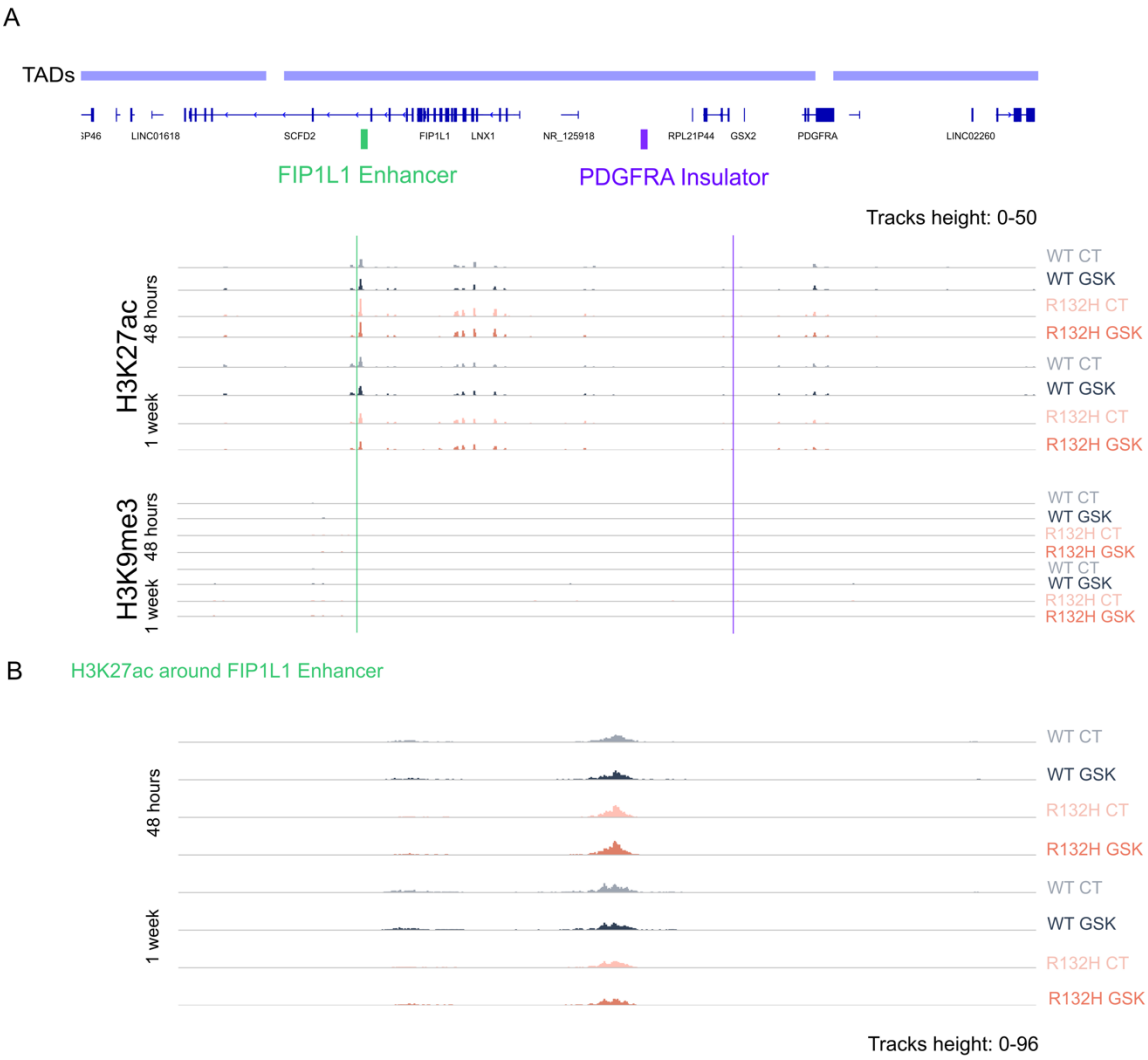


Figure 8.13. Distribution of H3K27ac and H3K9me3 around the PDGFRA locus. *A) Top: Diagram representing astrocytes TADs distribution, RefSeq gene track and the genomic positions of the FIP1L1 enhancer and Insulator, as described in Flavahan et al., 2016. Bottom: Distribution of H3K27ac and H3K9me3 around the PDGFRA locus. Bigwig files were normalized to 1x coverage, then tracks were grouped and scaled to the same height in the IGV genome browser. A purple and green line indicate the position of the Insulator and the FIP1L1 enhancer, respectively. B) Zoom in on the FIP1L1 enhancer and H3K27ac distribution in the 48 hours or one week ChIPSeq datasets.*

Analysis of other chromatin marks around the PDGFRA locus shows no H3K9me3 distribution, but a H3K27ac peak is present near the FIP1L1 enhancer (Figure 8.12A). However, distribution of H3K27ac at the enhancer does not show any significant difference across the experimental conditions (WT Ctrl Tag counts = 513.145; R132H Ctrl Tag counts: 493.08, n.s., Figure 8.11B).

CTCF ChIPSeq results do not confirm loss of CTCF around the PDGFRA insulator as described in Flavahan et al., 2016. Treatment with GSK864, after 48 hours or one week, does not affect

CTCF levels, suggesting that while the molecule efficiently restores the metabolic defect, it does little to correct the epigenetic impairments.

8.8 No detection of a CTCF peak around Flavahan's annotated DNAJC12 boundary

We next looked for CTCF loss over the annotated boundary for DNAJC12, another gene described in the publication which we find upregulated in R132H cells in both our RNASeq datasets and qPCRs. Unlike PDGFRA, at this locus the authors only provide a single-nucleotide coordinate indicating the putative boundary (Figure 8.13C, red box), but do not annotate the full CTCF binding site sequence, nor they give any indications regarding its methylation-sensitive status. Annotated coordinates place the boundary in the middle of the *RUFY2* gene, separating *DNAJC12* from a downstream group of genes, including *TET1*. According to the authors, upon boundary loss, gene expression levels of *DNAJC12* correlate well with those of *CCAR1*, whilst *TET1* levels lose correlation with its neighbours *STOX1* and *DDX21*. *DNAJC12* is the only gene from this locus detected in RNASeq (Figure 8.7 and 8.8).

Visualization of the HiC contact map from IMR90 cells, as used in Flavahan (2016) to map TAD boundaries, confirms the presence of a boundary between two TADs around the middle of the *RUFY2* gene (Figure 8.13A). Interestingly, *CCAR1*, whose expression levels gain correlation with DNAJC12 upon IDH1 mutation according to the Flavahan model, is also positioned at a boundary between two sub-TADs, one containing *TET1* and the other *STOX1* and *DDX21*. In Flavahan (2016), the expression of these three genes is described to lose correlation upon IDH1 mutation, but the authors assume they are placed in the same TAD.

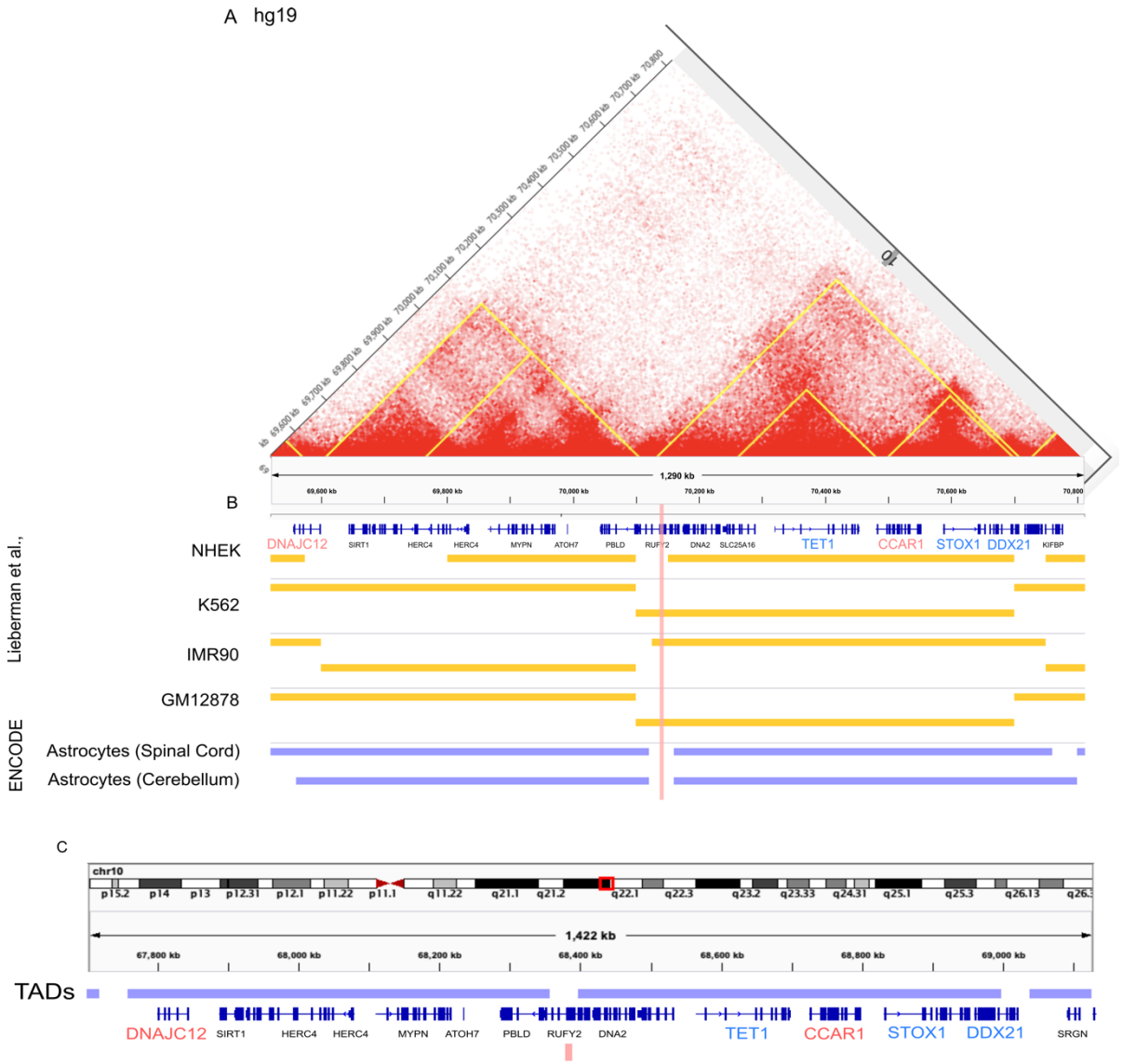


Figure 8.14. DNAJC12 locus description. *A) HiC contact map of the DNAJC12 locus in hg19. Contact map of IMR90 cells was visualized using a web instance of Juicebox, using publicly available data* (Lieberman-Aiden, Van Berkum, et al., 2009; J. T. Robinson et al., 2018)*. Yellow lines represent IMR90 domains as annotated in Rao et al., 2015. B) IGV visualization of the DNAJC12 locus (hg19). Refseq genes have been colored according to the correlation of their expression levels according to Flavahan (2016). Indeed, according to the authors, upon IDH1 mutation DNAJC12 expression gains correlation with that of CCAR1 (red), while expression of TET1 correlates less with that of STOX1 and DDX21 (blue). Publicly available TADs annotations from NHEK, K562, IMR90 and GM12878 (yellow tracks) are the same used in Flavahan (2016). Two TADs track from different types of Astrocytes from ENCODE have been added, as they are closest to our cellular model (purple)* (Dunham et al., 2012; Lieberman-Aiden, Van Berkum, et al., 2009)*. A red line indicated the boundary annotated in Flavahan et al., 2016. C) DNAJC12 locus in hg38 genome build. Chromosome 10 ideogram, TADs annotation (purple), Refseq Gene tracks and boundary localization as annotated in Flavahan et al 2016, after conversion to hg38 with LiftOver (see Figure 8.10). According to their paper, upon boundary loss (red rectangle, middle of RUFY2 gene), gene expression levels of DNAJC12 correlate better with those of CCAR1 (red), normally placed in a separate TAD, and as such showing different expression patterns. In contrast, levels of genes belonging to the same TAD, such as TET1, STOX1 and DDX21 (blue) lose correlation. In our dataset, however, only DNAJC12 is differentially expressed in RNASeq.*

Unlike around the PDGFRA insulator, analysis of CTCF distribution in the 48-hour and one week datasets finds no clear peak around the hg38-converted coordinates for the boundary (chr10: 68380243-68380244), which falls exactly between two TADs (Figure 8.14). Indeed, peak calling did not detect anything around this very coordinates, and normalized coverage was very low in both datasets, as indicated by the range of values in IGV tracks (Figure 8.14F and G). Zooming out on the locus, we find two CTCF peaks at the edges of each TAD, however only one of them is detected by peak calling and while it is decreasing in the IDH1 mutant, the difference is not statistically significant (48 hours: WT Tag counts = 108.13, R132H Tag counts = 73.93, ns, Figure 8.14D and E). Analysis of other chromatin marks revealed an absence of H3K9me3 and a small peak of H3K27ac around the RUFY2 promoter. Although there were no differences between the experimental samples at 48 hours (WT Ctrl Tag counts = 62.43; R132H Ctrl Tag counts: 77.24, n.s.), in the one week dataset a significant increase in its H3K27 acetylation levels was detected (Log2FC: 0.79, P.adjusted = 1.53e-28; Figure 8.14C).

The lack of a proper peak around the annotated coordinates given, combined with the poor detection of other genes in this locus aside from *DNAJC12*, suggest that DNAJC12 locus might not actually be subjected to boundary loss, or if it is, this loss is not due to a decrease in CTCF binding.

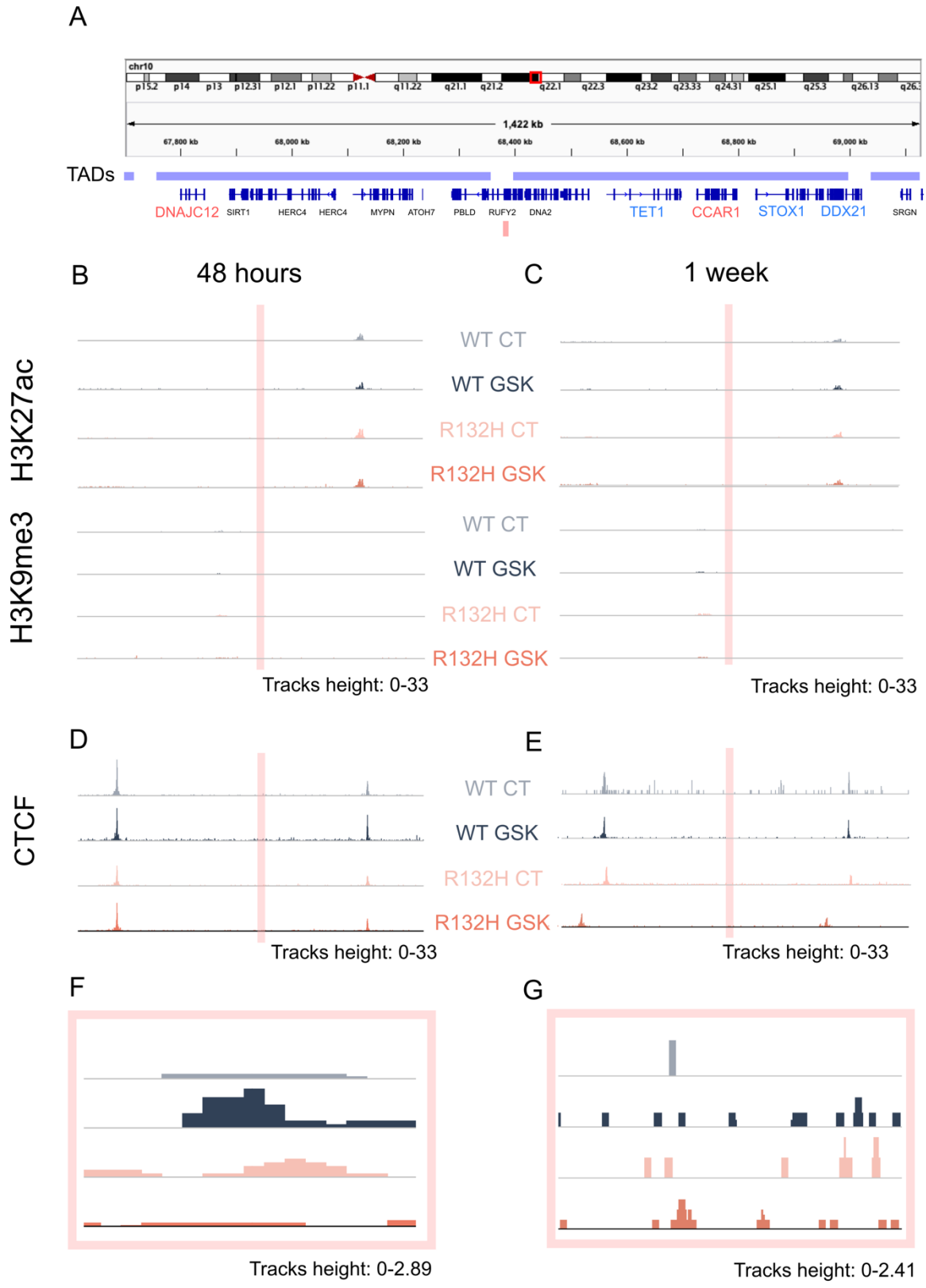


Figure 8.15. Lack of CTCF peak around annotated boundary for DNAJC12 locus. *A) Schematic description of DNAJC12 locus. B-C) Distribution of H3K27ac and H3K9me3 marks around the genomic coordinates that surround the putative boundary site (red bar) in either the 48-hour (B) or one week (C) datasets. D-E) Distribution of CTCF around the genomic coordinates surrounding the DNAJC12 boundary site described in Flavahan et al., 2016. F-G) Zoom-in on the annotated boundary (red bar) in the 48-hour (F) or one week (G) CTCF ChIPSeq datasets. No peak was detected through*

peak calling. All bigwig files have been normalized to 1x coverage as described in Figure 8.11, then resulting IGV tracks have been grouped together and auto-scaled to the range specified at the bottom.

8.9 No CTCF binding site at PTPRN Flavahan-predicted boundary

Thus far, we confirmed CTCF binding on the predicted insulator in the PDGFRA locus, even though its levels were not statistically different between WT or IDH1 mutant cells, but we did not detect any CTCF at the predicted coordinates in the DNAJC12 locus.

PTPRN is another gene in common with the Flavahan annotation, whose levels are significantly decreased in R132H cells at both timepoints. Visualization of the HiC contact map in IMR90 reveals a complex higher order chromatin structure. PTPRN and the neighbouring TUBA4A appear to be localised within the same sub-TAD, with TTLL4 residing in a separate TAD (Figure 8.15A). This structure is maintained in the IGV TADs tracks for IMR90, K562, NHEK and GM178128 cells (Figure 8.15B, yellow tracks), but not in astrocytes (purple tracks), the closest cell type to U87 cells. To further complicate the picture, two different boundaries are annotated as being lost upon IDH1 mutation, highlighted in green in Figure 8.15B. According to Flavahan et al., in an IDH1 mutation setting, expression of *PTPRN* correlates well with *TTLL4* (red), but not with its upstream neighbour *TUBA4A* (blue). Given the multiple boundaries and complex hierarchical structure of TADs, it is difficult to understand if and how this particular locus is rearranged upon IDH1 mutation without a dedicated chromosome conformation capture experiment, although one hypothesis could be that by losing both boundaries (green), the whole locus could be forming a single TAD, with one boundary starting upstream of *TTLL4* (219.5 Mb) and another one downstream of *PTPRN*.

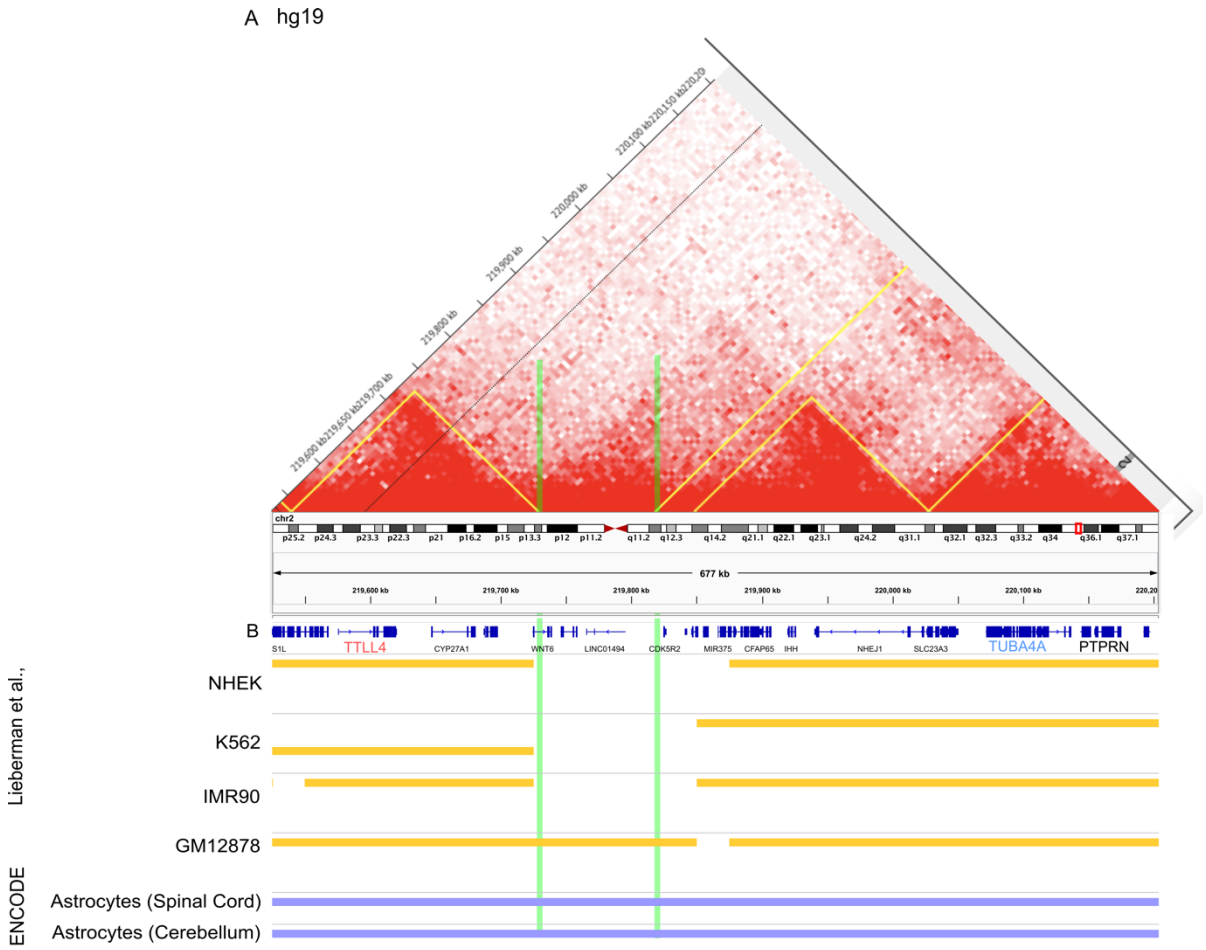


Figure 8.16. Distribution of TADs around the PTPRN locus (hg19). *A) HiC contact map of IMR90 cells around the PTPRN locus, visualized with Juicebox and annotated with IMR90 domains as described in Figure 6.13. Green lines represent boundaries annotated in Flavahan et al., 2016. B) IGV visualization of the locus, with tracks following the same order as Figure 13. Genes gaining correlation with PTPRN expression are colored in red, while those losing correlation are indicated in blue.*

We next investigated the CTCF binding pattern around the Flavahan-annotated boundaries in this genomic region, annotated in the 2016 papers at Chr2:218955278-218955279 (boundary 1, Figure 8.16A) and Chr2:218865278-218865279. According to the bioinformatic analysis presented in the paper, *PTPRN* expression levels increase upon IDH1 mutation, correlating less with the neighbouring *TUBA4A* and more with the upstream *TTLL4*. However, in our cellular model *PTPRN* expression significantly decreases in R132H cells compared to the WT, perhaps due to differences in the transcriptome of U87 cells compared to patient-derived samples. Indeed, H3K27ac distribution around this locus is generally decreased in R132H cells at both

timepoints (Figure 8.16A). Further inspection in the IGV Genome browser revealed that, much like DNAJC12, no clear CTCF peak is present at the predicted boundary 1 or 2 (Figure 8.16A, red boxes). Zooming into the annotated boundary locus (1), a very small peak nearby can be found, even though no significant difference is detected amongst experimental samples at either timepoints (Figure 8.16B and C). A similar result is found zooming into the second annotated boundary (2): a CTCF peak is nearby, however it bears no significant differences between the experimental samples at either timepoints (Figure 8.16D and E). Moreover, while according to Flavahan's annotation *PTPRN* and *TLL4* are supposed to be in two separate TADs, both astrocyte TADs annotations place these genes within the same domain (Figure 8.16A).

Thus, *PTPRN* is yet another example of discrepancies between astrocytes TADs annotation and Flavahan-predicted boundaries. Thus, much like *PDGFRA* and *DNAJC12*, it cannot be used as a readout for loss of insulation induced by IDH1 mutation.

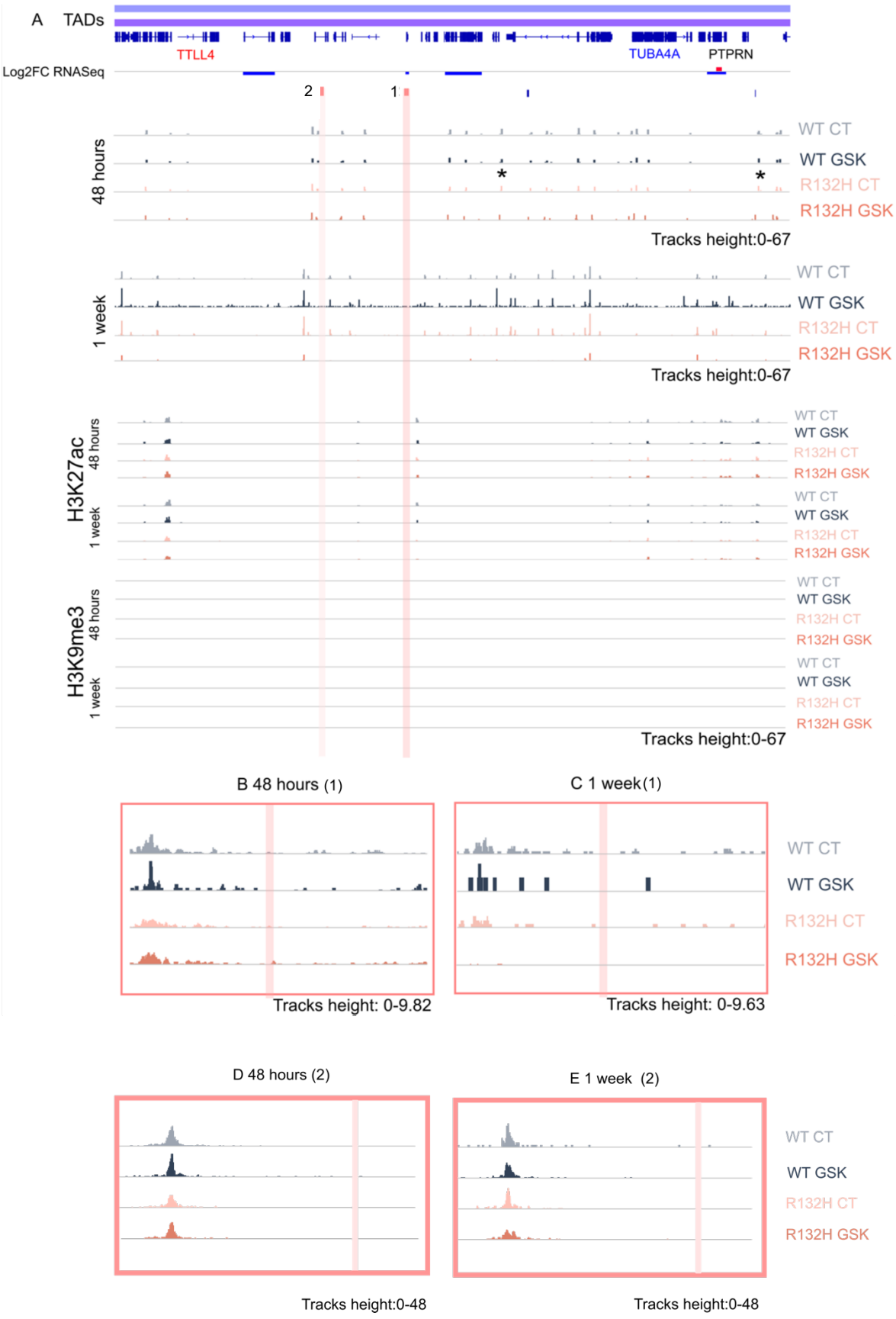


Figure 8.17 No clear CTCF binding site at PTPRN predicted boundary. *A) IGV tracks annotating PTPRN locus, in the following order: two Astrocytes TADs annotation, as presented in Figure 8.9; RefSeq track with genes losing(blue) or gaining (red) correlation with PTPRN upon IDH1 mutation, as*

predicted by Flavahan et al.2016; Barplot of Log2FC of differential gene expression from 48 hours RNASeq; CTCF, H3K27ac and H3K9me3 distribution around the PTPRN locus. Position of the predicted insulator has been highlighted with a red box. B) Zoom-in on the predicted insulator (1) locus at 48 hours. C) Zoom into the insulator locus (1) at one week. D) Zoom-in on the predicted insulator (2) at 48 hours. E) Zoom-in at the predicted insulator (2) at one week. Tracks in A through E belonging to the same dataset have all been grouped and scaled to the common height written below.

8.10 No significant change in CTCF binding detected at the SOX2 locus

Modrek et al. confirmed the model proposed by Flavahan in their three-hit glioma model around the *SOX2* gene (Modrek, Golub, Khan, Bready, Prado, Bowman, Deng, Zhang, Rocha, Raviram, Lazaris, Stafford, Leroy, et al., 2017). Within low-grade astrocytoma, upon mutation of IDH1 and loss of both *ATRX* and *p53*, loss of CTCF binding causes a dissociation between SOX2 and its enhancer, thus reducing its expression levels and blocking differentiation (Figure 8.18A). According to the authors, such interaction is strictly dependent on SOX2 expression levels, and is cell type-specific (Modrek et al., 2017). Indeed, HiC contact maps from publicly available data in two different cell types, human embryonic stem cells and IMR90, shows a contact between SOX2 (purple line) and a downstream promoter (green line) only in the former, where this gene is expressed at high levels (Figure 8.17A, black circle). Thus, their three-hit model should recapitulate a domain distribution of IMR90 cells, with the difference that, upon IDH1 mutation and loss of ATRX and p53, this contact, present in a WT setting, is now lost.

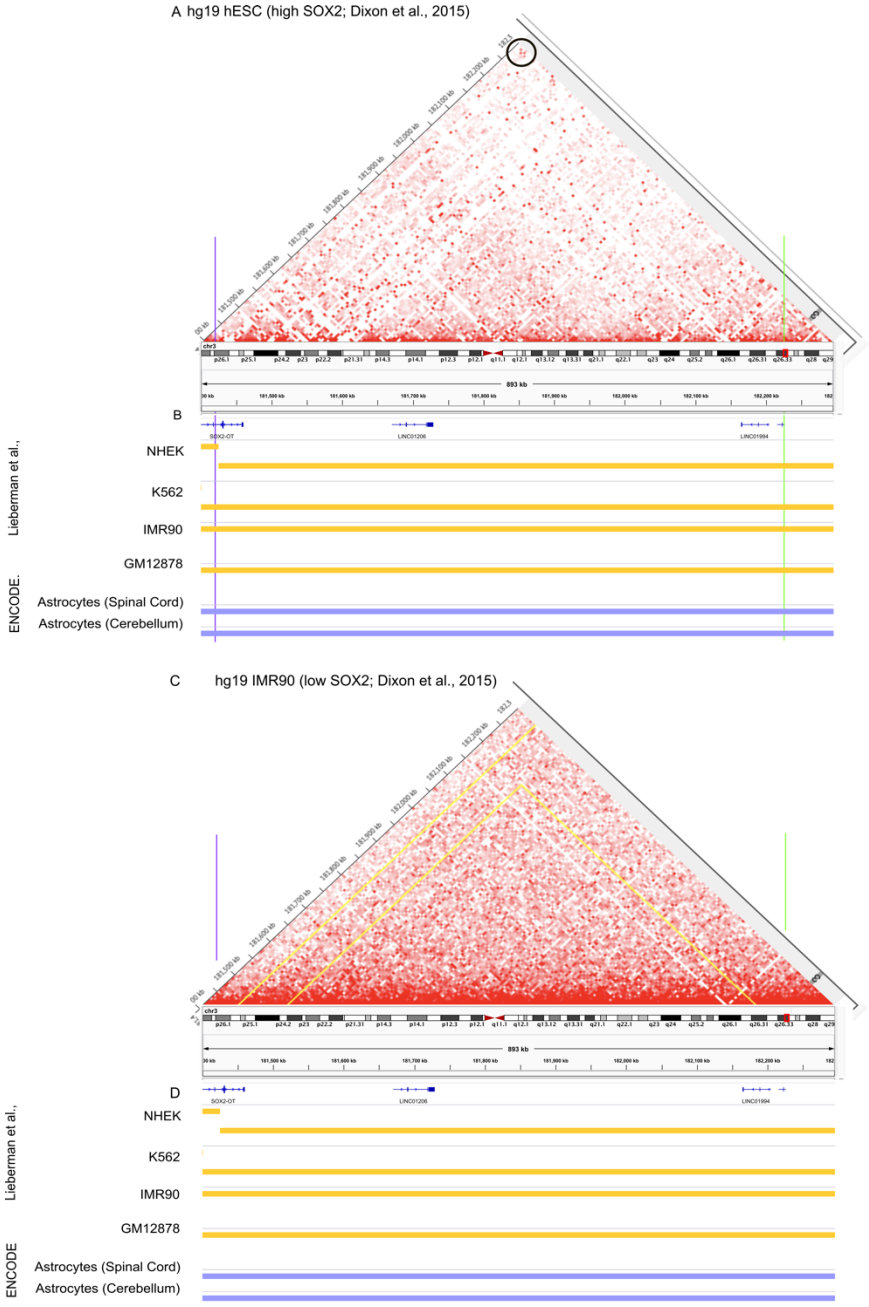


Figure 8.18. SOX2 locus shows cell type-dependent TAD boundaries. *A) HiC contact map showing TADs distribution in human Embryonic stem cells (hESC), as published in* (Dixon et al., 2015)*. The data has been used in Modrek et al. 2107 to show the presence of a contact between SOX2 (purple line) and its downstream enhancer (green line). Such contact is enclosed in a circle. According to the authors, this contact is only present in cells expressing high levels of SOX2. HiC map has been visualized with an online instance of Juicebox, as previously described (Figure 6.10). B and D) TADs annotation in publicly available data: NHEK, K562, IMR90, GM12878 and two Astrocytes tracks from ENCODE* (Dunham et al., 2012; Lieberman-Aiden, Miiro, et al., 2009)*. C) HiC contact map of the SOX2 locus in IMR90 cells, which show low expression of SOX2. Data from Dixon et al., 2015 was visualized with Juicebox, as previously described (Figure 8.10). Yellow lines represent IMR90 domains annotation* (Rao et al., 2014)*.*

We next verified whether a similar mechanism is occurring in our cellular model. In contrast with the Modrek paper, *SOX2* expression is not statistically different between WT and R132H U87 cells (48 hours: Log2 fold-change = -0.67, P.adjusted = 0.26; not detected in one week dataset). Even though two CTCF peaks are detected by ChIPSeq at the coordinates annotated in the paper, its binding is not significantly lost upon IDH1 mutation (48 hours: Peak 10167 WT CT Tag counts = 310.07, R132H CT = 223.88, Peak 10168 WT CT Tag counts = 160.86, R132H CT = 118.4; one week: WT CT= 476.01, R132H CT = 391.78, only one peak annotated; Figure 8.18B). Furthermore, no H3K27ac nor H3K9me3 are detected at any of the timepoints tested (Figure 8.18B).

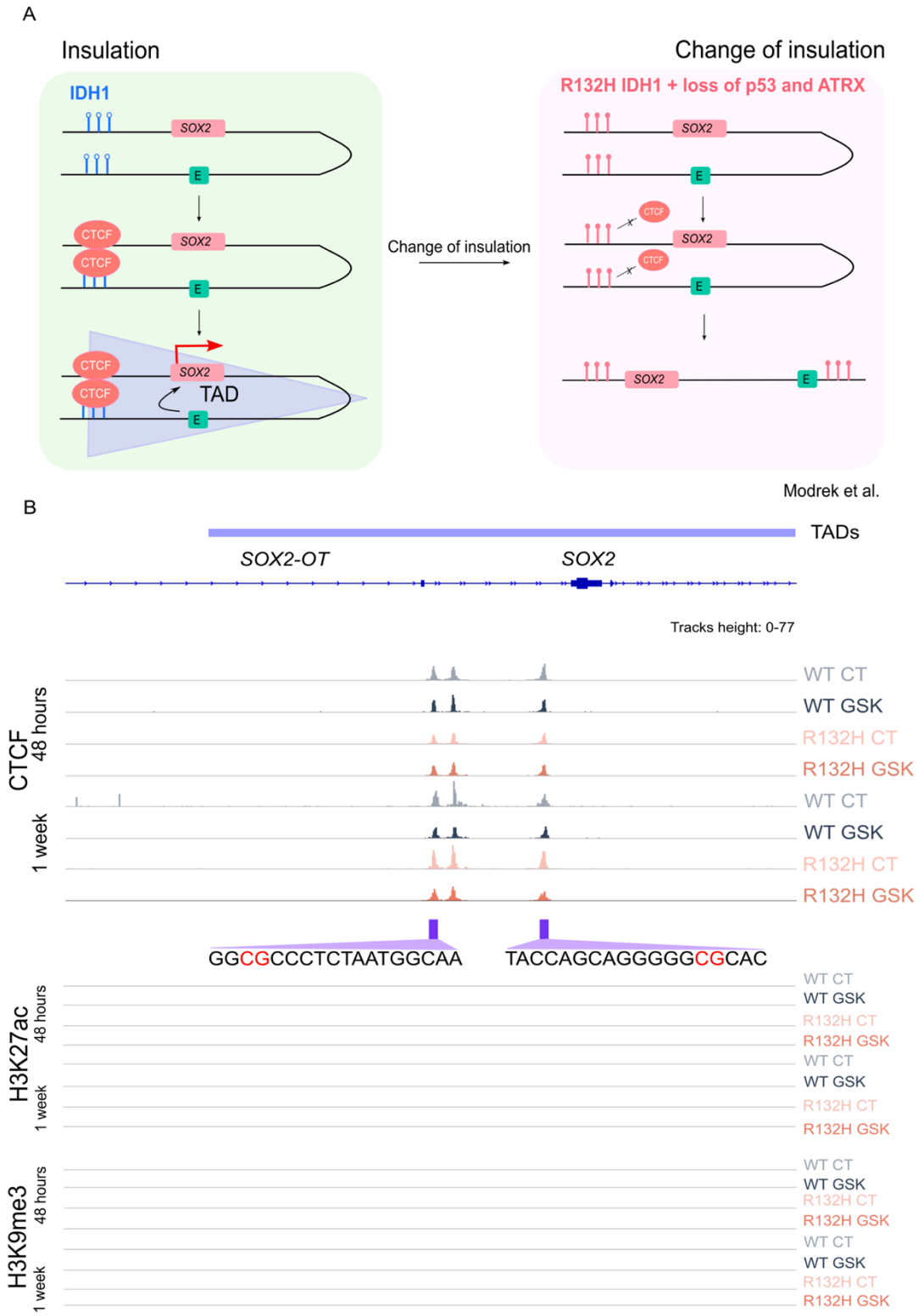


Figure 8.19. CTCF binding remains unchanged between WT or R132H U87 cells at the SOX2 locus *A) Graphical summary of the mechanism proposed by Modrek et al. around the SOX2 locus. Briefly,in normal conditions the SOX2 promoter is directly in contact with its enhancer (E), thanks to CTCF binding forming a loop. Upon IDH1 mutation, loss of ATRX and p53, following DNA hypermethylation, CTCF binding is impaired, thus failing to form the loop that would bring SOX2 promoter and its enhancer in physical contact. This leads to a decreased expression of SOX2, inducing a block of differentiation in low-grade astrocytoma cells* (Modrek, Golub, Khan, Bready, Prado, Bowman, Deng, Zhang, Rocha, Raviram, Lazaris, Stafford, Leroy, et al., 2017; Raineri & Mellor, 2018). *B) IGV tracks of 48 hours and one week CTCF ChIPSeq around the SOX2 locus,*

followed by distribution od H3K27ac and H3K9me3. Bigwig files have been normalized to 1x coverage, then grouped and scaled to the same height (0-77). Purple boxes indicate the CTCF binding sites of interest, as described in the Modrek paper, together with their sequence. DNA methylation sites are highlighted in red.

Notably, the cellular model in their paper bears not only *IDH1* mutation, but loss of *ATRX* and *p53*, which is bound to disrupt the epigenetic state and induce greater alterations in gene expression, while our U87 cells still express both (Table 21). This difference could explain the discrepancy seen between the results obtained.

Table 21. Expression levels of ATRX and p53 are maintained in U87 WT and R132H cells.

Expression levels of ATRX and p53 in both RNASeq datasets. Levels are expressed by their "average expression", corresponding to the baseMean parameter for the one week dataset, representing the average of normalized counts for all samples, and to the reverse of the average log2 expression levels for the 48 hours dataset. Log2-fold change and P.adjusted values for each gene are also provided.

	Average Expression	Log2 fold-change	P.adjusted
48 hours			
ATRX	13.99	0.44	0.59
P53	36.75	-0.10	0.8
One week			
ATRX	15.69	-0.69	0.73
P53	45.34	-0.07	0.98

Nevertheless, despite a trend towards decreasing, our data cannot confirm any significant loss of CTCF binding in R132H cells at the coordinates reported, or any decrease in SOX2 expression, as described in Modrek et al.

8.11 Future directions

At the time of writing, a genome-wide DNA methylation sequencing using TAPS is on queue to be sequenced (see Chapter 6, note 6B). These results will be key to understand if and where the mechanism of loss of insulation proposed by Flavahan is happening in our cellular model. Indeed, the combination of DNA methylation data with CTCF binding will allow the definition of DNA methylation-sensitive CTCF binding sites. Gene expression data around loci showing loss of CTCF binding due to DNA hypermethylation will help in detecting candidate loci to test for loss of insulation between TADs in our cellular model, provided that any changes in expression are detected. To further corroborate the choice of a candidate locus, in a first instance

the distribution of higher order domains at loci of interest will be annotated using publicly available data in astrocytes, as that is the closest cell type to U87. Such in silico analysis will be used to make an initial selection of loci of interest. Ultimately, however, the definite proof of loss of insulation between TADs in U87 upon IDH1 mutation will derive from a chromosome conformation capture experiment, which will be used to validate the appropriate candidate loci.

8.12 Discussion

Analysis of CTCF binding around Flavahan et al. annotated boundaries has shown no overlap at the PTPRN or DNAJC12 boundaries, and the presence of a CTCF peak at the PDGFRA Insulator, even though its binding was not significantly affected by IDH1 mutation. Furthermore, gene expression results are somewhat in contrast with what has been published, as we detected no five-fold change increase in PDGFRA levels, but its increase at 48 hours, which was not confirmed by qPCR, and a decrease at one week in R132H control cells. While DNAJC12 levels were in line with results from the paper, PTPRN showed a significant downregulation at both timepoints tested. Taken together, results from CTCF ChIP and RNASeq do not support Flavahan's model, at least in our U87 WT and R132H cell lines.

While RNAseq suggested a potential increase of PDGFRA levels in R132H cells, this difference was not validated by qPCR results. This gene is considered to be frequently altered in glioma patients, but its basal expression in U87 cell line could be very low, as indicated by the high Ct values measured by qPCR, which would explain the high biological variability seen through Real Time PCR. Furthermore, CTCF ChIPSeq results pointed at a slight decrease of its binding on the Insulator target sequence, although the difference was not statistically significant and was not reproduced in the one week dataset. Thus, in these cells PDGFRA is not a good candidate to use as readout for confirming Flavahan's model. At the time of writing, DNA methylation results by TAPS are still ongoing as the experiment has been delayed because

of the Coronavirus pandemic, but might be key in understanding whether the model proposed by Flavahan can stand or not. The model suggests loss of CTCF binding upon methylation of CGs nucleotides in its target sequence, so it may be that in our cell line there is no great reduction of CTCF binding or a consistent increase in PDGFRA levels simply because the insulator is not methylated.

Investigation of other loci annotated in the Flavahan publication, such as DNAJC12 and PTPRN, did not provide clear results, as no CTCF peak was found around the annotated boundary. While genomic coordinates in the publication are based on hg19 annotation, our datasets have all been aligned to hg38, and coordinates conversion between the two genomes has been obtained using the LiftOver online tool. To exclude problems with LiftOver, a further check of the PDGFRA locus using a range of TADs annotations from the same publicly available datasets used in Flavahan et al., 2016 was added, only to find that the boundaries of such TADs coincide with the astrocytes TADs we used (Figure 8.9). If we are to exclude any problems with LiftOver function, as the coordinates given for the PDGFRA Insulator were indeed matching with our CTCF track, it may be that the boundary loss at DNAJC12 or PTPRN locus does not involve CTCF binding, at least in our experimental model.

When looking for the annotated boundaries described in the 2016 paper, another common discrepancy is that the alleged boundary coordinate does not fall at the end of a TAD. This could be due to the fact that the authors refers to domains without necessarily classifying them into TADs or sub-TADs. Indeed, further visualization of HiC maps from the same publicly available data used in Flavahan (2016) indicates that the Insulator upstream PDGFRA is delimiting a sub-TAD (Figure 8.10), and many of the gene pairs gaining or losing correlation are falling into separate sub-TADs, as opposed to a single TAD (eg. TET1, STOX1 and DDX21, Figure 8.13). Furthermore, many publications have shown that TADs tend to be stable across different cell types, an assumption accepted by the authors themselves, as they have used a plethora of high throughput conformation capture (Hi-C) datasets to help predict the boundaries

in their model (Barutcu et al., 2018; Dixon, Gorkin, & Ren, 2016; W A Flavahan et al., 2016). Yet, in at least two of the loci evaluated, PTPRN and SOX2, TADs distribution seemed to be cell type-dependent. Indeed, PTPRN locus spanned a single domain in two different sets of Astrocytes, in contrast with the two domains seen in a panel of other cell types (Figure 8.15). Moreover, the contact between SOX2 promoter and its downstream enhancer is strictly dependent on the need for SOX2 expression: while present in human embryonic stem cells, where its expression promotes differentiation, it is absent in IMR90 lung fibroblasts, which show low levels of SOX2 (Figure 8.17, (Modrek, Golub, Khan, Bready, Prado, Bowman, Deng, Zhang, Rocha, Raviram, Lazaris, Stafford, Leroy, et al., 2017). Thus, another possible explanation for the lack of CTCF binding at certain boundaries could be that U87 cells show at least a subset of cell-specific domains, differing from those annotated in Flavahan et al., 2016. To further complicate the picture, the authors are only considering TADs which have bound CTCF at their boundaries, but this corresponds to a specific subset of genomic domains, while the TADs annotations publicly available might be representing all TADs, or the biggest ones, as opposed to specific, CTCF associated TADs or sub-TADs.

Another possibility is that in the U87 cell line this mechanism is still happening, just not at the boundaries predicted by Flavahan. In their paper, RNASeq analyses have been carried out by using several publicly available datasets from glioma patient-derived samples, while our experimental model is a commercially available glioblastoma cell line. Since RNASeq results and correlations of expression levels between genes standing within the same TAD or in two neighbouring ones are the first step upon which they build their whole model, it may be that starting from a cell line with a different trascriptome leads to different results.

Another detail to consider is that patient-derived samples represent gliomas at different stages or subtypes, and they might present other mutations or genetic alterations aside from IDH1, which might further affect their transcriptome or epigenetic landscape. The same is true for the

cellular model in Modrek et al, showing a three hit composed of IDH1 mutation, loss of ATRX and p53. Perhaps multiple hits are required before this mechanism actually occurs, and our cell lines are not at such an advanced stage of cancer development.

Furthermore, the main assumption that disruption in higher order structures affects gene expression has recently been questioned. Findings analyzing gene expression following higher order chromatin disruption either in *trans*, through deletion of CTCF or cohesin, or in *cis,* through highly rearranged chromosomes in Drosophila, show that, while inducing multiple rearrangements in TADs organization, little effect is seen at the gene expression level (Ghavi-Helm et al., 2019; Nora et al., 2017; Rao et al., 2017). This suggests an uncoupling between higher order chromatin structures and gene expression, implying that other mechanisms might be at play in maintaining appropriate gene expression (Ghavi-Helm et al., 2019).

CTCF ChIP results around annotated boundaries for PDGFRA and DNAJC12 are also in contrast with a generalized CTCF loss in the R132H cells compared to their WT counterpart: around 2000 CTCF peaks were significantly decreased in IDH1 mutant cells (48-hour: 1630; one week: 2692, Chapter 4). This might suggest that CTCF binding on DNA is indeed methylation sensitive, but its decrease does not necessarily coincide with loss of insulation between TADs, or at least not at the coordinates suggested by the 2016 paper in our cell lines. DNA methylation sequencing results will be important in proving this hypothesis, as they will not only inform us of the DNA methylation levels around the Flavahan predicted boundaries, but will also help us in identify potential sites elsewhere in the genome of U87 cells where we do see an increase of DNA methylation coupled with loss of CTCF binding.

In both datasets, GSK864 treatment did little to rescue CTCF binding or gene expression levels. As the compound inhibits preferably the IDH1 mutant protein, its action is very specific and any epigenetic change would be only an indirect consequence of its activity. Furthermore, examples in literature have shown that IDH1 mutation greatly impacts methylation levels, but

due to the long turnover times of this modification across histones and DNA, restoring normal levels takes longer than expected, and a long-term exposure to this mutation can lead to virtually irreversible changes (Turcan et al., 2018). U87 R132H cells are a commercially available cell line, so reverting these changes might require an even longer treatment before seeing a robust impact on their epigenetic landscape.

Thus, while our results cannot confirm loss of insulation caused by impaired CTCF binding at the boundaries suggested by Flavahan et al, or Modrek et al., future analysis of DNA methylation levels might shed more light as to whether this mechanism is happening, but at different loci compared to those already described, or not at all in our cellular model.

.

9. Effect of IDH1 R132H mutation on levels of H3K9me3 on Chromosome X

9.1 Heterochromatin role in genome organization within the nucleus

Within the nucleus, DNA is compacted into chromatin, defined as the ensemble of DNA and the proteins associated with it. Chromatin is generally divided into two classes: euchromatin, linked with active gene expression and a looser nucleosome compaction to allow for the binding of the transcriptional machinery on the DNA, and heterochromatin, wherein nucleosomes are more condensed and thus generally associated with low levels, if at all, of gene expression (Allshire & Madhani, 2018). Heterochromatin further divides into constitutive and facultative. Constitutive heterochromatin corresponds to regions whose silencing is maintained regardless of phases of the cell cycle or the cell type, such as centromeric regions or highly repetitive sequences, and is preferably marked by H3K9me3. On the contrary, facultative heterochromatin is locus- and cell type-specific, and can be marked by both H3K9me3 and H3K27me3 (Hawkins et al., 2010; Zhu et al., 2013). Thus, heterochromatin plays a functional role in the control of gene expression, as its presence or absence at defined genomic loci helps maintaining stem cells pluripotency or silencing genes in a cell-type specific manner in differentiated cells (Hawkins et al., 2010; Zhu et al., 2013).

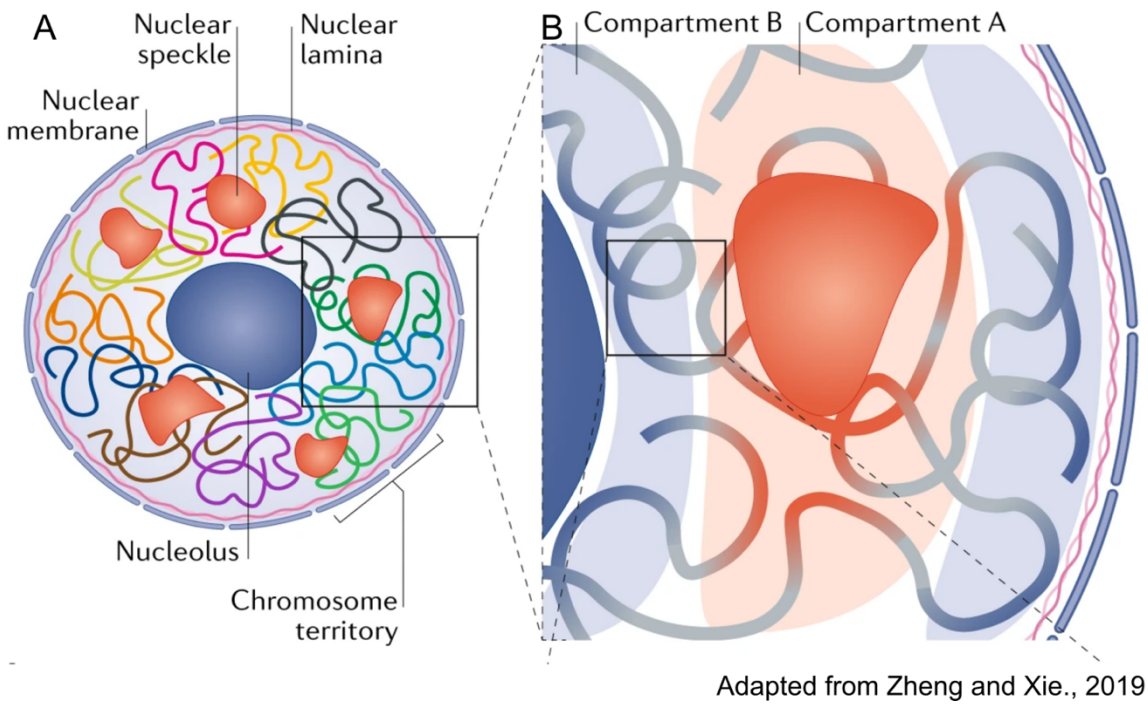


Figure 9.1. Schematic representation of nuclear and chromosome organization. *A) Diagram of an interphase nucleus. Nuclear membrane is lined with Nuclear lamina, composed of a network of intermediate filaments called lamins. The nuclear lamina is in direct contact with chromatin regions, organizing them into lamina-associated domains (LADs). B) Schematics of two main chromosome compartments, as detected in Hi-C mapping experiments. Compartment A (red) includes actively transcribed sequences in the DNA, more likely to interact with one another around nuclear speckles (here in dark red) and mainly found in the central core of the nucleus. Compartment B corresponds to transcriptionally silenced heterochromatin and, potentially due to its inactivity, is located to the nuclear periphery where it interacts with the nuclear lamina forming LADs. Image adapted from Zheng and Xie 2019* (Zheng & Xie, 2019).

Furthermore, microscopy and chromatin conformation capture studies have suggested a role for heterochromatin in establishing chromosome structure and nuclear architecture (Figure 9.1). The nuclear lamina lines the inner nuclear membrane and is composed of intermediate filament proteins named lamins (Stuurman, Heins, & Aebi, 1998). Genomic regions associated with it are termed nuclear lamina associated domains, or NADs (Pickersgill et al., 2006). Several studies have suggested that such regions might correspond to heterochromatic DNA, as they are marked by H3K9me3, have a low gene density and are often transcriptionally inactive, potentially explaining their localization at the nuclear periphery (Reddy, Zullo, Bertolino, & Singh, 2008; B. Wen et al., 2009).

A second layer of organization provided by heterochromatin is the distinction between chromosomal compartments. Hi-C experiments have highlighted the presence of two distinct compartments, A and B (Lieberman-Aiden, Miiro, et al., 2009). While the former corresponds to active regions in the genome, often interacting with one another and found within the central nuclear space or around nuclear speckles, the latter is made up of silenced DNA, overlaps with heterochromatin and is localized in proximity to the nuclear lamina (Bickmore & Van Steensel, 2013; S. Wang et al., 2016).

Results from H3K9me3 ChIP-Seq suggested its localised loss around selected chromosomes upon IDH1 mutation (Chapter 6, Figure 6.3). Given that this histone modification is frequently found in heterochromatic regions, and the role heterochromatin plays in structural organization within the nucleus, in this Chapter we sought out to validate this result in vitro; to investigate whether this loss corresponds to heterochromatic regions and to assess its impact on nuclear architecture or chromosome organization.

9.2 Loss of H3K9me3 on chromosomes 18 and X

Results from H3K9me3 ChIP-Sequencing revealed a general increase of this histone modification in cells bearing the IDH1 mutation (48 hours: 6739 peaks enriched in R132H U87; one week: 4906, Chapter 6). This is unsurprising, given that higher 2HG levels induced by IDH1 mutation have been shown to inhibit histone demethylases (Lu et al., 2012; Turcan et al., 2012).

Counterintuitively, we also noticed a consistent number of statistically significant H3K9me3 peaks decreased (48 hours: 2660 peaks decreased in R132H compared to WT U87; one week: 3306). Analysis of the number of H3K9me3 peaks increased or decreased for each chromosome indicates that, while the majority of chromosomes generally show a higher proportion of peaks increased, others are losing H3K9me3 peaks, including chromosomes 18 and X. At 48 hours, this difference is particularly dramatic on the X chromosome in the control IDH1 R132H cells,

which is losing around 500 H3K9me3 peaks and gaining less than 20 (Figure 7.2). Notably, a different number of peaks are found increased or decreased in R132H control compared to WT control cells between the 48 hours and one week datasets. This effect could be due to the fact that control samples are still being treated with 10 µl of DMSO, and perhaps after a week of culturing this is affecting their H3K9me3 distribution. Another difference between the 48 hours and the one week dataset is that in the latter, cells were split three times before chromatin harvest, to make up for their high proliferation rates. Thus, the different number of H3K9me3 peaks detected might also reflect the dynamics of this histone modification over multiple rounds of cell division.

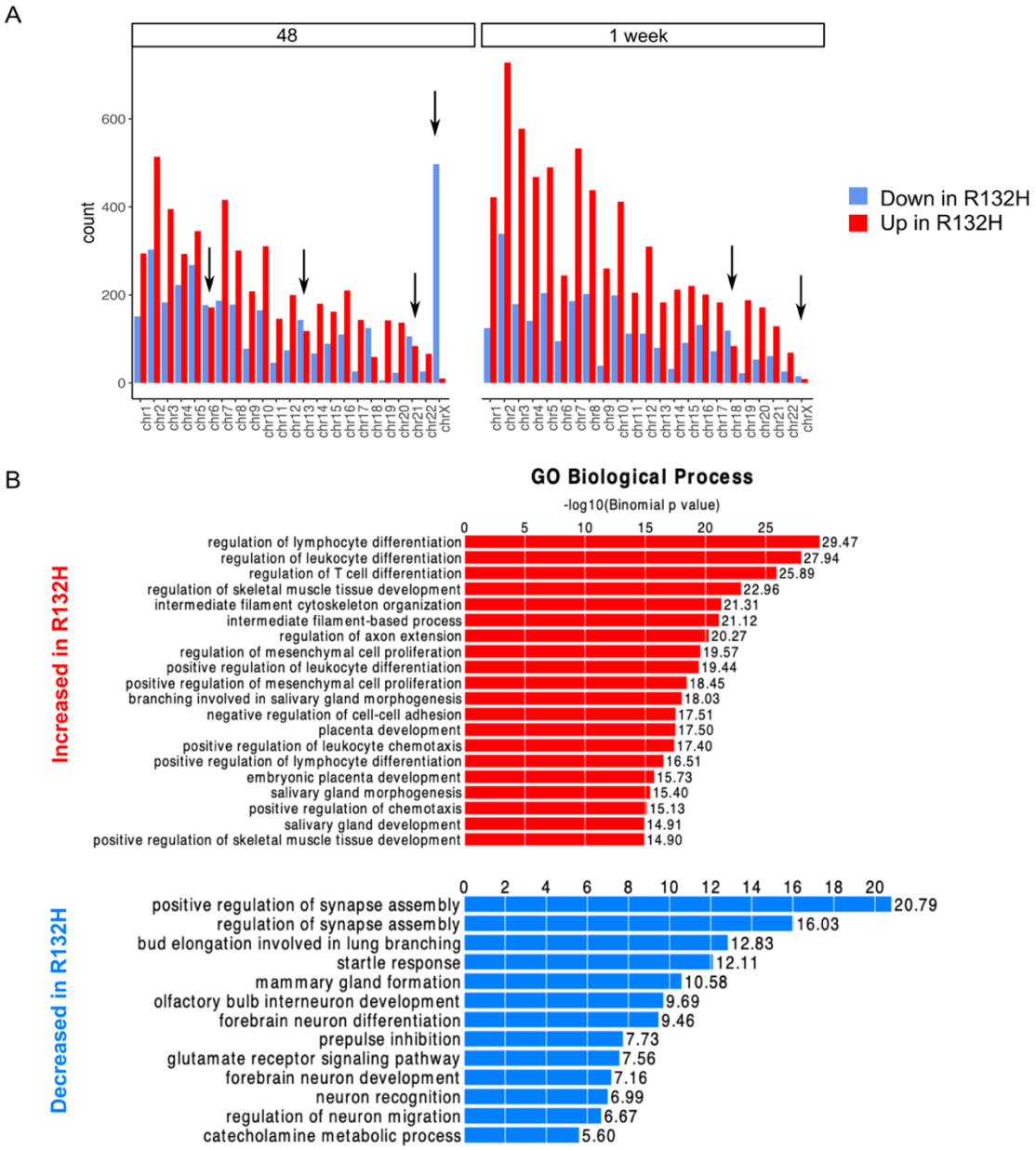


Figure 9.2 Some chromosomes show loss of H3K9me3 in R132H cells. *A) Barplot showing counts of H3K9me3 peaks increased (red) or decreased (blue) in R132H cells. Broad peaks were called using MACS2 callPeak, then further processed as detailed in the Methods section (Gaspar, 2018). Black arrows point at bars corresponding to chromosome X. B) Histogram of -log10 p.value of gene ontology terms enriched in the genomic regions corresponding to increased (red) or decreased (blue) H3K9me3 peaks in R132H cells. Histograms and Gene Ontology enrichment analysis on differentially enriched loci were produced with the GREAT online tool (http://great.stanford.edu/public/html/index.php).*

We next searched for potential functions linked to the genomic regions associated with H3K9me3 peaks increased (red) or decreased (blue) in R132H using the online tool GREAT ((Mclean et al., 2010), Figure 9.2B). Regions corresponding to an increase in K9me3 are associated with gene ontology categories related to differentiation and development. Indeed,

H3K9me3-mediated maintenance of heterochromatin ensures the activation of the appropriate genes during development and cell fate determination (J. Chen et al., 2013; Matoba et al., 2014). These results are in also line with findings in literature, as IDH1 mutations have been linked with a block in differentiation as a consequence of the hypermethylator phenotype they promote (Figueroa et al., 2010; Modrek, Golub, Khan, Bready, Prado, Bowman, Deng, Zhang, Rocha, Raviram, Lazaris, Stafford, LeRoy, et al., 2017; Okoye-Okafor et al., 2015). Gene ontologies associated to H3K9me3 peaks lost also relate to development, but seem to be more focused on neuronal function, as opposed to the haematopoietic or immune system.

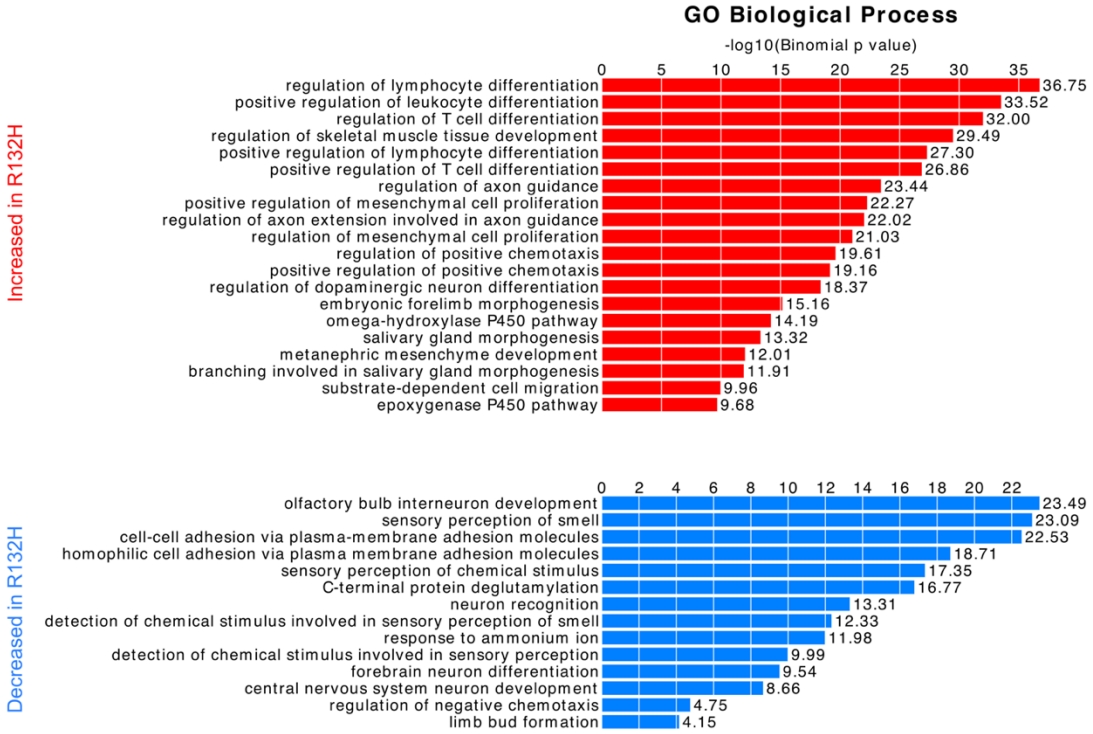


Figure 9.3 GO enrichment analysis on H3K9me3 differential peaks in the one week dataset. *Bar plot of the log10 p.value of gene ontology terms enriched in the genomic regions corresponding to increase(red) or decreased (blue) H3K9me3 peaks in R132H control cells at one week. Histograms and Gene Ontology enrichment analysis on differentially enriched loci were produced as described in Figure 9.2*

Very similar results were found upon GO analysis of the one week H3K9me3 dataset: a high number of significant terms related to differentiation of hematopoietic and muscle cell lineages

linked to regions with increased H3K9me3 in R132H cells, and ontologies regarding neuronal function or differentiation associated with regions with decreased K9 trimethylation (Figure 9.3). Restricting the search specifically to the H3K9me3 peaks decreased on chromosome X resulted in no significant GO terms.

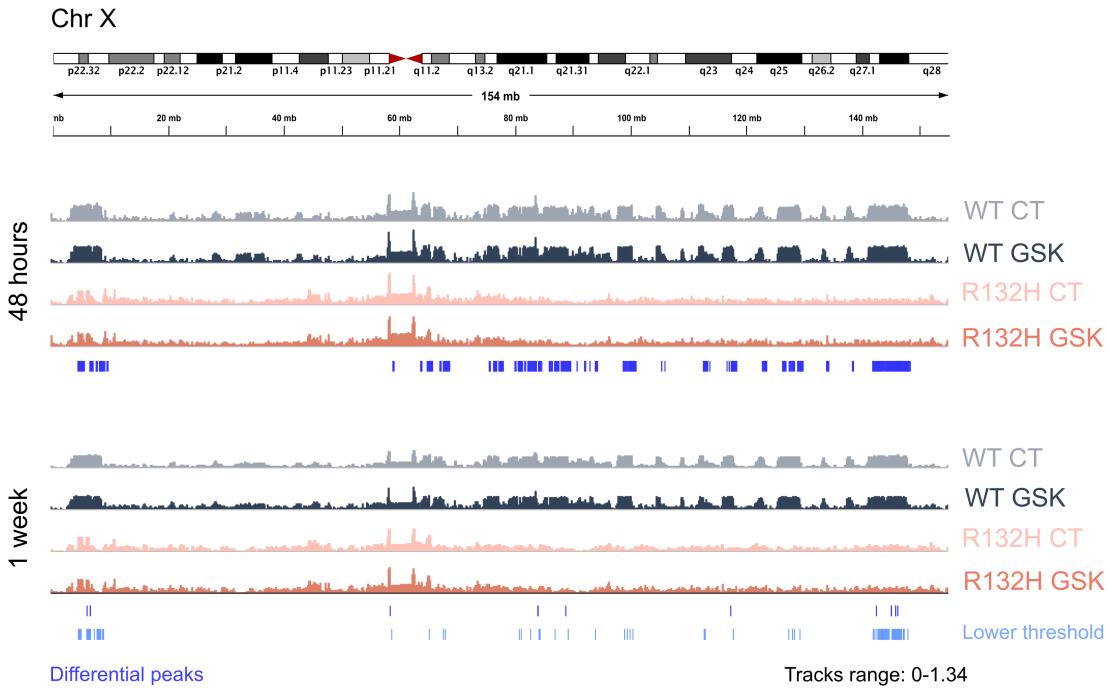


Figure 9.4 H3K9me3 distribution on Chromosome X in WT or R132H IDH1 U87 cells*. IGV tracks representing chromosome X ideogram, WT Ctrl (grey) or R132H Ctrl (light red) H3K9me3 tracks at 48 hours (top) or one week (bottom). A blue track underneath the samples highlights statistically significant differential peaks, as obtained from MACS2 getDifferentialPeaks (see Materials and Methods). All H3K9me3 tracks have been grouped and autoscaled to the range of values 0 – 1.34 on IGV.*

Analysis of chromosome X on a genome browser (IGV) highlighted another characteristic of these peaks. Rather than being broad peaks of localized H3K9me3, IDH1 mutated cells are losing entire blocks of K9me3 (Figure 9.4), potentially corresponding to heterochromatin. This effect is present in both datasets, although more statistically significant peaks are detected in the 48 hours one, perhaps due to the restrictive filtering parameters used in ChIPSeq analysis. Indeed, only peaks with a total tag counts greater than 32 in each sample have been selected for differential analysis, as per Turcan et al., 2018. As expected, lowering this threshold to include peaks with at least an average tag count of 32 across samples finds more differentially decreased

peaks in the one week dataset (Figure 9.4, light blue bars), so it would seem that the lower threshold analysis reflects better the effect seen in IGV. However, selecting peaks with a minimum count of 32 in each sample would ensure more reliable results and exclude loci with excessively low read counts. As this parameter has been maintained throughout all the ChIPSeq analyses, across both timepoints, we ultimately decided to prioritize comparability amongst different datasets. Notably, treatment with GSK864 was ineffective in rescuing this phenotype, perhaps because of the turnover times required for histone demethylation to be restored (Turcan et al., 2018).

While chromosome X is the one most affected by H3K9me3 loss in the IDH1 mutant, other autosomal chromosomes, such as chromosomes 6, 18 and 21, show a similar effect in more localized areas, specifically around heterochromatic bands or telomeric regions (Figure 9.5).

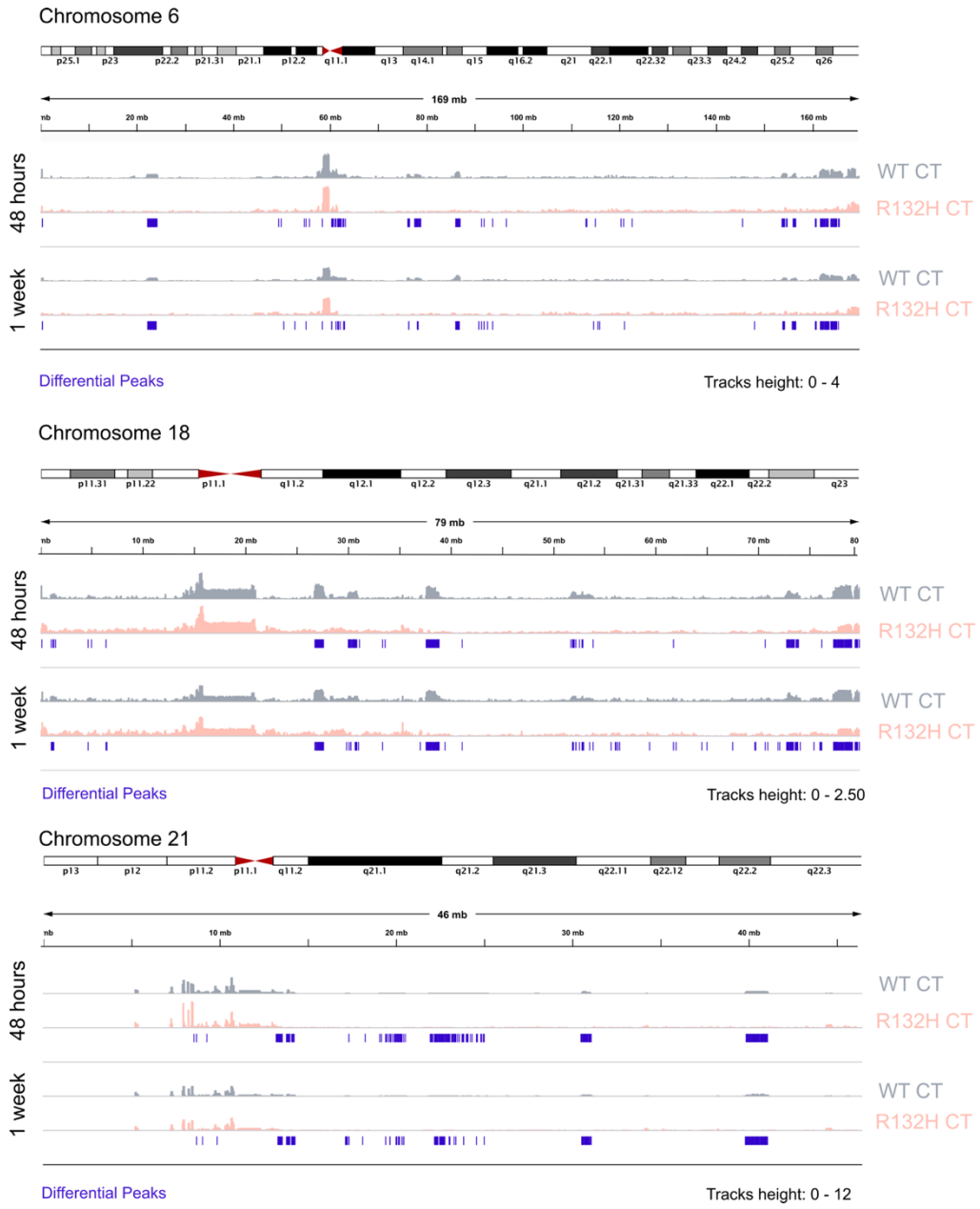


Figure 9.5 Distribution of H3K9me3 peaks lost across three autosomal chromosomes*. IGV tracks showing H3K9me3 distribution in WT control (grey) or R132H control (pink) in chromosomes 6 (top), 18 (middle) or 21 (bottom). Differential peaks are highlighted with blue bars underneath the appropriate tracks. Tracks from the two datasets have been grouped and scaled to the same height, indicated at the bottom. Bigwig files were normalized to 1x coverage and differentially enriched peaks were calculated as described in the Methods section.*

Thus, IDH1 mutant cells show a general increase of H3K9me3, but a peculiar loss of entire blocks of this histone modification, specifically along chromosome X.

9.3 Loss of H3K9me3 around Transposable Elements

Annotation of differential peaks with MACS2 suggested that the majority of decreased peaks across chromosome X are falling along transposable or repetitive elements (Figure 9.6). These sequences commonly coincide with heterochromatic regions as a way to protect the genome from their aberrant activation (J. He et al., 2019; Rowe et al., 2013). Indeed, mutations rising from the integration or transposition of these loci represent a threat to genome stability. We hypothesised that loss of H3K9me3 and potentially heterochromatin along these regions of chromosome X could cause an increased expression of transposable elements in IDH1 mutated cells.

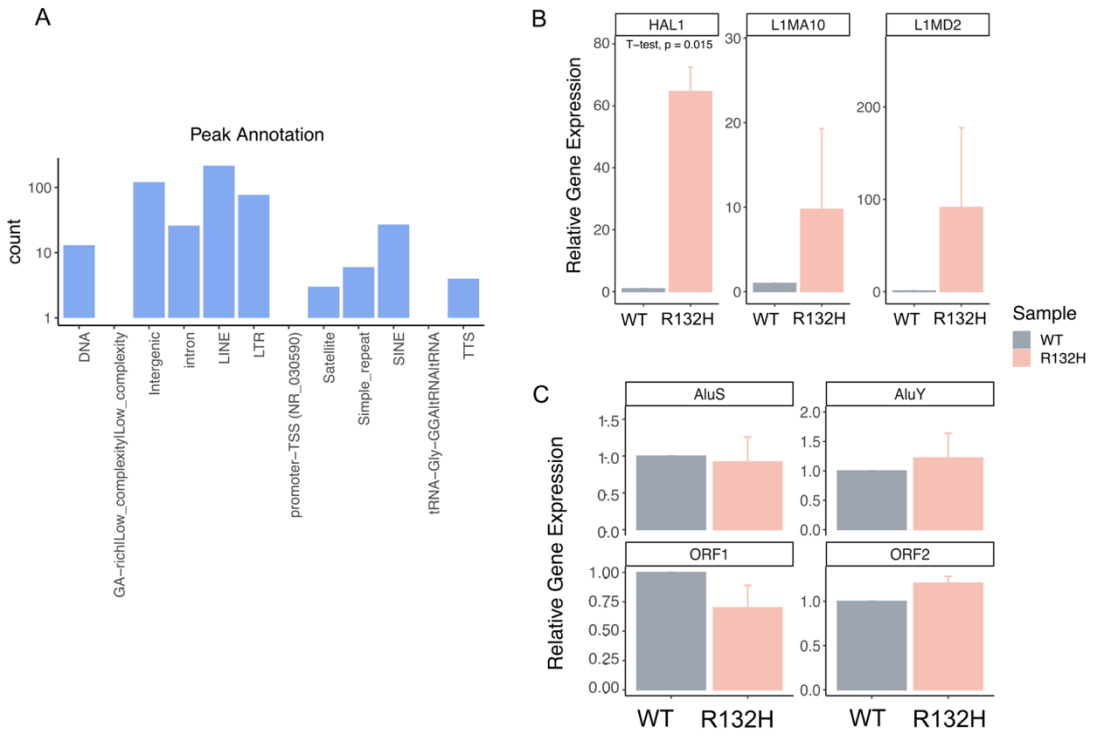


Figure 9.6 Annotation of decreased H3K9me3 peaks along chromosome X*. A) Barplot showing the number of peaks for each genomic annotation (x axis). The majority of peaks seems to fall within repetitive sequences, such as long interspersed nuclear elements (LINE), short interspersed nuclear elements (SINE), long terminal repeats (LTR), DNA Tigger transposons (DNA), simple DNA repeat (Simple repeat). Genomic annotation for each peak was obtained with MACS2 AnnotatePeak. B) Relative gene expression levels of transposable elements. Top: qPCR results obtained with primer pairs annealing to specific H3K9me3 peaks lost around LINEs in chromosome X. C) General levels of LINE1 (ORF1, ORF2) or SINEs (AluY, AluS). Relative expression of all genes were calculated according to the 2^{-ddCt} method, using the mean of two housekeeping genes (ACTB, RPLP0) as reference and the WT sample as calibrator. Statistical significance levels were assessed with a Student's T-test, and p-values*

of significant comparisons are shown on the plot. Data are expressed as mean ± SEM for N = 3 biological replicates.

As repetitive sequences are usually discarded in very early stages of RNA-Seq analysis, we could not extract any information regarding their expression from our datasets. We decided to address this question using quantitative Real-Time PCR (qPCR). The high degree of sequence similarity amongst repetitive elements of the same class (LINEs, SINEs, Alu or LTR/ERV) makes primer design around a specific region difficult. We thus used two different strategies: a set of generic primers to amplify general levels of LINE1, as the main representative of the LINE category, AluS or AluY, both SINEs. As a second strategy, we designed chromosome X-specific primers around three decreased peaks: HAL1 (peak 61632), L1MA10 (peak 25871), and L1MD2 (peak 27232), all L1 LINEs. Predicted non-specific target amplification for each primer pair was assessed by Primer-blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) and primers selected had significantly less non-specific targets compared to others tested. Results from qPCR experiments indicated similar relative levels of SINEs and LINE1 (Figure 9.6C) in R132H compared to WT U87 cells, in the three biological replicates tested (LINE1 ORF1 p.value = 0.25; LINE1 ORF2 p.value = 0.11; AluY p.value = 0.65; AluS p.value = 0.84). Absolute levels of all the SINEs and LINEs were reasonably high, with average Ct around 20 for LINE1 ORF1 and 2, and around 15 for AluS and AluY, indicating that in both cell types there is a high basal expression of all three (Table 22).

Table 22. Raw Ct values as detected in qPCR with the LINE1 ORF1, ORF2, AluY and AluS primer sets*. Raw Ct values for the primer sets LINE1 ORF1, LINE1 ORF2, AluY and AluS in each of the three biological replicates tested. In a qPCR experiment, the Ct (cycle threshold) value corresponds to the number of PCR cycles required for the fluorescence in a sample to exceed the background threshold. Thus, the more highly expressed the gene is, the lower the Ct value will be, as fewer rounds of PCR amplification will be required to have enough molecules to produce the fluorescence needed to exceed the threshold. Conversely, the lower the starting amount of a certain sample, the more rounds of PCR will be needed to produce sufficient fluorescence, the higher the Ct value will be.*

Sample	ORF1	ORF2	AluY	AluS
WT N1	20.52	26.06	15.59	15.55
WT N1	20.62	26.3	15.2	15.57

WT N1	20.84	26.4	15.11	15.42
R123H N1	21.62	26.85	15.27	16.03
R123H N1	21.59	27.21	15.27	15.37
R123H N1	21.8	27.27	15.22	15.97
WT N2	21.35	26.99	15.5	15.66
WT N2	21.54	27.15	15.45	16.22
WT N2	22.36	26.98	15.36	15.88
R132H N2	26.54	25.95	16.05	16.45
R132H N2	21.37	26.83	16.34	16.39
R132H N2	21.43	26.39	15.65	16.45
WT N3	20.66	26.29	14.76	14.39
WT N3	20.76	26.79	14.98	14.67
WT N3	20.69	26.08	14.89	14.87
R132H N3	21.17	26.56	14.88	15.29
R132H N3	20.87	25.97	14.92	15.89
R132H N3	21.47	26.23	n.a.	15.61

Primer pairs spanning areas where H3K9me3 is lost in R132H cells on chromosome X resulted in the robust amplification of HAL1 (p.value = 0.015, Figure 9.6B, central panel), but primers for L1MD2 and L1MA10 presented considerable variability amongst biological replicates and their results should thus be interpreted carefully. Indeed, examination of their raw Ct values shows that both HAL1 and L1MA10 have extremely low basal expression, with an average Ct between 35 (L1MA10) or 38 (HAL1) in the WT sample, likely meaning that such transcripts are not being expressed. Most samples were still detectable within the 45 cycles of qPCR programmed but show a high variability across biological samples. Raw Ct values for L1MD2

clearly suggest that this primer pair was not efficiently amplified, and thus should not be considered (Table 23).

Table 23. Raw Ct values for chromosome X-annealing primer sets*. Cycle threshold values for chromosome X specific transposable elements: L1MD2, HAL1 and L1MA10.*

Sample	L1MD2	HAL1	L1MA10
WT N1	--	38.93	33.86
WT N1	--	42.47	35.03
WT N1	40.96	34.32	36.84
R123H N1	40.1	32.95	40.94
R123H N1	--	33.92	--
R123H N1	36.95	32.81	33.95
WT N2	41.94	40.39	36.49
WT N2	--	--	38.32
WT N2	39.55	34.33	36.85
R132H N2	--	30.88	40.62
R132H N2	--	32.36	40.06
R132H N2	32.46	30.11	41.37
WT N3	--	40.17	--
WT N3	--	31.83	--
WT N3	--	32.58	41.83
R132H N3	39.97	--	37.78
R132H N3	41.76	29.38	37.09
R132H N3	38.29	29.29	36.56

Results show that, while the general levels of LINEs and SINEs are still comparable between R132H and WT cells, the low basal expression detected using the chromosome X-specific

primers is indication of the fact that this data is not robust enough to confirm our hypothesis of increased expression of repetitive elements on Chromosome X.

9.4 Investigating morphological changes in the nucleus and chromatin marks distribution of R132H cells using 3D-SIM

Results so far have highlighted a localized decrease of H3K9me3 at heterochromatic loci along a small subset of chromosomes in R132H cells. Investigation of the expression levels of transposable elements falling within these loci has hinted at their increase upon IDH1 mutation. Thus, to assess whether this decrease in H3K9me3 is indeed happening at heterochromatic loci, and whether this loss affects nuclear architecture of R132H cells, we next performed immunofluorescence of WT or R132H U87 cells. Antibody staining against H3K9me3, H3K27ac and H4K5ac were performed to assess differences in their abundance and nuclear localization. Images were processed using the 3D-Structure illumination microscopy (3D-SIM) technique, in collaboration with Dr. Lothar Schermelleh.

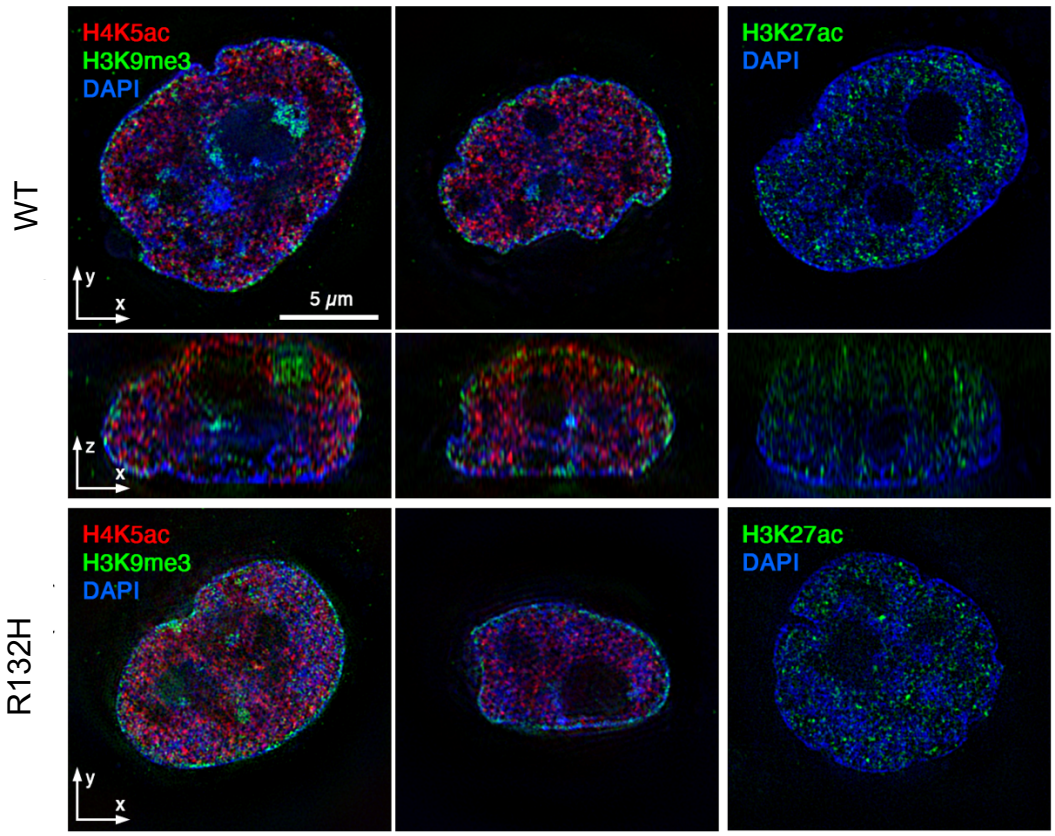


Figure 9.7 Optimization of H3K9me3, H3K27ac and H4K5ac immunofluorescence in WT or R132H U87 cells. *Representative images from a 3D-SIM optimization experiment show staining of the three*

chromatin marks of interest. WT U87 cells are shown on the first two rows, across their x and y axis (left to right, first row) or x and z axis (top to bottom, central row). R132H cells are displayed in the bottom row.

At the time of writing, a first optimization experiment has been analysed and two more biological repeats have been performed, but not analysed. Preliminary results show a successful staining of all antibodies tested, with H3K9me3 lining the outer perimeter of the nucleus and H4K5ac and H3K27ac being localized mostly in its centre (Figure 9.7). However, the low density of cells seeded caused difficulties in reaching the numbers required for statistical testing, thus a second repeat is needed in order to make a statement regarding global differences in chromatin morphology between the two cell lines.

9.5 Discussion

Results from H3K9me3 ChIP-Seq have highlighted a localised loss of this epigenetic mark around chromosome X in R132H cells, potentially representing selective heterochromatin loss. Given that such an effect is in contrast with the generalised increase of histone methylation detected in R132H cells and reported in literature, it should be extensively validated by both in silico and in vitro experiments. Future experiments comparing nuclear architecture between WT or R132H cells will reveal whether these changes are reflected in a different heterochromatin distribution upon IDH1 mutation. Indeed, 3D-SIM is a super-resolution microscopy technique allowing for visualization of general nuclear architecture and of chromatin marks of interest in single cells (Kraus et al., 2017). This data will complement the ChIP-Seq datasets, as it could confirm not only whether there is a generalized increase in H3K9me3 and a redistribution of acetylation, but it should also detect changes in these mark's distribution and whether they impact on nuclear architecture. H3K9me3 heterochromatin greatly overlaps with nuclear lamina-associated domains (NADs), genomic regions in physical contact with the nuclear lamina, found in the nuclear periphery (Reddy et al., 2008; B. Wen et al., 2009). Thus, should any of these H3K9me3 loss-affected loci fall within NADs, 3D-SIM

analysis would show a loss of H3K9me3 signal around the nuclear perimeter, suggesting an altered nuclear architecture upon IDH1 mutation.

Notably, given that the greatest H3K9me3 loss occurs on chromosome X and that our cell line is derived from a male subject, only one copy of this chromosome is present in the cells, which would make visualization of chromosome X-specific loci difficult.

Another way to confirm the existence of such H3K9me3-deprived loci would be to analyse more H3K9me3 publicly available datasets in IDH1 mutated cells, or patient-derived IDH1 mutated samples from the TCGA database, to see if the effect is also seen in other experimental models. A first analysis could be done using the same pipeline from Turcan et al. 2018, but trying a second bioinformatics protocol and assessing whether the same effect is seen on chromosome X would further validate our results.

H3K9me3-marked heterochromatin is bound by heterochromatin protein 1 (HP1), so another useful experiment would be a ChIP-Seq against this protein to verify whether this loss is indeed occurring at heterochromatic loci (Bannister et al., 2001).

Even though further studies are needed to validate this selective decrease in vitro, results from H3K9me3 ChIP-Seq have informed a new line of research, aimed at investigating nuclear architecture and higher order DNA structural differences upon IDH1 mutation.

10. Discussion

10.1 Summary

This thesis presents two different projects linked by the common theme of investigating the nexus between chromatin and metabolism.

A first chapter presents results concerning the effect of a panel of weight loss drugs on obese rats, with particular focus on changes to their gut microbiome. Somewhat surprisingly, common pharmacological treatments for obesity, such as Bupropion and Naltrexone, did not induce consistent weight loss. Instead, rats belonging to these groups appear to have gained, rather than lost, weight, while Sibutramine successfully reduced the weight of the rats by decreasing their appetite (Chapter 2). Furthermore, analysis of the rats' gut microbiome through shotgun metagenomics highlighted a potentially beneficial impact of Sibutramine treatment, as their microbiome composition by the end of the treatment is significantly different both compared to the beginning of the treatment and to control rats. Differential analysis revealed a significant increase in the abundance of species belonging to the Bacteroidetes phylum, and a concomitant shift of the Bacteroidetes to Firmicutes ratio towards the former, while in obese patients the ratio favours Firmicutes. Functional analysis showed a significant decrease in the abundance of genes coding for proteins of the flagellum, a bacterial component that has been linked to inflammation, further suggesting a beneficial impact of the treatment on the gut microbiome of obese rats. In contrast with previously published studies, Sibutramine treatment did not significantly alter any metabolic pathway, and Naltrexone treatment did not significantly impact on the rats microbiome, at least in our experimental setup.

The second project focused on the impact of IDH1 R132H mutation on the epigenetic and metabolic landscape of the glioblastoma cell line U87.

Metabolic characterization of U87 cells, either WT or bearing the IDH1 R132H mutation, and further treated or untreated with the IDH1 inhibitor GSK864, revealed broad metabolic alterations as a result of the IDH1 R132H mutation. Indeed, R132H cells show a staggering increase in 2HG levels, as well as several TCA cycle intermediates, such as AcetylCoA, Fumarate and Citrate. While treatment with GSK864 successfully rescued the level of 2HG after a 48 hours treatment, it had no impact on the levels of the other altered metabolites (Chapter 4).

An epigenetic characterization of the cellular model through RNASeq and genome-wide mapping of the insulator protein CTCF, H3K9me3 and H3K27ac further highlighted a series of alterations to gene expression, as well as the chromatin landscape upon IDH1 mutation. The effect of the IDH1 inhibitor on these epigenetic defects was tested at two different timepoints: 48 hours and one week. Indeed, in both timepoints R132H cells showed changes in transcriptome as a result of the mutation. Interestingly, while treatment with GSK864 did not induce many changes after 48 hours, it did alter gene expression upon a week-long treatment. Notably, while successfully correcting the metabolic defect by decreasing the levels of 2HG after 48 hours, and inducing gene expression changes promoting protein synthesis and AcylCoA metabolism after a week, treatment with GSK864 had little effect on the epigenetic alterations induced by IDH1 mutation, at both timepoints tested. H3K9me3 was generally increased, even though chromosome X and other chromosomes to a lesser extent, showed a characteristic localised loss of this histone modification in R132H cells, generating a new line of research within the project (Chapter 6 and 9). The increase in AcetylCoA levels instructed our mapping of H3K27ac, revealing a rearrangement of this chromatin mark after 48 hours, and its generalized increase in the one week dataset (Chapter 7). Finally, CTCF binding was decreased genome-wide in R132H cells but did not change as levels of D-2-HG decreased (Chapter 6).

A third line of research aimed at validating the mechanism suggested in Flavahan et al., who hypothesised that loss of CTCF binding due to increased DNA methylation would cause loss of insulation between otherwise separated topologically associated domains upon IDH1 mutation (Chapter 8). By merging RNASeq results with a genome-wide CTCF mapping and a publicly available TAD annotation, no clear indication of loss of insulation could be seen at the loci indicated by Flavahan. Results from a genome-wide DNA methylation sequencing experiment are pending and were severely delayed due to the Covid19 pandemic, but will be key in revealing whether CTCF binding is lost because of DNA hypermethylation and whether the mechanism proposed by Flavahan et al., 2016 is indeed happening, but in different loci in U87 WT or R132H cells compared to the ones described in their paper, or not at all.

10.2 Future directions

A wide range of chromatin-modifying enzymes use metabolic cofactors or substrates to modify the epigenome, with different outcomes on gene expression. This creates a link between metabolism and epigenetics, but whether it is enough to modify the former in order to induce changes in the latter is still an open question.

Results from the metabolic and epigenetic characterization of U87 WT or R132H cells suggest that, while the metabolic defect was successfully rescued upon treatment with an IDH1 inhibitor, it was not enough to revert the epigenetic alterations, especially those related to DNA or histone methylation (Chapters 4 to 7). Indeed, these modifications are notoriously persistent, given that they constitute the epigenetic memory and their patterns can be transmitted to other generations (Bird, 2002; C. Wang et al., 2018; Zheng et al., 2016). Perhaps, in the context of IDH1 mutation, the metabolic defect is indeed driving the epigenetic changes, but once this new epigenetic memory is established, it is much harder to revert it back to its original state. This introduces a temporal variable in the relationship between chromatin and epigenetics, taking into account the time needed to re-establish the original epigenetic state. In the case of IDH1 mutation, this implies further experiments aimed at defining how long it takes to revert the inhibition on DNA

and histones demethylases: whether these inhibited proteins need to be discarded, re-transcribed and translated, or if there is a threshold 2HG or α-ketoglutarate concentration after which they are active again. It is known that different lysine demethylases have different affinities for 2HG, so it is likely that the time needed to rescue their activity will differ from protein to protein, but the overall time needed to overcome the excessive DNA and histone methylation upon IDH1 mutation is still unknown (Black et al., 2012).

Interestingly, the epigenomic alterations seemed to be somewhat uncoupled from gene expression changes, as treatment with GSK864 induced changes in the transcriptome, without impacting the epigenome. Furthermore, analysis of H3K27ac changes around promoters and enhancers and of expression levels of their target genes indicated that an increase in this histone mark is not always linked to an increase in gene expression, in contrast with literature findings (Berger, 2007; Kouzarides, 2007; Shogren-Knaak et al., 2006).

Work in this thesis focused on post-translational modifications on histones, but did not address whether the generalized increase in methylation or acetylation is affecting other types of proteins in the cell. Indeed, both modifications can impact on the activity of several transcription factors (TFs), with effects ranging from increased stability, leading to activation of transcriptional programmes, stimulation of DNA binding activity, and inhibition, with subsequent decrease in target genes transcription (W. Gu & Roeder, 1997; Levy et al., 2011; Patel et al., 2004; Xie et al., 2011). As post-translational modifications on TFs are transient in nature, it may be that, if these proteins are affected by the increase in methylation or acetylation, this effect would be observed at a much smaller time scale compared to the 48 hours or one week timepoints. Thus, a proteomics screening for methylated and acetylated proteins in R132H cells could provide important information to further explore the relationship between metabolic defects, epigenetic alterations and gene expression in the context of IDH1 mutation.

Linking metabolism to higher order DNA structures rearrangements, Flavahan et al., 2016 suggested loss of insulation between neighbouring TADs as a further consequence of IDH1 mutation, due to DNA methylation forbidding CTCF binding on its target sequences. Data collected so far did not support this hypothesis, at least in the U87 R132H cellular model or at the two timepoints tested (Chapter 8). Genome-wide DNA methylation sequencing will reveal whether the underlying cause for the decrease in CTCF binding is indeed the DNA methylation of its target sequences. This could better inform the selection of candidate loci to test in a chromosome conformation capture experiment, to get definite proof of the loss of insulation between TADs indirectly induced by IDH1 mutation, or lack thereof. Furthermore, mapping of H3K9me3 in R132H cells revealed an unexpected decrease of this histone modification in restricted areas of the genome, more strikingly along chromosome X, potentially suggesting some rearrangement of heterochromatin upon IDH1 mutation (Chapter 9). Visualization of H3K9me3 nuclear distribution by 3D-SIM will help define whether there is a different pattern in R132H cells.

A second project aimed and investigating the nexus between chromatin and metabolism on a more systemic level, by analysing the impact of a panel of weight-loss drugs on the gut microbiome of obese rats (Chapter 2). While consistent weight loss, concomitant to a change in microbiome composition, was observed only after treatment with Sibutramine, further investigation of the ways in which the transformed microbiota is affecting the host metabolism is still needed. To this end, a metabolomic analysis of the blood of Sibutramine-treated rats would shed light on the kind of metabolites that the microbiota is producing, leading to a better understanding of the connection between the gut microbiota and its host metabolism, in the context of obesity.

10.3 Concluding remarks

Recent research has highlighted a mutual influence of epigenetics and metabolism, both at the cellular level, as many epigenetic writers and erasers use, or are inhibited by, metabolites as substrate or co-factors, and at a systemic one, through gut microbiota-derived metabolites used by the host. Work presented here investigates this relationships at both levels, in two separate projects. Shotgun metagenomics analysis of the gut microbiome of obese rats treated with a panel of weight-loss drugs showed a concomitant change in microbiome composition and function upon treatment with Sibutramine, laying the foundation for experiments aimed at elucidating the contribution of the transformed microbiome on the host metabolism.

In the context of IDH1 mutation, results highlight a temporal factor that needs to be taken into account, as correcting the metabolic defect was not sufficient to rescue the epigenetic alterations. Furthermore, a longer treatment with an IDH1 inhibitor induced changes in the transcriptome, but not in the altered epigenome, suggesting that, in this cellular model, gene expression changes are concomitant to, but not immediately instructed by, the epigenetic alterations induced by IDH1 mutation. This adds a layer of complexity to the connection between chromatin and metabolism, which future research will need to focus on, in order to design better therapeutic strategies for the many diseases in which alterations in the latter induce deleterious defects in the former.

11. Appendix

1. List of Tables

Table 1. Positioning of best characterized histone modifications along genes and relative correlation with the transcriptional state. 16
Table 2. Current pharmacotherapies for obesity. 31
Table 3. Treatment groups of the in vivo experiment. 36
Table 4. Average weight before (Day -3) or at the end of the experiment (Day 42). 42
Table 5. ANOVA results for combined food intake consumption. 48
Table 6. Primer design. 72
Table 7 Antibodies used in ChIP Seq. 74
Table 8 Primers for quality check of ChIPSeq. 75
Table 9. Metabolites monitored in GC/MS experiment. 81
Table 10. Transitions of unlabelled metabolites monitored in LC/MS/MS. 82
Table 11. Schematic of Immunofluorescence staining for WT and R132H IDH1 U87 cells. ..84
Table 12. List of antibodies used in immunofluorescence sample preparation for 3D-SIM 85
Table 13 Current targeted therapies for IDH1/2-mutant tumours 113
Table 14 Number of differentially expressed genes after 48-hours treatment with GSK864. 116
Table 15. Number of differentially expressed genes after 1 week of GSK864 treatment. 125
Table 16. Summary of main histone modifications targeted by α-Ketogluratate dependent demethylases 144
Table 17. Gene expression and significance levels of genes around the F11R locus. 200
Table 18. Gene expression and significance levels of genes from the GSTM locus. 202
Table 19. Raw Ct values of PDGFRA and DNAJC12 expression as detected in qPCR. 204
Table 20. Enrichment levels of CTCF peaks around the PDGFRA Insulator. 215
Table 21. Expression levels of ATRX and p53 are maintained in U87 WT and R132H cells. 230
Table 22. Raw Ct values as detected in qPCR with the LINE1 ORF1, ORF2, AluY and AluS primer sets. 246
Table 23. Raw Ct values for chromosome X-annealing primer sets. 248
Table 24. Primers and Barcodes used for library preparation and sequencing. 260
Table 25. Percentage of unassigned reads per sample for representative genes fliC, flgE and motA 260
Table 26. Labelled Acetyl-CoA transitions monitored. 261
Table 27. Labelled SAM transitions monitored. 262
Table 28. Parameters of LC/MS labelled run 262

2. *Supplementary image 11.8*

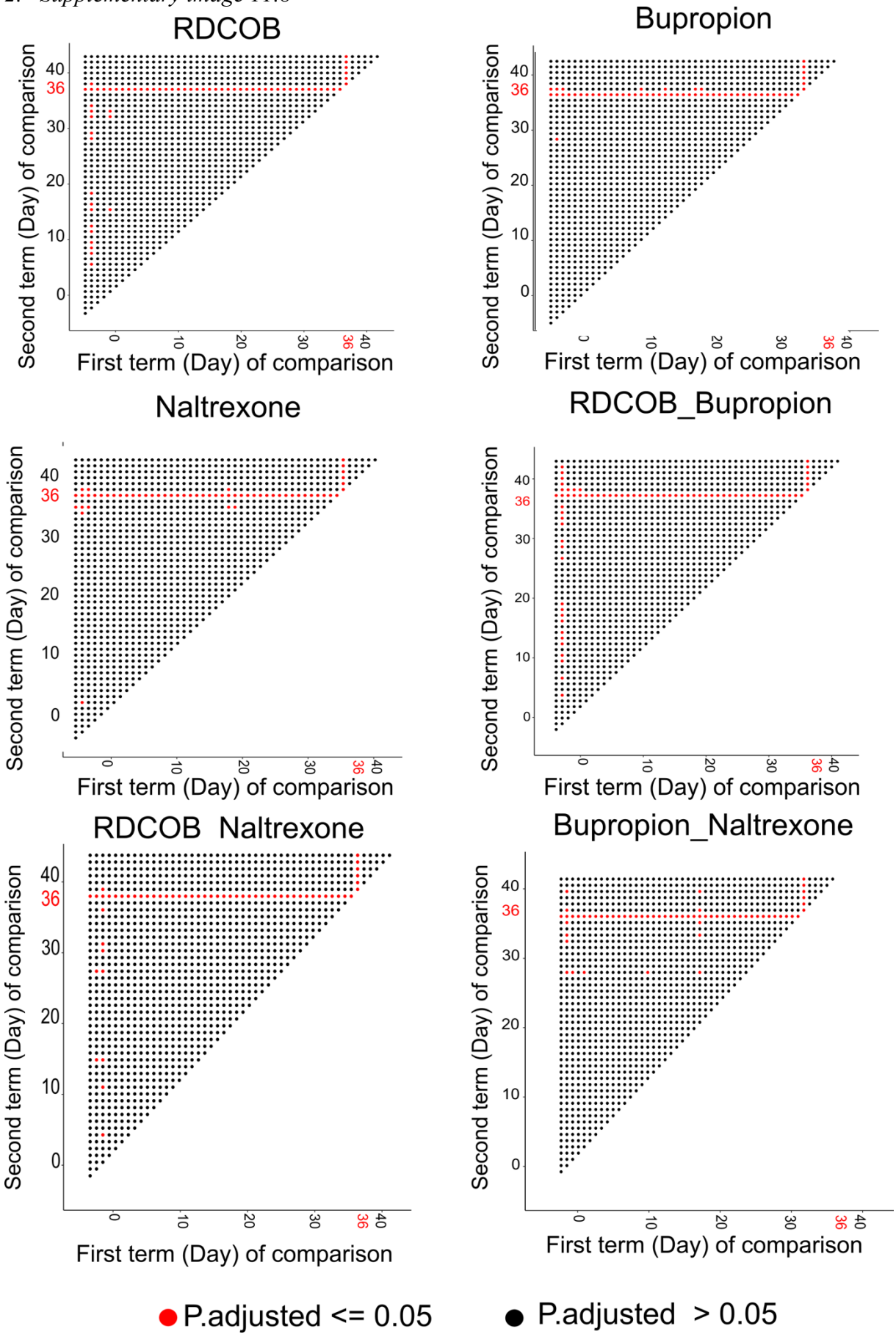


Figure 11.1. Tukey's post-hoc test following ANOVA statistical test on Timepoint reveals significance mostly due to fasting on Day 36. *Visualization of the p.values (P.adj <= 0.05, red; P.adj > 0.05, black) in the Tukey's post hoc test following ANOVA statistical test on timepoint. Food intake on a given day (first term of comparison, x axis) was compared to another (y axis) in the Tukey's post-hoc test. The statistical test performed gives a significant P.value, indicating which days are contributing to the significance of the Timepoint comparison. A closer look at the plots for most of the treatment groups, except Sibutramine, reveals that Timepoint is coming up as a significant term in ANOVA test due to the fasting occurred on Day 36, in preparation of the OGTT test.*

Primer	Sequence
Ion Torrent Adapter P1	ATCACCGACTGCCCATAGAGAGGCTGAGAC
Barcode 1	CTAAGGTAAC
Barcode 2	TAAGGAGAAC
Barcode 3	AAGAGGATTC
Barcode 4	TACCAAGATC
Barcode 5	CAGAAGGAAC
Barcode 6	CTGCAAGTTC
Barcode 7	TTCGTGATTC
Barcode 8	TTCCGATAAC
Barcode 9	TGAGCGGAAC
Barcode 10	CTGACCGAAC
Barcode 11	TCCTCGAATC
Barcode 12	TAGGTGGTTC
Barcode 13	TCTAACGGAC
Barcode 14	TTGGAGTGTC
Barcode 15	TCTAGAGGTC
Barcode 16	TCTGGATGAC
Fragment Library Adapter A	TTCCATCTCATCCCTGCGTGTCTCCGACTCAG
Fragment Library Adapter A Rev	AAGGTAGAGTAGGGACGCACAGAGGCTGAGTC

Table 24. Primers and Barcodes used for library preparation and sequencing.

Table 25. Percentage of unassigned reads per sample for representative genes fliC, flgE and motA

	Unassigned reads (%)		
Sample	fliC	motA	flgE
4	53.6082	33.3333	NA

14	44.3439	100	0
25	24.9229	13.5647	3.44828
33	69.3111	41.6452	36.2031
40	45.866	100	47.5207
55	51.7922	67.4242	85.3261
62	56.2468	0	7.69231
75	30.7377	82.3529	52.1866
85	42.9543	33.6134	38.3754
4H	11.8442	4.08602	0
14H	100	NA	NA
25H	32.8909	3.41686	7.20461
33H	60.6393	51.0417	27.6786
40H	57.453	38.0952	32.8125
55H	80	NA	0
62H	6.18557	0	0
75H	30.4888	25.7062	10.8168
85H	63.6459	26.875	56.5972

Table 26. Labelled Acetyl-CoA transitions monitored.

To correctly measure levels of labelled Acetyl-CoA in the samples, all possible transitions containing labelled fragments from Acetyl-CoA precursor ions have been calculated, then monitored at the m/z described below.

Compound	Precursor Ion Mass (m/z)	Transition (m/z)
810 + A + B + C	819	819 → 138 819 → 435 819 → 305
810 + A	812	812 → 138 812 → 430 812 → 303
810 + B	815	815 → 136

		815 → 433 815 → 303
810 + C	812	812 → 136 812 → 428 812 → 305
810 + A + B	817	817 → 138 817 → 435 817 → 303
810 + A + C	814	814 → 138 814 → 430 814 → 305
810 + B + C	817	817 → 136 817 → 433 817 → 305

Table 27. Labelled SAM transitions monitored.

Detectable fragments from SAM were defined as described in Chapter 4. This table recapitulates the possible combination of different labelled fragments that have been monitored in the LC/MS experiment at the m/z described.

Compound	Precursor Ion Mass (m/z)	Transitions (m/z)
399 + A + B + C	407	407→ 102 407 → 306
399 + A	401	401 → 97 401 → 300
399 + B	404	404 → 102 404 → 303
399 + C	400	400 → 97 400 → 299
399 + A + B	406	406 → 102 406 → 305
399 + A + C	402	402 → 97 402 → 301
399 + C + B	405	405 → 102 405 → 304

Table 28. Parameters of LC/MS labelled run

Compound	Transition	Collision Energy	Polarity
Ac_CoA_3	819⟶305	36	Positive
Ac_CoA_1	819⟶138	60	Positive
Ac_CoA_14	817⟶435	25	Positive

Ac_CoA_20	817⟶433	25	Positive
Ac_CoA_21	817⟶305	36	Positive
Ac_CoA_15	817⟶303	36	Positive
Ac_CoA_13	817⟶138	60	Positive
Ac_CoA_19	817⟶136	60	Positive
Ac_CoA_8	815⟶433	25	Positive
Ac_CoA_9	815⟶303	36	Positive
Ac_CoA_7	815⟶136	60	Positive
Ac_CoA_17	814⟶430	25	Positive
Ac_CoA_18	814⟶305	36	Positive
Ac_CoA_16	814⟶138	60	Positive
Ac_CoA_5	812⟶430	25	Positive
Ac_CoA_11	812⟶428	25	Positive
Ac_CoA_12	812⟶305	36	Positive
Ac_CoA_6	812⟶303	36	Positive
Ac_CoA_4	812⟶138	60	Positive
Ac_CoA_10	812⟶136	60	Positive
Ac_CoA_3u	810⟶428	25	Positive
Ac_CoA_1u	810⟶303	36	Positive
Ac_CoA_2u	810⟶136	60	Positive
SAM_2	407⟶306	4	Positive
SAM_1	407⟶102	32	Positive
SAM_10	406⟶305	4	Positive
SAM_9	406⟶102	32	Positive
SAM_14	405⟶304	4	Positive
SAM_13	405⟶102	32	Positive
SAM_6	404⟶303	4	Positive
SAM_5	404⟶102	32	Positive
SAM_12	402⟶301	4	Positive
SAM_11	402⟶97	32	Positive
SAM_4	401⟶300	4	Positive
SAM_3	401⟶97	32	Positive
SAM_8	400⟶299	4	Positive
SAM_7	400⟶97	32	Positive
SAM_3u	399⟶298	4	Positive
SAM_1u	399⟶250	12	Positive
SAM_2u	399⟶97	32	Positive

Bibliography

Abeshouse, A., Ahn, J., Akbani, R., Ally, A., Amin, S., Andry, C. D., … Zmuda, E. (2015). The Molecular Taxonomy of Primary Prostate Cancer. *Cell*, *163*(4), 1011–1025.

https://doi.org/10.1016/J.CELL.2015.10.025

Adam, J., Yang, M., Soga, T., & Pollard, P. J. (2014). Rare insights into cancer biology. *Oncogene*, *33*(20), 2547–2556. https://doi.org/10.1038/onc.2013.222

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., … Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, *46*(W1), W537–W544. https://doi.org/10.1093/nar/gky379

Allshire, R. C., & Madhani, H. D. (2018, April 1). Ten principles of heterochromatin formation and function. *Nature Reviews Molecular Cell Biology*. Nature Publishing Group. https://doi.org/10.1038/nrm.2017.119

Amary, M. F., Bacsi, K., Maggiani, F., Damato, S., Halai, D., Berisha, F., … Flanagan, A. M. (2011). IDH1 and IDH2 mutations are frequent events in central chondrosarcoma and central and periosteal chondromas but not in other mesenchymal tumours. *The Journal of Pathology*, *224*(3), 334–343. https://doi.org/10.1002/path.2913

Amatangelo, M. D., Quek, L., Shih, A., Stein, E. M., Roshal, M., David, M. D., … Levine, R. L. (2017). Enasidenib induces acute myeloid leukemia cell differentiation to promote clinical response. *Blood*, *130*(6), 732–741. https://doi.org/10.1182/blood-2017-04-779447

Amir, R. E., Van Den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., & Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature Genetics*, *23*(2), 185–188. https://doi.org/10.1038/13810

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638

Andrews, S. (n.d.). Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. Retrieved January 7, 2020, from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Ardlie, K. G., DeLuca, D. S., Segrè, A. V., Sullivan, T. J., Young, T. R., Gelfand, E. T., … Lockhart. (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*, *348*(6235), 648–660. https://doi.org/10.1126/science.1262110

Ariyannur, P. S., Moffett, J. R., Madhavarao, C. N., Arun, P., Vishnu, N., Jacobowitz, D. M., … Namboodiri, A. M. A. (2010). Nuclear-cytoplasmic localization of acetyl coenzyme a synthetase-1 in the rat brain. *Journal of Comparative Neurology*, *518*(15), 2952–2977. https://doi.org/10.1002/cne.22373

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., … Bork, P. (2011). Enterotypes of the human gut microbiome. *Nature*, *473*(7346), 174–180. https://doi.org/10.1038/nature09944

Bäckhed, F., Ding, H., Wang, T., Hooper, L. V., Gou, Y. K., Nagy, A., … Gordon, J. I. (2004). The gut microbiota as an environmental factor that regulates fat storage. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(44), 15718–15723. https://doi.org/10.1073/pnas.0407076101

Badur, M. G., Muthusamy, T., Parker, S. J., Ma, S., McBrayer, S. K., Cordes, T., … Metallo, C. M. (2018). Oncogenic R132 IDH1 Mutations Limit NADPH for De Novo Lipogenesis through (D)2-Hydroxyglutarate Production in Fibrosarcoma Cells. *Cell Reports*, *25*(4), 1018-1026.e4. https://doi.org/10.1016/J.CELREP.2018.09.074

Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C., & Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature*, *410*(6824), 120–124. https://doi.org/10.1038/35065138

Barker, C. S., Meshcheryakova, I. V., Kostyukova, A. S., & Samatey, F. A. (2010). FliO Regulation of FliP in the Formation of the Salmonella enterica Flagellum. *PLoS Genetics*, *6*(9), e1001143. https://doi.org/10.1371/journal.pgen.1001143

Barnes, D. E., Tomkinson, A. E., Lehmann, A. R., Webster, A. D. B., & Lindahl, T. (1992). Mutations in the DNA ligase I gene of an individual with immunodeficiencies and cellular hypersensitivity to DNA-damaging agents. *Cell*, *69*(3), 495–503. https://doi.org/10.1016/0092-8674(92)90450-Q

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., … Zhao, K. (2007). High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell*, *129*(4), 823–837. https://doi.org/10.1016/j.cell.2007.05.009

Bartlett, T. E., Zaikin, A., Olhede, S. C., West, J., Teschendorff, A. E., & Widschwendter, M. (2013). Corruption of the Intra-Gene DNA Methylation Architecture Is a Hallmark of Cancer. *PLoS ONE*, *8*(7). https://doi.org/10.1371/journal.pone.0068285

Barutcu, A. R., Maass, P. G., Lewandowski, J. P., Weiner, C. L., & Rinn, J. L. (2018). A TAD boundary is preserved upon deletion of the CTCF-rich Firre locus. *Nature Communications*, *9*(1), 1444. https://doi.org/10.1038/s41467-018-03614-0

Beagan, J. A., & Phillips-Cremins, J. E. (2020). On the existence and functionality of topologically associating domains. *Nature Genetics*, *52*(1), 8–16. https://doi.org/10.1038/s41588-019-0561-1

Bell, A. C., & Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*, *405*(6785), 482–485. https://doi.org/10.1038/35013100

Berger, S. L. (2007, May 24). The complex language of chromatin regulation during transcription. *Nature*. Nature Publishing Group. https://doi.org/10.1038/nature05915

Bergeron, J. R. (2016). Structural modeling of the flagellum MS ring protein FliF reveals similarities to the type III secretion system and sporulation complex. *PeerJ*, *2016*(2). https://doi.org/10.7717/peerj.1718

Bernstein, B. E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D. K., Huebert, D. J., … Lander, E. S. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*, *120*(2), 169–181. https://doi.org/10.1016/j.cell.2005.01.001

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., … Lander, E. S. (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell*, *125*(2), 315–326. https://doi.org/10.1016/j.cell.2006.02.041

Bernstein, E., Duncan, E. M., Masui, O., Gil, J., Heard, E., & David Allis, C. (2006). Mouse Polycomb Proteins Bind Differentially to Methylated Histone H3 and RNA and Are Enriched in Facultative Heterochromatin. *MOLECULAR AND CELLULAR BIOLOGY*, *26*(7), 2560–2569. https://doi.org/10.1128/MCB.26.7.2560-2569.2006

Bertram, K., Agafonov, D. E., Liu, W. T., Dybkov, O., Will, C. L., Hartmuth, K., … Lührmann, R. (2017). Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature*, *542*(7641), 318–323. https://doi.org/10.1038/nature21079

Bickmore, W. A., & Van Steensel, B. (2013). Genome Architecture: Domain Organization of Interphase Chromosomes. *Cell*, *152*, 1270–1284. https://doi.org/10.1016/j.cell.2013.02.001

Bird, A. (2002, January 1). DNA methylation patterns and epigenetic memory. *Genes and Development*. Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/gad.947102

Black, J. C., Van Rechem, C., & Whetstine, J. R. (2012). Histone Lysine Methylation Dynamics: Establishment, Regulation, and Biological Impact. *Molecular Cell*, *48*(4), 491–507. https://doi.org/10.1016/J.MOLCEL.2012.11.006

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bornfeldt, K. E., & Tabas, I. (2011, November 2). Insulin resistance, hyperglycemia, and atherosclerosis. *Cell Metabolism*. Elsevier. https://doi.org/10.1016/j.cmet.2011.07.015

Brat, D. J., Verhaak, R. G. W., Aldape, K. D., Yung, W. K. A., Salama, S. R., Cooper, L. A. D., … Zhang, J. (2015b). Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *New England Journal of Medicine*, *372*(26), 2481–2498. https://doi.org/10.1056/NEJMoa1402121

Brennan, C. W., Verhaak, R. G. W., Mckenna, A., Campos, B., Noushmehr, H., Salama, S. R., … Network, T. R. (2013). The Somatic Genomic Landscape of Glioblastoma. *Cell*, *30*(155), 462–462. https://doi.org/10.1016/j.cell.2013.09.034

Bulusu, V., Tumanov, S., Michalopoulou, E., van den Broek, N. J., MacKay, G., Nixon, C., … Kamphorst, J. J. (2017). Acetate Recapturing by Nuclear Acetyl-CoA Synthetase 2 Prevents Loss of Histone Acetylation during Oxygen and Serum Limitation. *Cell Reports*, *18*(3), 647–658. https://doi.org/10.1016/j.celrep.2016.12.055

Cai, L., Sutter, B. M., Li, B., & Tu, B. P. (2011). Acetyl-CoA Induces Cell Growth and Proliferation by Promoting the Acetylation of Histones at Growth Genes. *Molecular Cell*, *42*(4), 426–437. https://doi.org/10.1016/j.molcel.2011.05.004

Cairncross, G., Wang, M., Shaw, E., Jenkins, R., Brachman, D., Buckner, J., … Mehta, M. (2013). Phase III trial of chemoradiotherapy for anaplastic oligodendroglioma: Long-term results of RTOG 9402. *Journal of Clinical Oncology*, *31*(3), 337–343. https://doi.org/10.1200/JCO.2012.43.2674

Carlson M. (2019). GO.db: A set of annotation maps describing the entire Gene Ontology. R package version 3.8.2. https://doi.org/10.18129/B9.BIOC.GO.DB

Carvalho, F. A., Koren, O., Goodrich, J. K., Johansson, M. E. V., Nalbantoglu, I., Aitken, J. D., … Gewirtz, A. T. (2012). Transient inability to manage proteobacteria promotes chronic gut inflammation in TLR5-deficient mice. *Cell Host and Microbe*, *12*(2), 139–152. https://doi.org/10.1016/j.chom.2012.07.004

Castaner, O., Goday, A., Park, Y. M., Lee, S. H., Magkos, F., Shiow, S. A. T. E., & Schröder, H. (2018). The gut microbiome profile in obesity: A systematic review. *International Journal of Endocrinology*, *2018*. https://doi.org/10.1155/2018/4095789

Chan, A. C., Iwashima, M., Turck, C. W., & Weiss, A. (1992). ZAP-70: A 70 kd protein-tyrosine kinase that associates with the TCR ζ chain. *Cell*, *71*(4), 649–662. https://doi.org/10.1016/0092-8674(92)90598-7

Chen, J., Liu, H., Liu, J., Qi, J., Wei, B., Yang, J., … Pei, D. (2013). H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nature Genetics*, *45*(1), 34–42. https://doi.org/10.1038/ng.2491

Chen, X., Xie, R., Gu, P., Huang, M., Han, J., Dong, W., … Lin, T. (2019). Long noncoding RNA LBCs inhibits self-renewal and chemoresistance of bladder cancer stem cells through epigenetic silencing of SOX2. *Clinical Cancer Research*, *25*(4), 1389–1403. https://doi.org/10.1158/1078-0432.CCR-18-1656

Choi, J., Djebbar, S., Fournier, A., & Labrie, C. (2014). The co-chaperone DNAJC12 binds to Hsc70 and is upregulated by endoplasmic reticulum stress. *Cell Stress and Chaperones*, *19*(3), 439–446. https://doi.org/10.1007/s12192-013-0471-6

Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., … Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science (New York, N.Y.)*, *325*(5942), 834–840. https://doi.org/10.1126/science.1175371

Chowdhury, R., Yeoh, K. K., Tian, Y., Hillringhaus, L., Bagg, E. A., Rose, N. R., … Kawamura, A. (2011). The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Reports*, *12*(5), 463–469. https://doi.org/10.1038/embor.2011.43

Ciccone, D. N., Su, H., Hevi, S., Gay, F., Lei, H., Bajko, J., … Chen, T. (2009). KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. *Nature*,

461(7262), 415–418. https://doi.org/10.1038/nature08315

Compton, S. A., Tolun, G., Kamath-Loeb, A. S., Loeb, L. A., & Griffith, J. D. (2008). The Werner syndrome protein binds replication fork and Holliday junction DNAs as an oligomer. *Journal of Biological Chemistry*, *283*(36), 24478–24483. https://doi.org/10.1074/jbc.M803370200

Cremona, C. A., Sarangi, P., Yang, Y., Hang, L. E., Rahman, S., & Zhao, X. (2012). Extensive DNA Damage-Induced Sumoylation Contributes to Replication and Repair and Acts in Addition to the Mec1 Checkpoint. *Molecular Cell*, *45*(3), 422–432. https://doi.org/10.1016/j.molcel.2011.11.028

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., … Jaenisch, R. (2010a). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(50), 21931–21936. https://doi.org/10.1073/pnas.1016071107

Cryan, J. F., & Dinan, T. G. (2012, October). Mind-altering microorganisms: The impact of the gut microbiota on brain and behaviour. *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn3346

Csibi, A., Lee, G., Yoon, S. O., Tong, H., Ilter, D., Elia, I., … Blenis, J. (2014). The mTORC1/S6K1 pathway regulates glutamine metabolism through the eif4b-dependent control of c-Myc translation. *Current Biology*, *24*(19), 2274–2280. https://doi.org/10.1016/j.cub.2014.08.007

Dai, Z., Ramesh, V., & Locasale, J. W. (2020). The evolving metabolic landscape of chromatin biology and epigenetics. *Nature Reviews Genetics*, 1–17. https://doi.org/10.1038/s41576-020-0270-8

Dalby, M. J., Ross, A. W., Walker, A. W., & Morgan, P. J. (2017). Dietary Uncoupling of Gut Microbiota and Energy Harvesting from Obesity and Glucose Tolerance in Mice. *Cell Reports*, *21*(6), 1521–1533. https://doi.org/10.1016/j.celrep.2017.10.056

Dalby, M., Rennie, S., & Andersson, R. (2017). FANTOM5 transcribed enhancers in hg38. https://doi.org/10.5281/ZENODO.556775

Dang, L., White, D. W., Gross, S., Bennett, B. D., Bittinger, M. A., Driggers, E. M., … Su, S. M. (2009). Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature*. https://doi.org/10.1038/nature08617

De La Serre, C. B., Ellis, C. L., Lee, J., Hartman, A. L., Rutledge, J. C., & Raybould, H. E. (2010). Propensity to high-fat diet-induced obesity in rats is associated with changes in the gut microbiota and gut inflammation. *American Journal of Physiology - Gastrointestinal and Liver Physiology*, *299*(2). https://doi.org/10.1152/ajpgi.00098.2010

De Smet, C., Lurquin, C., Lethé, B., Martelange, V., & Boon, T. (1999). DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a

CpG-rich promoter. *Molecular and Cellular Biology*, *19*(11), 7327–7335. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10523621

Decorsière, A., Mueller, H., Van Breugel, P. C., Abdul, F., Gerossier, L., Beran, R. K., … Strubin, M. (2016). Hepatitis B virus X protein identifies the Smc5/6 complex as a host restriction factor. *Nature*, *531*(7594), 386–389. https://doi.org/10.1038/nature17170

Degnan, P. H., Taga, M. E., & Goodman, A. L. (2014, November 4). Vitamin B12 as a modulator of gut microbial ecology. *Cell Metabolism*. Cell Press. https://doi.org/10.1016/j.cmet.2014.10.002

Dekker, J., Marti-Renom, M. A., & Mirny, L. A. (2013). Exploring the three-dimensional organization of genomes: Interpreting chromatin interaction data. *Nature Reviews Genetics*. https://doi.org/10.1038/nrg3454

DiNardo, C. D., Schimmer, A. D., Yee, K. W. L., Hochhaus, A., Kraemer, A., Carvajal, R. D., … Wei, A. (2016). A Phase I Study of IDH305 in Patients with Advanced Malignancies Including Relapsed/Refractory AML and MDS That Harbor IDH1R132 Mutations. *Blood*, *128*(22). Retrieved from http://www.bloodjournal.org/content/128/22/1073?sso-checked=true

Dixon, J. R., Gorkin, D. U., & Ren, B. (2016). Chromatin Domains: The Unit of Chromosome Organization. *Molecular Cell*, *62*(5), 668–680. https://doi.org/10.1016/j.molcel.2016.05.018

Dixon, J. R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J. E., Lee, A. Y., … Ren, B. (2015). Chromatin architecture reorganization during stem cell differentiation. *Nature*, *518*(7539), 331–336. https://doi.org/10.1038/nature14222

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., … Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. https://doi.org/10.1038/nature11082

Dogan, S., Chute, D. J., Xu, B., Ptashkin, R. N., Chandramohan, R., Casanova-Murphy, J., … Berger, M. F. (2017). Frequent IDH2 R172 mutations in undifferentiated and poorly-differentiated sinonasal carcinomas. *The Journal of Pathology*, *242*(4), 400–408. https://doi.org/10.1002/path.4915

Donohoe, D. R., Collins, L. B., Wali, A., Bigler, R., Sun, W., & Bultman, S. J. (2012). The Warburg Effect Dictates the Mechanism of Butyrate-Mediated Histone Acetylation and Cell Proliferation. *Molecular Cell*, *48*(4), 612–626. https://doi.org/10.1016/j.molcel.2012.08.033

Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., … Lochovsky, L. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414), 57–74. https://doi.org/10.1038/nature11247

Durinck S, Spellman P, Birney E, & Huber W. (2009). “Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt.” *Nature*

Protocols, *4*, 1184–1191. https://doi.org/10.18129/B9.BIOC.BIOMART

El-Sharnouby, S., Fischer, B., Magbanua, J. P., Umans, B., Flower, R., Choo, S. W., … White, R. (2017). Regions of very low H3K27me3 partition the Drosophila genome into topological domains. *PLOS ONE*, *12*(3), e0172725. https://doi.org/10.1371/journal.pone.0172725

Essafi, A., Webb, A., Berry, R. L., Slight, J., Burn, S. F., Spraggon, L., … Hohenstein, P. (2011). A Wt1-Controlled Chromatin Switching Mechanism Underpins Tissue-Specific Wnt4 Activation and Repression. *Developmental Cell*, *21*(3), 559–574. https://doi.org/10.1016/J.DEVCEL.2011.07.014

Farthing, C. R., Ficz, G., Ng, R. K., Chan, C. F., Andrews, S., Dean, W., … Reik, W. (2008). Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. *PLoS Genetics*, *4*(6). https://doi.org/10.1371/journal.pgen.1000116

Figueroa, M. E., Abdel-Wahab, O., Lu, C., Ward, P. S., Patel, J., Shih, A., … Melnick, A. (2010). Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. https://doi.org/10.1016/j.ccr.2010.11.015

Flavahan, W A, Drier, Y., Liau, B. B., Gillespie, S. M., Venteicher, A. S., Stemmer-Rachamimov, A. O., … Bernstein, B. E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*, *529*(7584), 110–114. https://doi.org/10.1038/nature16490

Flavahan, William A, Drier, Y., Johnstone, S. E., Hemming, M. L., Tarjan, D. R., Hegazi, E., … Bernstein, B. E. (2019a). Altered chromosomal topology drives oncogenic programs in SDH-deficient GIST. *Nature*. https://doi.org/10.1038/s41586-019-1668-3

Flavahan, William A, Drier, Y., Liau, B. B., Gillespie, S. M., Venteicher, A. S., Stemmer-Rachamimov, A. O., … Bernstein, B. E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*, *529*. https://doi.org/10.1038/nature16490

Fothergill, E., Guo, J., Howard, L., Kerns, J. C., Knuth, N. D., Brychta, R., … Hall, K. D. (2016). Persistent metabolic adaptation 6 years after "The Biggest Loser" competition. *Obesity*, *24*(8), 1612–1619. https://doi.org/10.1002/oby.21538

Freitas, M. R. P., Malheiros, S. M. F., Stávale, J. N., Biassi, T. P., Zamunér, F. T., Begnami, M. D. F. de S., … Vettore, A. L. (2013). Expression of cancer/testis antigens is correlated with improved survival in glioblastoma. *Oncotarget*, *4*(4), 636–646. https://doi.org/10.18632/oncotarget.950

Gaglio, D., Metallo, C. M., Gameiro, P. A., Hiller, K., Danna, L. S., Balestrieri, C., … Chiaradonna, F. (2011). Oncogenic K-Ras decouples glucose and glutamine metabolism to support cancer cell growth. *Molecular Systems Biology*, *7*. https://doi.org/10.1038/msb.2011.56

Galanty, Y., Belotserkovskaya, R., Coates, J., Polo, S., Miller, K. M., & Jackson, S. P. (2009). Mammalian SUMO E3-ligases PIAS1 and PIAS4 promote responses to DNA double-strand breaks. *Nature*, *462*(7275), 935–939. https://doi.org/10.1038/nature08657

Gao, P., Tchernyshyov, I., Chang, T. C., Lee, Y. S., Kita, K., Ochi, T., … Dang, C. V. (2009). C-Myc suppression of miR-23a/b enhances mitochondrial glutaminase expression and glutamine metabolism. *Nature*, *458*(7239), 762–765. https://doi.org/10.1038/nature07823

Gao, T., & Qian, J. (2020). EnhancerAtlas 2.0: An updated resource with enhancer annotation in 586 tissue/cell types across nine species. *Nucleic Acids Research*, *48*(D1), D58–D64. https://doi.org/10.1093/nar/gkz980

Gareau, J. R., & Lima, C. D. (2010, December 23). The SUMO pathway: Emerging mechanisms that shape specificity, conjugation and recognition. *Nature Reviews Molecular Cell Biology*. Nature Publishing Group. https://doi.org/10.1038/nrm3011

Gaspar, J. M. (2018). Improved peak-calling with MACS2. *BioRxiv*, 496521. https://doi.org/10.1101/496521

Gelman, S. J., Naser, F., Mahieu, N. G., McKenzie, L. D., Dunn, G. P., Chheda, M. G., & Patti, G. J. (2018). Consumption of NADPH for 2-HG Synthesis Increases Pentose Phosphate Pathway Flux and Sensitizes Cells to Oxidative Stress. *Cell Reports*, *22*(2), 512–522. https://doi.org/10.1016/J.CELREP.2017.12.050

Ghavi-Helm, Y., Jankowski, A., Meiers, S., Viales, R. R., Korbel, J. O., & Furlong, E. E. M. (2019). Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nature Genetics*, *51*(8), 1272–1282. https://doi.org/10.1038/s41588-019-0462-3

Gibney, E. R., & Nolan, C. M. (2010, July 12). Epigenetics and gene expression. *Heredity*. Nature Publishing Group. https://doi.org/10.1038/hdy.2010.54

Gu, P., Chen, X., Xie, R., Xie, W., Huang, L., Dong, W., … Lin, T. (2019). A novel AR translational regulator lncRNA LBCS inhibits castration resistance of prostate cancer. *Molecular Cancer*, *18*(1). https://doi.org/10.1186/s12943-019-1037-8

Gu, W., & Roeder, R. G. (1997). Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell*, *90*(4), 595–606. https://doi.org/10.1016/S0092-8674(00)80521-8

Haberland, M., Montgomery, R. L., & Olson, E. N. (2009, January). The many roles of histone deacetylases in development and physiology: Implications for disease and therapy. *Nature Reviews Genetics*. Nat Rev Genet. https://doi.org/10.1038/nrg2485

Hall, J. E., Da Silva, A. A., Do Carmo, J. M., Dubinion, J., Hamza, S., Munusamy, S., … Stec, D. E. (2010, June 4). Obesity-induced hypertension: Role of sympathetic nervous system, leptin, and melanocortins. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.R110.113175

Hamamoto, R., Furukawa, Y., Morita, M., Iimura, Y., Silva, F. P., Li, M., … Nakamura, Y. (2004). SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. *Nature Cell Biology*, *6*(8), 731–740. https://doi.org/10.1038/ncb1151

Hanssen, L. L. P., Kassouf, M. T., Oudelaar, A. M., Biggs, D., Preece, C., Downes, D. J., … Higgs, D. R. (2017). Tissue-specific CTCF–cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. *Nature Cell Biology*, *19*(8), 952–961. https://doi.org/10.1038/ncb3573

Hark, A. T., Schoenherr, C. J., Katz, D. J., Ingram, R. S., Levorse, J. M., & Tilghman, S. M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature*, *405*(6785), 486–489. https://doi.org/10.1038/35013106

Hartmann, C., Hentschel, B., Wick, W., Capper, D., Felsberg, J., Simon, M., … Von Deimling, A. (2010). Patients with IDH1 wild type anaplastic astrocytomas exhibit worse prognosis than IDH1-mutated glioblastomas, and IDH1 mutation status accounts for the unfavorable prognostic effect of higher age: Implications for classification of gliomas. *Acta Neuropathologica*, *120*(6), 707–718. https://doi.org/10.1007/s00401-010-0781-z

Hata, K., Okano, M., Lei, H., & Li, E. (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development*, *129*(8).

Hawkins, R. D., Hon, G. C., Lee, L. K., Ngo, Q., Lister, R., Pelizzola, M., … Ren, B. (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell*, *6*(5), 479–491. https://doi.org/10.1016/j.stem.2010.03.018

Haws, S. A., Yu, D., Ye, C., Wille, C. K., Nguyen, L. C., Krautkramer, K. A., … Denu, J. M. (2020). Methyl-Metabolite Depletion Elicits Adaptive Responses to Support Heterochromatin Stability and Epigenetic Persistence. *Molecular Cell*, *78*(2), 210-223.e8. https://doi.org/10.1016/j.molcel.2020.03.004

He, J., Fu, X., Zhang, M., He, F., Li, W., Abdul, M. M., … Hutchins, A. P. (2019). Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells. *Nature Communications*, *10*(1), 1–13. https://doi.org/10.1038/s41467-018-08006-y

He, Y., Hariharan, M., Gorkin, D. U., Dickel, D. E., Luo, C., Castanon, R. G., … Ecker, J. R. (2020). Spatiotemporal DNA methylome dynamics of the developing mouse fetus. *Nature*, *583*(7818), 752–759. https://doi.org/10.1038/s41586-020-2119-x

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., … Ren, B. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, *459*(7243), 108–112. https://doi.org/10.1038/nature07829

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., … Glass, C. K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-

Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell*, *38*(4), 576–589. https://doi.org/10.1016/j.molcel.2010.05.004

Heymsfield, S. B., & Wadden, T. A. (2017). Mechanisms, Pathophysiology, and Management of Obesity. *New England Journal of Medicine*, *376*(3), 254–266. https://doi.org/10.1056/NEJMra1514009

Hoffman, A. E., Zheng, T., Ba, Y., & Zhu, Y. (2008). The circadian gene NPAS2, a putative tumor suppressor, is involved in DNA damage response. *Molecular Cancer Research*, *6*(9), 1461–1468. https://doi.org/10.1158/1541-7786.MCR-07-2094

Hou, Y. P., He, Q. Q., Ouyang, H. M., Peng, H. S., Wang, Q., Li, J., … Yin, A. H. (2017). Human Gut Microbiota Associated with Obesity in Chinese Children and Adolescents. *BioMed Research International*, *2017*. https://doi.org/10.1155/2017/7585989

Howe, F. S., Fischl, H., Murray, S. C., & Mellor, J. (2017). Is H3K4me3 instructive for transcription activation? *BioEssays*, *39*(1), 1–12. https://doi.org/10.1002/bies.201600095

Hu, P., Wu, S., Sun, Y., Yuan, C.-C., Kobayashi, R., Myers, M. P., & Hernandez, N. (2002). Characterization of Human RNA Polymerase III Identifies Orthologues for Saccharomyces cerevisiae RNA Polymerase III Subunits. *Molecular and Cellular Biology*, *22*(22), 8044–8055. https://doi.org/10.1128/mcb.22.22.8044-8055.2002

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., Von Mering, C., & Bork, P. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Molecular Biology and Evolution*, *34*(8), 2115–2122. https://doi.org/10.1093/molbev/msx148

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., … White, O. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, *486*(7402), 207–214. https://doi.org/10.1038/nature11234

Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, *11*, 119. https://doi.org/10.1186/1471-2105-11-119

Hylemon, P. B., Zhou, H., Pandak, W. M., Ren, S., Gil, G., & Dent, P. (2009, August 1). Bile acids as regulatory molecules. *Journal of Lipid Research*. American Society for Biochemistry and Molecular Biology. https://doi.org/10.1194/jlr.R900007-JLR200

Inagaki, T., Moschetta, A., Lee, Y. K., Peng, L., Zhao, G., Downes, M., … Kliewer, S. A. (2006). Regulation of antibacterial defense in the small intestine by the nuclear bile acid receptor. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3920–3925. https://doi.org/10.1073/pnas.0509592103

Iyer, A. K., Zhang, Y.-H., & McCabe, E. R. B. (2006). Dosage-sensitive sex reversal adrenal hypoplasia congenita critical region on the X chromosome, gene 1 (DAX1) (NR0B1) and small heterodimer partner (SHP) (NR0B2) form homodimers individually, as well as DAX1-SHP heterodimers. *Molecular Endocrinology (Baltimore, Md.)*, *20*(10), 2326–

2342. https://doi.org/10.1210/me.2005-0383

Izquierdo-Garcia, J. L., Cai, L. M., Chaumeil, M. M., Eriksson, P., Robinson, A. E., Pieper, R. O., … Ronen, S. M. (2014). Glioma Cells with the IDH1 Mutation Modulate Metabolic Fractional Flux through Pyruvate Carboxylase. https://doi.org/10.1371/journal.pone.0108289

Izquierdo-Garcia, J. L., Viswanath, P., Eriksson, P., Cai, L., Radoul, M., Chaumeil, M. M., … Ronen, S. M. (2015). IDH1 Mutation Induces Reprogramming of Pyruvate Metabolism. *Cancer Research*, *75*(15), 2999–3009. https://doi.org/10.1158/0008-5472.CAN-15-0840

Izquierdo-Garcia, J. L., Viswanath, P., Eriksson, P., Chaumeil, M. M., Pieper, R. O., Phillips, J. J., & Ronen, S. M. (2015). Metabolic Reprogramming in Mutant IDH1 Glioma Cells. *PLOS ONE*, *10*(2), e0118781. https://doi.org/10.1371/journal.pone.0118781

Janku, F., Yap, T. A., & Meric-Bernstam, F. (2018, May 1). Targeting the PI3K pathway in cancer: Are we making headway? *Nature Reviews Clinical Oncology*. Nature Publishing Group. https://doi.org/10.1038/nrclinonc.2018.28

Johnson, B. E., Mazor, T., Hong, C., Barnes, M., Aihara, K., McLean, C. Y., … Costello, J. F. (2014). Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science*, *343*(6167), 189–193. https://doi.org/10.1126/science.1239947

Jones, P. A. (2012, July 29). Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*. Nature Publishing Group. https://doi.org/10.1038/nrg3230

Jones, P. A., & Baylin, S. B. (2002). The fundamental role of epigenetic events in cancer. *Nature Reviews Genetics*. Nature Publishing Group. https://doi.org/10.1038/nrg816

Kadaveru, K., Protiva, P., Greenspan, E. J., Kim, Y. I., & Rosenberg, D. W. (2012). Dietary methyl donor depletion protects against intestinal tumorigenesis in ApcMin/+ mice. *Cancer Prevention Research*, *5*(7), 911–920. https://doi.org/10.1158/1940-6207.CAPR-11-0544

Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M., & Saitou, M. (2013). Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *EMBO Journal*, *32*(3), 340–353. https://doi.org/10.1038/emboj.2012.331

Kalcheva, N., Albala, J., O'Guin, K., Rubino, H., Garner, C., & Shafit-Zagardo, B. (1995). Genomic structure of human microtubule-associated protein 2 (MAP-2) and characterization of additional MAP-2 isoforms. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(24), 10894–10898. https://doi.org/10.1073/pnas.92.24.10894

Kanehisa, M., & Goto, S. (2000, January 1). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*. Oxford University Press. https://doi.org/10.1093/nar/28.1.27

Kassambara, A. (n.d.). "ggplot2" Based Publication Ready Plots [R package ggpubr version

0.2.4].

Kassambara, A. (2020). rstatix: Pipe-Friendly Framework for Basic Statistical Tests. R package version 0.6.0.

Ke, J., An, Y., Cao, B., Lang, J., Wu, N., & Zhao, D. (2020). Orlistat-Induced Gut Microbiota Modification in Obese Mice. *Evidence-Based Complementary and Alternative Medicine*, *2020*. https://doi.org/10.1155/2020/9818349

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360. https://doi.org/10.1038/nmeth.3317

Kim, K. K., Chamberlin, H. M., Morgan, D. O., & Kim, S. H. (1996, October). Three-dimensional structure of human cyclin H, a positive regulator of the CDK-activating kinase. *Nature Structural Biology*. https://doi.org/10.1038/nsb1096-849

Kim, T. H., Barrera, L. O., Zheng, M., Qu, C., Singer, M. A., Richmond, T. A., … Ren, B. (2005). A high-resolution map of active promoters in the human genome. *Nature*, *436*(7052), 876–880. https://doi.org/10.1038/nature03877

Kinnaird, A., Zhao, S., Wellen, K. E., & Michelakis, E. D. (2016, October 24). Metabolic control of epigenetics in cancer. *Nature Reviews Cancer*. Nature Publishing Group. https://doi.org/10.1038/nrc.2016.82

Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X. S., & Ahringer, J. (2009). Differential chromatin marking of introns and expressed exons by H3K36me3. *Nature Genetics*, *41*(3), 376–381. https://doi.org/10.1038/ng.322

Kouzarides, T. (2007, February 23). Chromatin Modifications and Their Function. *Cell*. Elsevier. https://doi.org/10.1016/j.cell.2007.02.005

Kraus, F., Miron, E., Demmerle, J., Chitiashvili, T., Budco, A., Alle, Q., … Markaki, Y. (2017). Quantitative 3D structured illumination microscopy of nuclear structures. *Nature Protocols*, *12*(5), 1011–1028. https://doi.org/10.1038/nprot.2017.020

Krautkramer, K. A., Kreznar, J. H., Romano, K. A., Vivas, E. I., Barrett-Wilt, G. A., Rabaglia, M. E., … Denu, J. M. (2016). Diet-Microbiota Interactions Mediate Global Epigenetic Programming in Multiple Host Tissues. *Molecular Cell*, *64*(5), 982–992. https://doi.org/10.1016/j.molcel.2016.10.025

Krupenko, N. I., Holmes, R. S., Tsybovsky, Y., & Krupenko, S. A. (2015). Aldehyde dehydrogenase homologous folate enzymes: Evolutionary switch between cytoplasmic and mitochondrial localization. *Chemico-Biological Interactions*, *234*, 12–17. https://doi.org/10.1016/j.cbi.2014.12.022

Langmead, B., & Salzberg, S. L. (2012b). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lee, D.-H., Jin, S.-G., Cai, S., Chen, Y., Pfeifer, G. P., & O'Connor, T. R. (2005). Repair of

Methylation Damage in DNA and RNA by Mammalian AlkB Homologues. *Journal of Biological Chemistry*, *280*(47), 39448–39459. https://doi.org/10.1074/JBC.M509881200

Lee, J. M., Wagner, M., Xiao, R., Kim, K. H., Feng, D., Lazar, M. A., & Moore, D. D. (2014). Nutrient-sensing nuclear receptors coordinate autophagy. *Nature*, *516*(729), 112–115. https://doi.org/10.1038/nature13961

Lee, J. V., Carrer, A., Shah, S., Snyder, N. W., Wei, S., Venneti, S., … Wellen, K. E. (2014). Akt-dependent metabolic reprogramming regulates tumor cell Histone acetylation. *Cell Metabolism*, *20*(2), 306–319. https://doi.org/10.1016/j.cmet.2014.06.004

Lerin, C., Rodgers, J. T., Kalume, D. E., Kim, S. hee, Pandey, A., & Puigserver, P. (2006). GCN5 acetyltransferase complex controls glucose metabolism through transcriptional repression of PGC-1α. *Cell Metabolism*, *3*(6), 429–438. https://doi.org/10.1016/j.cmet.2006.04.013

Levy, D., Kuo, A. J., Chang, Y., Schaefer, U., Kitson, C., Cheung, P., … Gozani, O. (2011). Lysine methylation of the NF-κ B subunit RelA by SETD6 couples activity of the histone methyltransferase GLP at chromatin to tonic repression of NF-κ B signaling. *Nature Immunology*, *12*(1), 29–36. https://doi.org/10.1038/ni.1968

Ley, R. E., Bäckhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D., & Gordon, J. I. (2005). Obesity alters gut microbial ecology. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(31), 11070–11075. https://doi.org/10.1073/pnas.0504978102

Ley, R. E., Turnbaugh, P. J., Klein, S., & Gordon, J. I. (2006). Microbial ecology: Human gut microbes associated with obesity. *Nature*, *444*(7122), 1022–1023. https://doi.org/10.1038/4441022a

Li, B., Carey, M., & Workman, J. L. (2007, February 23). The Role of Chromatin during Transcription. *Cell*. Cell. https://doi.org/10.1016/j.cell.2007.01.015

Li, Daofeng, Hsu, S., Purushotham, D., Sears, R. L., & Wang, T. (2019). WashU Epigenome Browser update 2019. *Nucleic Acids Research*, *47*(W1), W158–W165. https://doi.org/10.1093/nar/gkz348

Li, Dinghua, Liu, C. M., Luo, R., Sadakane, K., & Lam, T. W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, *31*(10), 1674–1676. https://doi.org/10.1093/bioinformatics/btv033

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, X., Egervari, G., Wang, Y., Berger, S. L., & Lu, Z. (2018, September 1). Regulation of chromatin and gene expression by metabolic enzymes and metabolites. *Nature Reviews*

Molecular Cell Biology. Nature Publishing Group. https://doi.org/10.1038/s41580-018-0029-7

Liao, Y., Smyth, G. K., & Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Lieberman-Aiden, E., Miiro, G., Serwanga, J., Pozniak, A., McPhee, D., Manigart, O., … Lander, E. S. (2009). Principles of the Human Genome Comprehensive Mapping of Long-Range Interactions Reveals Folding Supporting Online Material Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome, *289*. https://doi.org/10.1126/science.1178746

Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., … Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, *326*(5950), 289–293. https://doi.org/10.1126/science.1181369

Liu, R., Hong, J., Xu, X., Feng, Q., Zhang, D., Gu, Y., … Wang, W. (2017). Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention. *Nature Medicine*, *23*(7), 859–868. https://doi.org/10.1038/nm.4358

Liu, X., Kato, Y., Kaneko, M. K., Sugawara, M., Ogasawara, S., Tsujimoto, Y., … Takagi, M. (2013). Isocitrate dehydrogenase 2 mutation is a frequent event in osteosarcoma detected by a multi-specific monoclonal antibody MsMab-1. *Cancer Medicine*, *2*(6), 803–814. https://doi.org/10.1002/cam4.149

Locasale, J. W. (2013, August 4). Serine, glycine and one-carbon units: Cancer metabolism in full circle. *Nature Reviews Cancer*. Nature Publishing Group. https://doi.org/10.1038/nrc3557

Lock, L. F., Takagi, N., & Martin, G. R. (1987). Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. *Cell*, *48*(1), 39–46. https://doi.org/10.1016/0092-8674(87)90353-9

Longo, D. L., Heymsfield, S. B., & Wadden, T. A. (2017). Mechanisms, Pathophysiology, and Management of Obesity. *N Engl J Med The New England Journal of Medicine Downloaded n Engl j Med*, *3763*(376), 254–266. https://doi.org/10.1056/NEJMra1514009

Louis, D. N., Ohgaki, H., Otmar, ·, Wiestler, D., Cavenee, W. K., Burger, P. C., … Kleihues, P. (2007). The 2007 WHO Classification of Tumours of the Central Nervous System. *Acta Neuropathol*, *114*, 97–109. https://doi.org/10.1007/s00401-007-0243-4

Louis, D. N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W. K., … Ellison, D. W. (2016, June 1). The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathologica*. Springer Verlag. https://doi.org/10.1007/s00401-016-1545-1

Louis, S., Tappu, R. M., Damms-Machado, A., Huson, D. H., & Bischoff, S. C. (2016). Characterization of the gut microbial community of obese patients following a weight-loss intervention using whole metagenome shotgun sequencing. *PLoS ONE*, *11*(2), 1–18. https://doi.org/10.1371/journal.pone.0149564

Love, M. I., Huber, W., & Anders, S. (2011). Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. *Genome Biology*, *15*. https://doi.org/10.1186/s13059-014-0550-8

Love, M. I., Huber, W., & Anders, S. (2014b). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lu, C., Ward, P. S., Kapoor, G. S., Rohle, D., Turcan, S., Abdel-Wahab, O., … Thompson, C. B. (2012). IDH mutation impairs histone demethylation and results in a block to cell differentiation. https://doi.org/10.1038/nature10860

Luo, C., Hajkova, P., & Ecker, J. R. (2018, September 28). Dynamic DNA methylation: In the right place at the right time. *Science*. American Association for the Advancement of Science. https://doi.org/10.1126/science.aat6806

Luong, A., Hannah, V. C., Brown, M. S., & Goldstein, J. L. (2000). Molecular characterization of human acetyl-CoA synthetase, an enzyme regulated by sterol regulatory element-binding proteins. *Journal of Biological Chemistry*, *275*(34), 26458–26466. https://doi.org/10.1074/jbc.M004160200

M. Gagné, L., Boulay, K., Topisirovic, I., Huot, M.-É., & Mallette, F. A. (2017a). Oncogenic Activities of IDH1/2 Mutations: From Epigenetics to Cellular Signaling. *Trends in Cell Biology*, *27*(10), 738–752. https://doi.org/10.1016/J.TCB.2017.06.002

Macrae, C. J., McCulloch, R. D., Ylanko, J., Durocher, D., & Koch, C. A. (2008). APLF (C2orf13) facilitates nonhomologous end-joining and undergoes ATM-dependent hyperphosphorylation following ionizing radiation. *DNA Repair*, *7*(2), 292–302. https://doi.org/10.1016/j.dnarep.2007.10.008

Maddocks, O. D. K., Labuschagne, C. F., Adams, P. D., & Vousden, K. H. (2016). Serine Metabolism Supports the Methionine Cycle and DNA/RNA Methylation through De Novo ATP Synthesis in Cancer Cells. *Molecular Cell*, *61*(2), 210–221. https://doi.org/10.1016/j.molcel.2015.12.014

Maga, G., Crespan, E., Markkanen, E., Imhof, R., Furrer, A., Villani, G., … Van Loon, B. (2013). DNA polymerase δ-interacting protein 2 is a processivity factor for DNA polymerase λ during 8-oxo-7,8-dihydroguanine bypass. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(47), 18850–18855. https://doi.org/10.1073/pnas.1308760110

Masri, S., Rigor, P., Cervantes, M., Ceglia, N., Sebastian, C., Xiao, C., … Sassone-Corsi, P. (2014). Partitioning circadian transcription by SIRT6 leads to segregated control of cellular metabolism. *Cell*, *158*(3), 659–672. https://doi.org/10.1016/j.cell.2014.06.050

Matoba, S., Liu, Y., Lu, F., Iwabuchi, K. A., Shen, L., Inoue, A., & Zhang, Y. (2014). Embryonic development following somatic cell nuclear transfer impeded by persisting histone methylation. *Cell*, *159*(4), 884–895. https://doi.org/10.1016/j.cell.2014.09.055

McBrayer, S. K., Mayers, J. R., DiNatale, G. J., Shi, D. D., Khanal, J., Chakraborty, A. A., … Kaelin, W. G. (2018). Transaminase Inhibition by 2-Hydroxyglutarate Impairs Glutamate Biosynthesis and Redox Homeostasis in Glioma. *Cell*, *175*(1), 101-116.e25. https://doi.org/10.1016/J.CELL.2018.08.038

Mclean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., … Bejerano, G. (2010). A n A ly s i s GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology VOLUME*, *28*. https://doi.org/10.1038/nbt.1630

McLendon, R., Friedman, A., Bigner, D., Van Meir, E. G., Brat, D. J., Mastrogianakis, G. M., … Thomson, E. (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, *455*(7216), 1061–1068. https://doi.org/10.1038/nature07385

McMurdie, P. J., & Holmes, S. (2013). Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, *8*(4). https://doi.org/10.1371/journal.pone.0061217

Mi, H., Muruganujan, A., Ebert, D., Huang, X., & Thomas, P. D. (2018). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research*, *47*, 419–426. https://doi.org/10.1093/nar/gky1038

Mifsud, B., Tavares-Cadete, F., Young, A. N., Sugar, R., Schoenfelder, S., Ferreira, L., … Osborne, C. S. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature Genetics*, *47*(6), 598–606. https://doi.org/10.1038/ng.3286

Mikami, A., Ogita, T., Namai, F., Shigemori, S., Sato, T., & Shimosato, T. (2020). Oral administration of Flavonifractor plautii attenuates inflammatory responses in obese adipose tissue. *Molecular Biology Reports*, *1*, 3. https://doi.org/10.1007/s11033-020-05727-6

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., … Bernstein, B. E. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, *448*(7153), 553–560. https://doi.org/10.1038/nature06008

Minamino, T., Morimoto, Y. V., Kawamoto, A., Terashima, H., & Imada, K. (2018). *Salmonella* Flagellum. In *Salmonella - A Re-emerging Pathogen*. InTech. https://doi.org/10.5772/intechopen.73277

Modrek, A. S., Golub, D., Khan, T., Bready, D., Prado, J., Bowman, C., … Placantonakis, D. G. (2017). Low-Grade Astrocytoma Mutations in IDH1, P53, and ATRX Cooperate to Block Differentiation [...] Materials and Methods. *Cell Reports*, *21*. Retrieved from

https://www.cell.com/cms/attachment/2119084828/2088983705/mmc1.pdf

Mondesir, J., Willekens, C., Touat, M., & de Botton, S. (2016). IDH1 and IDH2 mutations as novel therapeutic targets: current perspectives. *Journal of Blood Medicine*, *7*, 171–180. https://doi.org/10.2147/JBM.S70716

Mullen, A. R., Wheaton, W. W., Jin, E. S., Chen, P. H., Sullivan, L. B., Cheng, T., … Deberardinis, R. J. (2012). Reductive carboxylation supports growth in tumour cells with defective mitochondria. *Nature*, *481*(7381), 385–388. https://doi.org/10.1038/nature10642

Mundel, P., Heid, H. W., Mundel, T. M., Krüger, M., Reiser, J., & Kriz, W. (1997). Synaptopodin: An actin-associated protein in telencephalic dendrites and renal podocytes. *Journal of Cell Biology*, *139*(1), 193–204. https://doi.org/10.1083/jcb.139.1.193

Myung, J. K., Cho, H. J., Kim, H., Park, C. K., Lee, S. H., Choi, S. H., … Park, S. H. (2014). Prognosis of glioblastoma with oligodendroglioma component is associated with the IDH1 mutation and MGMT methylation status. *Translational Oncology*, *7*(6), 712–719. https://doi.org/10.1016/j.tranon.2014.10.002

Nagaraj, R., Sharpley, M. S., Chi, F., Braas, D., Zhou, Y., Kim, R., … Banerjee, U. (2017). Nuclear Localization of Mitochondrial TCA Cycle Enzymes as a Critical Step in Mammalian Zygotic Genome Activation. *Cell*, *168*(1–2), 210-223.e11. https://doi.org/10.1016/j.cell.2016.12.026

Nakahashi, H., Kieffer Kwon, K.-R., Resch, W., Vian, L., Dose, M., Stavreva, D., … Casellas, R. (2013). A Genome-wide Map of CTCF Multivalency Redefines the CTCF Code. *CellReports*, *3*, 1678–1689. https://doi.org/10.1016/j.celrep.2013.04.024

Nardelli, C., Granata, I., D'Argenio, V., Tramontano, S., Compare, D., Guarracino, M. R., … Sacchetti, L. (2020). Characterization of the Duodenal Mucosal Microbiome in Obese Adult Subjects by 16S rRNA Sequencing. *Microorganisms*, *8*(4), 485. https://doi.org/10.3390/microorganisms8040485

Neddermann, P., Gallinari, P., Lettieri, T., Schmid, D., Truong, O., Hsuan, J. J., … Jiricny, J. (1996). Cloning and expression of human G/T mismatch-specific thymine-DNA glycosylase. *Journal of Biological Chemistry*, *271*(22), 12767–12774. https://doi.org/10.1074/jbc.271.22.12767

Neville, B. A., Sheridan, P. O., Harris, H. M. B., Coughlan, S., Flint, H. J., Duncan, S. H., … O'Toole, P. W. (2013, July 23). Pro-Inflammatory Flagellin Proteins of Prevalent Motile Commensal Bacteria Are Variably Abundant in the Intestinal Microbiome of Elderly Humans. *PLoS ONE*. Public Library of Science. https://doi.org/10.1371/journal.pone.0068919

Ninova, M., Tóth, K. F., & Aravin, A. A. (2019, October 1). The control of gene expression and cell identity by H3K9 trimethylation. *Development (Cambridge)*. Company of Biologists Ltd. https://doi.org/10.1242/dev.181180

Nishi, R., Okuda, Y., Watanabe, E., Mori, T., Iwai, S., Masutani, C., … Hanaoka, F. (2005). Centrin 2 Stimulates Nucleotide Excision Repair by Interacting with Xeroderma Pigmentosum Group C Protein. *Molecular and Cellular Biology*, *25*(13), 5664–5674. https://doi.org/10.1128/mcb.25.13.5664-5674.2005

Nora, E. P., Goloborodko, A., Valton, A. L., Gibcus, J. H., Uebersohn, A., Abdennur, N., … Bruneau, B. G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*, *169*(5), 930-944.e22. https://doi.org/10.1016/j.cell.2017.05.004

Noushmehr, H., Weisenberger, D. J., Diefes, K., Phillips, H. S., Pujara, K., Berman, B. P., … Cancer Genome Atlas Research Network. (2010). Identification of a CpG Island Methylator Phenotype that Defines a Distinct Subgroup of Glioma. *Cancer Cell*, *17*(5), 510–522. https://doi.org/10.1016/j.ccr.2010.03.017

Núñez, F. J., Mendez, F. M., Kadiyala, P., Alghamri, M. S., Savelieff, M. G., Garcia-Fabiani, M. B., … Castro, M. G. (2019). IDH1-R132H acts as a tumor suppressor in glioma via epigenetic up-regulation of the DNA damage response. *Science Translational Medicine*, *11*(479). https://doi.org/10.1126/scitranslmed.aaq1427

Ohka, F., Ito, M., Ranjit, M., Senga, T., Motomura, A., Motomura, K., … Natsume, A. (2014a). Quantitative metabolome analysis profiles activation of glutaminolysis in glioma with IDH1 mutation. *Tumor Biology*, *35*(6), 5911–5920. https://doi.org/10.1007/s13277-014-1784-5

Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, *99*(3), 247–257. https://doi.org/10.1016/S0092-8674(00)81656-6

Okoye-Okafor, U. C., Bartholdy, B., Cartier, J., Gao, E. N., Pietrak, B., Rendina, A. R., … Steidl, U. (2015). New IDH1 mutant inhibitors for treatment of acute myeloid leukemia. *Nature Chemical Biology*, *11*(11), 878–886. https://doi.org/10.1038/nchembio.1930

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Simpson, G. L., Solymos, P., … Wagner, H. (2016). vegan: Community Ecology Package, R package version 2.4-0. *R Package Version 2.2-1*. Retrieved from http://vegan.r-forge.r-project.org

Olsen, L., Hansen, T., Jakobsen, K. D., Djurovic, S., Melle, I., Agartz, I., … Werge, T. (2008). The estrogen hypothesis of Schizophrenia implicates glucose metabolism: Association study in three independent samples. *BMC Medical Genetics*, *9*. https://doi.org/10.1186/1471-2350-9-39

Ooi, S. K. T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., … Bestor, T. H. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature*, *448*(7154), 714–717. https://doi.org/10.1038/nature05987

Pansuriya, T. C., van Eijk, R., d'Adamo, P., van Ruler, M. A. J. H., Kuijjer, M. L., Oosting, J., … Bovée, J. V. M. G. (2011). Somatic mosaic IDH1 and IDH2 mutations are associated with enchondroma and spindle cell hemangioma in Ollier disease and

Maffucci syndrome. *Nature Genetics*, *43*(12), 1256–1261. https://doi.org/10.1038/ng.1004

Parashuraman, S., & D'Angelo, G. (2019, January 1). Visualizing sphingolipid biosynthesis in cells. *Chemistry and Physics of Lipids*. Elsevier Ireland Ltd. https://doi.org/10.1016/j.chemphyslip.2018.11.003

Park, J., Kim, M., Kang, S. G., Jannasch, A. H., Cooper, B., Patterson, J., & Kim, C. H. (2015). Short-chain fatty acids induce both effector and regulatory T cells by suppression of histone deacetylases and regulation of the mTOR-S6K pathway. *Mucosal Immunology*, *8*(1), 80–93. https://doi.org/10.1038/mi.2014.44

Patel, J. H., Du, Y., Ard, P. G., Phillips, C., Carella, B., Chen, C.-J., … McMahon, S. B. (2004). The c-MYC Oncoprotein Is a Substrate of the Acetyltransferases hGCN5/PCAF and TIP60. *Molecular and Cellular Biology*, *24*(24), 10826–10834. https://doi.org/10.1128/mcb.24.24.10826-10834.2004

Payer, L. M., & Burns, K. H. (2019, December 1). Transposable elements in human genetic disease. *Nature Reviews Genetics*. Nature Publishing Group. https://doi.org/10.1038/s41576-019-0165-8

Pellegatta, S., Valletta, L., Corbetta, C., Patanè, M., Zucca, I., Riccardi Sirtori, F., … Finocchiaro, G. (2015). Effective immuno-targeting of the IDH1 mutation R132H in a murine model of intracranial glioma. *Acta Neuropathologica Communications*, *3*(1), 4. https://doi.org/10.1186/s40478-014-0180-0

Pereira, G. C., Sanchez, L., Schaughency, P. M., Rubio-Roldán, A., Choi, J. A., Planet, E., … Goodier, J. L. (2018). Properties of LINE-1 proteins and repeat element expression in the context of amyotrophic lateral sclerosis. *Mobile DNA*, *9*(1), 1–30. https://doi.org/10.1186/s13100-018-0138-z

Pickersgill, H., Kalverda, B., De Wit, E., Talhout, W., Fornerod, M., & Van Steensel, B. (2006). Characterization of the Drosophila melanogaster genome at the nuclear lamina. *Nature Genetics*, *38*(9), 1005–1014. https://doi.org/10.1038/ng1852

Plourde, M., Ferland, A., Soucy, P., Hamdi, Y., Tranchant, M., Durocher, F., … Simard, J. (2009). Analysis of 17β-hydroxysteroid dehydrogenase types 5, 7, and 12 genetic sequence variants in breast cancer cases from French Canadian Families with high risk of breast and ovarian cancer. *Journal of Steroid Biochemistry and Molecular Biology*, *116*(3–5), 134–153. https://doi.org/10.1016/j.jsbmb.2009.05.005

Poirier, L. A., Wise, C. K., Delongchamp, R. R., & Sinha, R. (2001). Blood Determinations of S-Adenosylmethionine, S-Adenosylhomocysteine, and Homocysteine. *Cancer Epidemiology and Prevention Biomarkers*, *10*(6).

Popovici-Muller, J., Lemieux, R. M., Artin, E., Saunders, J. O., Salituro, F. G., Travins, J., … Yen, K. (2018). Discovery of AG-120 (Ivosidenib): A First-in-Class Mutant IDH1 Inhibitor for the Treatment of IDH1 Mutant Cancers. *ACS Medicinal Chemistry Letters*, *9*(4), 300–305. https://doi.org/10.1021/acsmedchemlett.7b00421

Postler, T. S., & Ghosh, S. (2017). Understanding the Holobiont: How Microbial Metabolites Affect Human Health and Shape the Immune System. *Cell Metabolism*, *26*(1), 110–130. https://doi.org/10.1016/J.CMET.2017.05.008

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., & Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, *470*(7333), 279–285. https://doi.org/10.1038/nature09692

Raineri, S., & Mellor, J. (2018). IDH1: Linking Metabolism and Epigenetics. *Frontiers in Genetics*, *9*, 493. https://doi.org/10.3389/fgene.2018.00493

Ramírez, F., Ryan, D. P., Bj¨, B., Grüning, B., Grüning, G., Bhardwaj, V., … Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Web Server Issue Published Online*, *44*. https://doi.org/10.1093/nar/gkw257

Ramírez, Ryan, Grüning, Bhardwaj, Kilpert, Richter, … Manke. (2016). deepTools2: a next generation web server for deep-sequencing data analysis | Nucleic Acids Research | Oxford Academic. *Nucleic Acids Research*, *44*(W1), W160–W165.

Rao, S. S. P., Huang, S. C., Glenn St Hilaire, B., Engreitz, J. M., Perez, E. M., Kieffer-Kwon, K. R., … Aiden, E. L. (2017). Cohesin Loss Eliminates All Loop Domains. *Cell*, *171*(2), 305-320.e24. https://doi.org/10.1016/j.cell.2017.09.026

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., … Aiden, E. L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, *159*(7), 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021

Reddy, K. L., Zullo, J. M., Bertolino, E., & Singh, H. (2008). Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature*, *452*(7184), 243–247. https://doi.org/10.1038/nature06727

Reitman, Z. J., Jin, G., Karoly, E. D., Spasojevic, I., Yang, J., Kinzler, K. W., … Yan, H. (2011). Profiling the effects of isocitrate dehydrogenase 1 and 2 mutations on the cellular metabolome. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(8), 3270–3275. https://doi.org/10.1073/pnas.1019393108

Ridaura, V. K., Faith, J. J., Rey, F. E., Cheng, J., Duncan, A. E., Kau, A. L., … Gordon, J. I. (2013). Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science*, *341*(6150). https://doi.org/10.1126/science.1241214

Rivera, C. M., & Ren, B. (2013, September 26). Mapping human epigenomes. *Cell*. Cell Press. https://doi.org/10.1016/j.cell.2013.09.011

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011, January). Integrative genomics viewer. *Nature Biotechnology*. NIH Public Access. https://doi.org/10.1038/nbt.1754

Robinson, J. T., Turner, D., Durand, N. C., Thorvaldsdóttir, H., Mesirov, J. P., & Aiden, E. L. (2018). Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data. *Cell*

Systems, *6*(2), 256-258.e1. https://doi.org/10.1016/j.cels.2018.01.001

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Rohle, D., Popovici-Muller, J., Palaskas, N., Turcan, S., Grommes, C., Campos, C., … Mellinghoff, I. K. (2013). An inhibitor of mutant IDH1 delays growth and promotes differentiation of glioma cells. *Science*, *340*(6132), 626–630. https://doi.org/10.1126/science.1236062

Rooks, M. G., & Garrett, W. S. (2016, June 1). Gut microbiota, metabolites and host immunity. *Nature Reviews Immunology*. Nature Publishing Group. https://doi.org/10.1038/nri.2016.42

Rowe, H. M., Kapopoulou, A., Corsinotti, A., Fasching, L., Macfarlan, T. S., Tarabay, Y., … Trono, D. (2013). TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. *Genome Research*, *23*(3), 452–461. https://doi.org/10.1101/gr.147678.112

Sacksteder, K. A., Morrell, J. C., Wanders, R. J. A., Matalon, R., & Gould, S. J. (1999). MCD encodes peroxisomal and cytoplasmic forms of malonyl-CoA decarboxylase and is mutated in malonyl-CoA decarboxylase deficiency. *Journal of Biological Chemistry*, *274*(35), 24461–24468. https://doi.org/10.1074/jbc.274.35.24461

Salamanca-Cardona, L., Shah, H., Poot, A. J., Correa, F. M., Di Gialleonardo, V., Lui, H., … Keshari, K. R. (2017). In Vivo Imaging of Glutamine Metabolism to the Oncometabolite 2-Hydroxyglutarate in IDH1/2 Mutant Tumors. *Cell Metabolism*, *26*(6), 830-841.e3. https://doi.org/10.1016/J.CMET.2017.10.001

Sampson, T. R., Debelius, J. W., Thron, T., Janssen, S., Shastri, G. G., Ilhan, Z. E., … Mazmanian, S. K. (2016). Gut Microbiota Regulate Motor Deficits and Neuroinflammation in a Model of Parkinson's Disease. *Cell*, *167*(6), 1469-1480.e12. https://doi.org/10.1016/j.cell.2016.11.018

Samuel, B. S., Shaito, A., Motoike, T., Rey, F. E., Backhed, F., Manchester, J. K., … Gordon, J. I. (2008). Effects of the gut microbiota on host adiposity are modulated by the short-chain fatty-acid binding G protein-coupled receptor, Gpr41. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(43), 16767–16772. https://doi.org/10.1073/pnas.0808567105

Sanderson, S. M., Gao, X., Dai, Z., & Locasale, J. W. (2019, November 1). Methionine metabolism in health and cancer: a nexus of diet and precision medicine. *Nature Reviews Cancer*. Nature Publishing Group. https://doi.org/10.1038/s41568-019-0187-8

Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C. T., … Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. *Nature*, *419*(6905), 407–411. https://doi.org/10.1038/nature01080

Schübeler, D. (2015, January 15). Function and information content of DNA methylation. *Nature*. Nature Publishing Group. https://doi.org/10.1038/nature14192

Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G., & Rauscher, F. J. (2002). SETDB1: A novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes and Development*, *16*(8), 919–932. https://doi.org/10.1101/gad.973302

Schultz, M. D., He, Y., Whitaker, J. W., Hariharan, M., Mukamel, E. A., Leung, D., … Ecker, J. R. (2015). Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature*, *523*(7559), 212–216. https://doi.org/10.1038/nature14465

Schumacher, T., Bunse, L., Pusch, S., Sahm, F., Wiestler, B., Quandt, J., … Platten, M. (2014). A vaccine targeting mutant IDH1 induces antitumour immunity. *Nature*, *512*(7514), 324–327. https://doi.org/10.1038/nature13387

Schwartz, Y. B., Linder-Basso, D., Kharchenko, P. V, Tolstorukov, M. Y., Kim, M., Li, H.-B., … Pirrotta, V. (2012). Nature and function of insulator protein binding sites in the Drosophila genome. *Genome Research*, *22*(11), 2188–2198. https://doi.org/10.1101/gr.138156.112

Schwiertz, A., Taras, D., Schäfer, K., Beijer, S., Bos, N. A., Donus, C., & Hardt, P. D. (2010). Microbiota and SCFA in lean and overweight healthy subjects. *Obesity*, *18*(1), 190–195. https://doi.org/10.1038/oby.2009.167

Sdelci, S., Rendeiro, A. F., Rathert, P., You, W., Lin, J. M. G., Ringler, A., … Kubicek, S. (2019). MTHFD1 interaction with BRD4 links folate metabolism to transcriptional regulation. *Nature Genetics*, *51*(6), 990–998. https://doi.org/10.1038/s41588-019-0413-z

Sergushichev, A. A. (2016). An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *BioRxiv*, 060012. https://doi.org/10.1101/060012

Shimazu, T., Hirschey, M. D., Newman, J., He, W., Shirakawa, K., Le Moan, N., … Verdin, E. (2013). Suppression of oxidative stress by β-hydroxybutyrate, an endogenous histone deacetylase inhibitor. *Science*, *339*(6116), 211–214. https://doi.org/10.1126/science.1227166

Shogren-Knaak, M., Ishii, H., Sun, J. M., Pazin, M. J., Davie, J. R., & Peterson, C. L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science*, *311*(5762), 844–847. https://doi.org/10.1126/science.1124000

Shyh-Chang, N., Locasale, J. W., Lyssiotis, C. A., Zheng, Y., Teo, R. Y., Ratanasirintrawoot, S., … Cantley, L. C. (2013). Influence of threonine metabolism on S-adenosylmethionine and histone methylation. *Science*, *339*(6116), 222–226. https://doi.org/10.1126/science.1226603

Silva, S., Altmannova, V., Eckert-Boulet, N., Kolesar, P., Gallina, I., Hang, L., … Lisby, M. (2016). SUMOylation of Rad52-Rad59 synergistically change the outcome of mitotic recombination. *DNA Repair*, *42*, 11–25. https://doi.org/10.1016/j.dnarep.2016.04.001

Smyth, G., Hu, Y., Ritchie, M., Silver, J., Wettenhall, J., McCarthy, D., … Choi, D. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*. https://doi.org/10.18129/b9.bioc.limma

Son, J., Lyssiotis, C. A., Ying, H., Wang, X., Hua, S., Ligorio, M., … Kimmelman, A. C. (2013). Glutamine supports pancreatic cancer growth through a KRAS-regulated metabolic pathway. *Nature*, *496*(7443), 101–105. https://doi.org/10.1038/nature12040

Soto-Reyes, E., & Recillas-Targa, F. (2010). Epigenetic regulation of the human p53 gene promoter by the CTCF transcription factor in transformed cell lines. *Oncogene*, *29*(15), 2217–2227. https://doi.org/10.1038/onc.2009.509

Srivastava, G., & Apovian, C. M. (2017a). Current pharmacotherapy for obesity. *Nature Publishing Group*, *14*. https://doi.org/10.1038/nrendo.2017.122

Stuurman, N., Heins, S., & Aebi, U. (1998). Nuclear lamins: Their structure, assembly, and interactions. *Journal of Structural Biology*, *122*(1–2), 42–66. https://doi.org/10.1006/jsbi.1998.3987

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., … Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102

Sulkowski, P. L., Corso, C. D., Robinson, N. D., Scanlon, S. E., Purshouse, K. R., Bai, H., … Bindra, R. S. (2017). 2-Hydroxyglutarate produced by neomorphic IDH mutations suppresses homologous recombination and induces PARP inhibitor sensitivity. *Science Translational Medicine*, *9*(375). https://doi.org/10.1126/scitranslmed.aal2463

Sulkowski, P. L., Oeck, S., Dow, J., Economos, N. G., Mirfakhraie, L., Liu, Y., … Glazer, P. M. (2020). Oncometabolites suppress DNA repair by disrupting local chromatin signalling. *Nature*, 1–6. https://doi.org/10.1038/s41586-020-2363-0

Sutendra, G., Kinnaird, A., Dromparis, P., Paulin, R., Stenson, T. H., Haromy, A., … Michelakis, E. D. (2014b). A Nuclear Pyruvate Dehydrogenase Complex Is Important for the Generation of Acetyl-CoA and Histone Acetylation. https://doi.org/10.1016/j.cell.2014.04.046

Suzuki, H., Aoki, K., Chiba, K., Sato, Y., Shiozawa, Y., Shiraishi, Y., … Ogawa, S. (2015). Mutational landscape and clonal architecture in grade II and III gliomas. *Nature Genetics*, *47*(5), 458–468. https://doi.org/10.1038/ng.3273

Szmola, R., & Sahin-Tóth, M. (2007). Chymotrypsin C (caldecrin) promotes degradation of human cationic trypsin: Identity with Rinderknecht's enzyme Y. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(27), 11227–11232. https://doi.org/10.1073/pnas.0703714104

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., … Rao, A.

(2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science (New York, N.Y.)*, *324*(5929), 930–935. https://doi.org/10.1126/science.1170116

Takahashi, H., McCaffery, J. M., Irizarry, R. A., & Boeke, J. D. (2006). Nucleocytosolic Acetyl-Coenzyme A Synthetase Is Required for Histone Acetylation and Global Transcription. *Molecular Cell*, *23*(2), 207–217. https://doi.org/10.1016/j.molcel.2006.05.040

Takekawa, N., Terahara, N., Kato, T., Gohara, M., Mayanagi, K., Hijikata, A., … Homma, M. (2016). The tetrameric MotA complex as the core of the flagellar motor stator from hyperthermophilic bacterium. *Scientific Reports*, *6*(1), 1–8. https://doi.org/10.1038/srep31526

Tarazona, S., Furí o-Tarí, P., Di Pietro, A., Jo Nueda, M., Ferrer, A., & Conesa, A. (2015). Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Research*, *43*(21). https://doi.org/10.1093/nar/gkv711

Tarazona, S., Nueda, M. J., & Ferrer, A. (2016). *NOISeq: Differential Expression in RNA-seq*.

Taylor, E. M., Broughton, B. C., Botta, E., Stefanini, M., Sarasin, A., Jaspers, N. G. J., … Lehmann, A. R. (1997). Xeroderma pigmentosum and trichothiodystrophy are associated with different mutations in the XPD (ERCC2) repair/transcription gene. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(16), 8658–8663. https://doi.org/10.1073/pnas.94.16.8658

Tenenbaum, D. (2019). KEGGREST: Client-side REST access to KEGG. R package version 1.24.1.

Thaiss, C. A., Itav, S., Rothschild, D., Meijer, M. T., Levy, M., Moresi, C., … Elinav, E. (2016). Persistent microbiome alterations modulate the rate of post-dieting weight regain. *Nature*, *540*(7634), 544–551. https://doi.org/10.1038/nature20796

Thienpont, B., Steinbacher, J., Zhao, H., D'Anna, F., Kuchnio, A., Ploumakis, A., … Lambrechts, Di. (2016). Tumour hypoxia causes DNA hypermethylation by reducing TET activity. *Nature*, *537*(7618), 63–68. https://doi.org/10.1038/nature19081

Thinnes, C. C., England, K. S., Kawamura, A., Chowdhury, R., Schofield, C. J., & Hopkinson, R. J. (2014). Targeting histone lysine demethylases — Progress, challenges, and the future. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1839*(12), 1416–1432. https://doi.org/10.1016/J.BBAGRM.2014.05.009

Tremaroli, V., Karlsson, F., Werling, M., Ståhlman, M., Kovatcheva-Datchary, P., Olbers, T., … Bäckhed, F. (2015). Roux-en-Y Gastric Bypass and Vertical Banded Gastroplasty Induce Long-Term Changes on the Human Gut Microbiome Contributing to Fat Mass Regulation. *Cell Metabolism*, *22*(2), 228–238. https://doi.org/10.1016/j.cmet.2015.07.009

Tsukada, Y.-I., Fang, J., Erdjument-Bromage, H., Warren, M. E., Borchers, C. H., Tempst, P.,

& Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. https://doi.org/10.1038/nature04433

Turcan, S., Makarov, V., Taranda, J., Wang, Y., Fabius, A. W. M., Wu, W., … Chan, T. A. (2018). Mutant-IDH1-dependent chromatin state reprogramming, reversibility, and persistence. *Nature Genetics*, *50*(1), 62–72. https://doi.org/10.1038/s41588-017-0001-z

Turcan, S., Rohle, D., Goenka, A., Walsh, L. A., Fang, F., Yilmaz, E., … Chan, T. A. (2012). IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature*, *483*. https://doi.org/10.1038/nature10866

Turnbaugh, P. J., Bäckhed, F., Fulton, L., & Gordon, J. I. (2008). Diet-Induced Obesity Is Linked to Marked but Reversible Alterations in the Mouse Distal Gut Microbiome. *Cell Host and Microbe*, *3*(4), 213–223. https://doi.org/10.1016/j.chom.2008.02.015

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., … Gordon, J. I. (2009). A core gut microbiome in obese and lean twins. *Nature*, *457*(7228), 480–484. https://doi.org/10.1038/nature07540

Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., & Gordon, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, *444*(7122), 1027–1031. https://doi.org/10.1038/nature05414

Uphoff, C. C., & Drexler, H. G. (2014). Detection of mycoplasma contamination in cell cultures. *Current Protocols in Molecular Biology*, *2014*. https://doi.org/10.1002/0471142727.mb2804s106

Utley, R. T., Ikeda, K., Grant, P. A., Côté, J., Steger, D. J., Eberharter, A., … Workman, J. L. (1998). Transcriptional activators direct histone acetyltransferase complexes to nucleosomes. *Nature*, *394*(6692), 498–502. https://doi.org/10.1038/28886

Van Bortle, K., Ramos, E., Takenaka, N., Yang, J., Wahi, J. E., & Corces, V. G. (2012). Drosophila CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains. *Genome Research*, *22*(11), 2176–2187. https://doi.org/10.1101/gr.136788.111

Van Driel, L. M. J. W., Eijkemans, M. J. C., De Jonge, R., De Vries, J. H. M., Van Meurs, J. B. J., Steegers, E. A. P., & Steegers-Theunissen, R. P. M. (2009). Body mass index is an important determinant of methylation biomarkers in women of reproductive ages. *Journal of Nutrition*, *139*(12), 2315–2321. https://doi.org/10.3945/jn.109.109710

Verdugo, R., & Orostica, K. (2020). chromPlot: Global visualization tool of genomic data. R package version 1.16.0.

Verhaak, R. G., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D., … Neil Hayes, D. (2010). Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*, *17*, 98–110. https://doi.org/10.1016/j.ccr.2009.12.020

Verhaak, R. G. W., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D., …

Cancer Genome Atlas Research Network. (2010). Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*, *17*(1), 98–110. https://doi.org/10.1016/j.ccr.2009.12.020

Vijay-Kumar, M., Aitken, J. D., Carvalho, F. A., Cullender, T. C., Mwangi, S., Srinivasan, S., … Gewirtz, A. T. (2010). Metabolie syndrome and altered gut microbiota in mice lacking toll-like receptor 5. *Science*, *328*(5975), 228–231. https://doi.org/10.1126/science.1179721

Waitkus, M. S., Diplas, B. H., & Yan, H. (n.d.). Isocitrate dehydrogenase mutations in gliomas Normal Biochemistry of Isocitrate Dehydrogenases. https://doi.org/10.1093/neuonc/nov136

Wang, C., Liu, X., Gao, Y., Yang, L., Li, C., Liu, W., … Gao, S. (2018). Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. https://doi.org/10.1038/s41556-018-0093-4

Wang, F., Meng, J., Zhang, L., Johnson, T., Chen, C., & Roy, S. (2018). Morphine induces changes in the gut microbiome and metabolome in a morphine dependence model. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-21915-8

Wang, S., Su, J. H., Beliveau, B. J., Bintu, B., Moffitt, J. R., Wu, C. T., & Zhuang, X. (2016). Spatial organization of chromatin domains and compartments in single chromosomes. *Science*, *353*(6299), 598–602. https://doi.org/10.1126/science.aaf8084

Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., Pääbo, S., Rebhan, M., & Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genetics*, *39*(4), 457–466. https://doi.org/10.1038/ng1990

Wellen, K. E., Hatzivassiliou, G., Sachdeva, U. M., Bui, T. V., Cross, J. R., & Thompson, C. B. (2009a). ATP-citrate lyase links cellular metabolism to histone acetylation. *Science*, *324*(5930), 1076–1080. https://doi.org/10.1126/science.1164097

Weller, M., Wick, W., & von Deimling, A. (2011). Isocitrate dehydrogenase mutations: A challenge to traditional views on the genesis and malignant progression of gliomas. *Glia*, *59*(8), 1200–1204. https://doi.org/10.1002/glia.21130

Wen, B., Wu, H., Shinkai, Y., Irizarry, R. A., & Feinberg, A. P. (2009). Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nature Genetics*, *41*(2), 246–250. https://doi.org/10.1038/ng.297

Wen, P. Y., & Reardon, D. A. (2016, February 1). Neuro-oncology in 2015: Progress in glioma diagnosis, classification and treatment. *Nature Reviews Neurology*. Nature Publishing Group. https://doi.org/10.1038/nrneurol.2015.242

White, G. R. M., Varley, J. M., & Heighway, J. (1998). Isolation and characterization of a human homologue of the latrophilin gene from a region of 1p31.1 implicated in breast

cancer. *Oncogene*, *17*(26), 3513–3519. https://doi.org/10.1038/sj.onc.1202487

WHO. (n.d.). Obesity and overweight. Retrieved April 13, 2020, from https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wikoff, W. R., Anfora, A. T., Liu, J., Schultz, P. G., Lesley, S. A., Peters, E. C., & Siuzdak, G. (2009). Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(10), 3698–3703. https://doi.org/10.1073/pnas.0812874106

Witt, H., Luck, W., Hennies, H. C., Classen, M., Kage, A., Lass, U., … Becker, M. (2000). Mutations in the gene encoding the serine protease inhibitor, Kazal type 1 are associated with chronic pancreatitis. *Nature Genetics*, *25*(2), 213–216. https://doi.org/10.1038/76088

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, *20*(1), 257. https://doi.org/10.1186/s13059-019-1891-0

Wu, X., & Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. *Nature Reviews Genetics*, *18*(9), 517–534. https://doi.org/10.1038/nrg.2017.33

Wysocka, J., Swigut, T., Xiao, H., Milne, T. A., Kwon, S. Y., Landry, J., … Allis, C. D. (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature*, *442*(7098), 86–90. https://doi.org/10.1038/nature04815

Xiao, M., Yang, H., Xu, W., Ma, S., Lin, H., Zhu, H., … Guan, K. L. (2012). Inhibition of a-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. *Genes and Development*, *26*(12), 1326–1338. https://doi.org/10.1101/gad.191056.112

Xie, Q., Bai, Y., Wu, J., Sun, Y., Wang, Y., Zhang, Y., … Yuan, Z. (2011). Methylation-mediated regulation of E2F1 in DNA damage-induced cell death. *Journal of Receptors and Signal Transduction*, *31*(2), 139–146. https://doi.org/10.3109/10799893.2011.552914

Xiong, Y., Miyamoto, N., Shibata, K., Valasek, M. A., Motoike, T., Kedzierski, R. M., & Yanagisawa, M. (2004). Short-chain fatty acids stimulate leptin production in adipocytes through the G protein-coupled receptor GPR41. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(4), 1045–1050. https://doi.org/10.1073/pnas.2637002100

Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S.-H., … Xiong, Y. (2011). Oncometabolite 2-Hydroxyglutarate Is a Competitive Inhibitor of α-Ketoglutarate-Dependent Dioxygenases. *Cancer Cell*, *19*(1), 17–30.

https://doi.org/10.1016/J.CCR.2010.12.014

Yan, H., Williams, D., Jin, G., Mclendon, R., Rasheed, B. A., Yuan, W., … Bigner, D. D. (2009). *IDH1 and IDH2 Mutations in Gliomas*. *N Engl J Med* (Vol. 360). Retrieved from https://www.nejm.org/doi/pdf/10.1056/NEJMoa0808710

Yen, K., Konteatis, Z., Sui, Z., Artin, E., Dang, L., Straley, K., … Popovici-Muller, J. (2018). Abstract B126: AG-881, a brain penetrant, potent, pan-mutant IDH (mIDH) inhibitor for use in mIDH solid and hematologic malignancies. In *Therapeutic Agents: Other Topics* (Vol. 17, pp. B126–B126). American Association for Cancer Research. https://doi.org/10.1158/1535-7163.TARG-17-B126

Yen, K., Travins, J., Wang, F., David, M. D., Artin, E., Straley, K., … Su, S.-S. M. (2017). AG-221, a First-in-Class Therapy Targeting Acute Myeloid Leukemia Harboring Oncogenic IDH2 Mutations. *Cancer Discovery*, *7*(5), 478–493. https://doi.org/10.1158/2159-8290.CD-16-1034

Zhang, H., DiBaise, J. K., Zuccolo, A., Kudrna, D., Braidotti, M., Yu, Y., … Krajmalnik-Brown, R. (2009). Human gut microbiota in obesity and after gastric bypass. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(7), 2365–2370. https://doi.org/10.1073/pnas.0812600106

Zhao, S., Yin, L., Guo, Y., Sheng, Q., & Shyr, Y. (2020). Heatmap3: An Improved Heatmap Package. R package version 1.1.7.

Zheng, H., Huang, B., Zhang, B., Xiang, Y., Du, Z., Xu, Q., … Xie, W. (2016). Resetting Epigenetic Memory by Reprogramming of Histone Modifications in Mammals. *Molecular Cell*, *63*(6), 1066–1079. https://doi.org/10.1016/j.molcel.2016.08.032

Zheng, H., & Xie, W. (2019, September 1). The role of 3D genome organization in development and cell differentiation. *Nature Reviews Molecular Cell Biology*. Nature Publishing Group. https://doi.org/10.1038/s41580-019-0132-4

Zhong, L., D'Urso, A., Toiber, D., Sebastian, C., Henry, R. E., Vadysirisack, D. D., … Mostoslavsky, R. (2010). The Histone Deacetylase Sirt6 Regulates Glucose Homeostasis via Hif1α. *Cell*, *140*(2), 280–293. https://doi.org/10.1016/j.cell.2009.12.041

Zhu, J., Adli, M., Zou, J. Y., Verstappen, G., Coyne, M., Zhang, X., … Bernstein, B. E. (2013). Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell*, *152*(3), 642–654. https://doi.org/10.1016/j.cell.2012.12.033

Ziller, M. J., Gu, H., Müller, F., Donaghey, J., Tsai, L. T. Y., Kohlbacher, O., … Meissner, A. (2013). Charting a dynamic DNA methylation landscape of the human genome. *Nature*, *500*(7463), 477–481. https://doi.org/10.1038/nature12433

Hoyles, L. (2009), PhD Thesis. *In vitro* examination of the effect of orlistat on the ability of the faecal microbiota to utilize dietary lipids