Aberrant gene expression induced by a high fat diet is linked to H3K9 acetylation in the promoter-proximal region

Núria Morral[1,2], Sheng Liu[1], Abass M. Conteh[2], Xiaona Chu[1], Yue Wang[1], X. Charlie Dong[2], Yunlong Liu[1,3], Amelia K. Linnemann[2,4], Jun Wan[1,3]

[1]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN; [2]Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN; [3]Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN; [4]Department of Pediatrics, Indiana University School of Medicine, Indianapolis, IN

Correspondence should be addressed to:

Núria Morral, PhD
Department of Medical and Molecular Genetics
Indiana University School of Medicine
975 West Walnut Street, room IB 130
Indianapolis, IN 46202
Phone: 317-2789039
e-mail: nmorralc@iu.edu

This is the author's manuscript of the article published in final edited form as:

Morral, N., Liu, S., Conteh, A. M., Chu, X., Wang, Y., Dong, X. C., Liu, Y., Linnemann, A. K., & Wan, J. (2021). Aberrant gene expression induced by a high fat diet is linked to H3K9 acetylation in the promoter-proximal region. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, 1864(3), 194691. https://doi.org/10.1016/j.bbagrm.2021.194691

ABSTRACT

Non-alcoholic fatty liver disease (NAFLD) is the most common chronic liver disease, with an estimated global prevalence of 1 in 4 individuals. Aberrant transcriptional control of gene expression is central to the pathophysiology of metabolic diseases. However, the molecular mechanisms leading to gene dysregulation are not well understood. Histone modifications play important roles in the control of transcription. Acetylation of histone 3 at lysine 9 (H3K9ac) is associated with transcriptional activity and is implicated in transcript elongation by controlling RNA polymerase II (RNAPII) pause-release. Hence, changes in this histone modification may shed information on novel pathways linking transcription control and metabolic dysfunction. Here, we carried out genome-wide analysis of H3K9ac in the liver of mice fed a control or a high-fat diet (an animal model of NAFLD), and asked whether this histone mark associates with changes in gene expression. We found that over 70% of RNAPII peaks in promoter-proximal regions overlapped with H3K9ac, consistent with a role of H3K9ac in the regulation of transcription. When comparing high-fat with control diet, approximately 17% of the differentially expressed genes were associated with changes in H3K9ac in their promoters, showing a strong correlation between changes in H3K9ac signal and gene expression. Overall, our data indicate that in response to a high-fat diet, dysregulated gene expression of a subset of genes may be attributable to changes in transcription elongation driven by H3K9ac. Our results point at an added mechanism of gene regulation that may be important in the development of metabolic diseases.

Keywords: histone acetylation; gene regulation; high-fat diet; liver; non-alcoholic fatty liver disease

INTRODUCTION

Non-alcoholic fatty liver disease (NAFLD) is the most common chronic liver disease, affecting ~25% of the global population [1]. NAFLD is associated with obesity, and is a risk factor for other metabolic diseases such as type 2 diabetes and cardiovascular disease [2, 3]. The liver is a critical tissue for energy homeostasis, switching its genetic programs to execute anabolic or catabolic functions in response to nutrient availability. The precise regulation of chromatin structure together with transcription factor activity enables the liver to activate or suppress gene networks in response to energy needs.

Nutrients and intermediates of cell metabolism serve as cofactors for chromatin-modifying enzymes to connect metabolic information with transcriptional control of gene expression [4]. The histone code states that DNA transcription is regulated in part by post-translational chemical modifications to histone proteins. Histone tails are extensively modified by 'writers', enzymes that utilize cellular metabolites such as acetyl-CoA, S-adenosylmethionine, or ATP as substrates, as well as ‘erasers’, enzymes that remove these modifications. Histone modifications act as recruiters of transcription factors and/or co-regulators to promote euchromatin or heterochromatin, activating or inactivating gene expression.

A key aspect of the pathophysiology of metabolic disease is abnormal transcriptional control of gene expression, and ongoing studies are providing evidence that epigenetic mechanisms contribute to its progression [4, 5]. Excess calorie consumption from carbohydrates and fats are main drivers of energy imbalance and alterations in metabolic pathways. Histone acetylation is highly sensitive to the availability of glucose-derived cytosolic acetyl-CoA [6-10]. In addition, several metabolites and cofactors generated through glycolysis, including acetyl-CoA, pyruvate and lactate, directly alter lysine acetyltransferase or deacetylase activity to link energy status with cellular and organismal homeostasis [11, 12]. Acetylation of histone 3 at lysine 9 (H3K9ac) is a marker of actively transcribing genes [13-15], and it has been shown to be

necessary for recruitment of the Super Elongation Complex to chromatin and transition from RNA Polymerase II (RNAPII) pause-release to transcript elongation [16, 17]. We questioned whether high-fat diet feeding elicits genome-wide alterations in H3K9ac, and using high-throughput technologies we explored the connection between this histone modification and gene expression in this mouse model of NAFLD.

2. Materials and methods

2.1. Animals

All animal studies were in accordance with the National Institutes of Health guidelines and were approved by the Indiana University School of Medicine Institutional Animal Care and Use Committee. Control mice from the IL6 receptor knockout colony (cre^+ wild type or cre^- flox/flox) were used for these studies [18]. The colony is maintained in a C57BL/6J background. Eight-week old male mice were fed a control (2018SX, 18% kcal fat, Envigo, a crude laboratory chow diet) or a high-fat diet (D12492, 60 kcal% fat, Research Diets; contains 7% kcal from fructose, as well as 279.6 mg/kg cholesterol, from lard and casein) for 10 weeks (n=4/group). This diet promotes hepatic steatosis and inflammation, relative to chow diet fed mice [19]. A group of mice from the colony was used to assess lean and fat mass, and to confirm glucose intolerance (Supplementary Fig. 1). Mice were kept in a BSL1 room and had free access to food and water. Animals were euthanized under fed (*ad libitum*) conditions, and tissue collection was initiated at 1:00 pm. Liver tissues were collected, rinsed in sterile PBS and snap frozen in liquid nitrogen.

2.2. ChIP-seq

Frozen liver tissue (50 to 100 mg) from three biological replicates per group was used to isolate chromatin using the MAGnify Chromatin Immunoprecipitation System kit (Invitrogen, Carlsbad, CA). Tissue was minced in cold D-PBS and chromatin was immediately crosslinked with 1% methanol-free formaldehyde for 6 min. Crosslinking was stopped with glycine to a final concentration of 0.125 M. Tissue was homogenized at 4°C in 1-mL syringes by passing ~20 times through 18G followed by 21G needles. Lysis buffer was added and cells were sonicated using Bioruptor UCD-300 (23 cycles of 30 seconds ON, 30 seconds OFF), to generate DNA fragments of an average size of 100-400 bp. Sonicated chromatin was immunoprecipitated overnight at 4°C with an antibody specific to H3K9ac (#17-658, Millipore, Burlington, MA). Prior

to library construction, the shearing quality of DNA was assessed by Bioanalyzer 2100 (Agilent Technologies Inc., Santa Clara, CA). Purified ChIP and input DNA samples were used for library preparation using Illumina TruSeq Nano DNA LT Library Prep Kit (Cat# FC-121-4001), including end-repair, dA-tailing, indexed adaptor ligation and amplification. Each resulting indexed library was quantified and its quality accessed by Qubit and Agilent Bioanalyzer. Multiple libraries were pooled in equal molarity. The pooled libraries were then denatured, and neutralized, before loading onto NextSeq 500 sequencer at 1.5 pM final concentration for 75-bp paired-end sequencing (Illumina, Inc.). Approximately 20 million reads per library were generated. A Phred quality score (Q score) was used to measure the quality of sequencing. More than 90% of the sequencing reads reached Q30 (99.9% base call accuracy). About 84% of the reads were uniquely mapped to mouse reference genome mm10 using Bowtie2. Peaks were identified with MACS2 and peaks from ENCODE blacklist were subsequently removed. UCSC genome browser tracks were generated to visualize the alignment and the peaks.

Peaks from multiple samples were merged to form a final set of unique regions using BEDTools (v2.29.0) [20]. Reads falling into the regions from different samples were counted using featureCounts (v 1.6.2) [21]. Differential analysis of binding signals between different conditions was carried out based on the reads within the regions between different conditions by using edgeR (v 3.24.3) with trimmed mean M-values normalization [22, 23].

RNAPII ChIP-Seq peaks from liver were retrieved from the GEO database (accession number GSE21696: GSM541303, GSM541306) [24], and were converted to mm10 using LiftOver [25]. Venn diagrams were plotted for overlaps between RNAPII peaks and H3K9ac peaks in control diet samples. Significance of overlaps of peaks was derived from Fisher exact test using BEDTools [20]. Two sample Kolmogorov-Smirnov Test was performed on H3K9ac peak signal overlapping versus not overlapping with RNAPII using the ks.test function in R [26].

2.3. RNA-seq

Total RNA was extracted using a RNeasy Maxi kit (Qiagen, Valencia, CA), following the manufacturer's protocol (n=4/group). RNA integrity was evaluated by Bioanalyzer 2100. mRNA libraries were generated from 100 ng RNA, using the KAPA mRNA Hyperprep kit (Roche, Indianapolis, IN). Paired end 75-bp reads were generated with the Illumina HiSeq4000 platform (~45 million reads/sample). Nearly 85% of reads were uniquely mapped to mouse genome reference mm10 using STAR (Spliced Transcripts Alignment to a Reference) [27]. To evaluate the quality of the RNA-seq data, number of reads that fall into different annotated regions (exonic, intronic, splicing junction, intergenic, promoter, UTR, etc.) of the reference genes were determined with bamUtils [28]. Low quality mapped reads (including reads mapped to multiple positions) were excluded and featureCounts [21] was used to quantify the gene level expression. Differential gene expression analysis was performed with edgeR [22].

2.4. ATAC-seq

ATAC-seq was conducted using two biological replicates from the control diet group using approximately 15 mg of frozen liver tissue. Tissue samples were cut into small pieces and were homogenized using a glass dounce tissue grinder (25 times with pestle A and 25 times with pestle B) in 2 ml of ice cold EZ PREP (Sigma, Cat #NUC-101) and incubated on ice for 5 min. Nuclei were centrifuged at 500 g for 5 min at 4°C, washed with 2 ml ice-cold EZ PREP and incubated on ice for another 5 min. After centrifugation, the nuclei were washed in 2 ml nuclei suspension buffer (NSB) consisting of 1xPBS, 0.01% BSA. Isolated nuclei were resuspended in 2 ml NSB, filtered through a 35 um cell strainer, and counted. Approximately 70,000 nuclei were pelleted at 500xg for 10 min and resuspended in Tn5 enzyme and transposase buffer (Illumina Nextera® DNA library preparation kit, FC-121-1030). The Nextera libraries were amplified using the Nextera® PCR master mix and KAPA biosystems HiFi hotstart readymix successively. AMPure XP beads (Beckman Coulter) were used to purify the transposed DNA and the amplified PCR products. All libraries were sequenced on a 100 cycle paired-end run on an

Illumina NOVAseq instrument. The resulting ATAC-seq libraries were sequenced on Illumina NovaSeq 6000 at CMG of Indiana University School of Medicine and paired-end 50-bp reads were generated (~160 million reads per sample). Illumina adapter sequences and low quality base calls were trimmed off the paired-end reads with Trim Galore v0.4.3. The resulting high-quality reads were aligned to the human reference genome hg38 using bowtie2 (version 2.3.2) 63 with parameters "-X 2000 --no-mixed --no-discordant". Duplicate reads were discarded with Picard (https://broadinstitute.github.io/picard/). Reads mapped to mitochondrial DNA (<3.5%) together with low mapping quality reads (MAPQ<10) were excluded from further analysis. MACS2 (version 2.1.0) was used to identify general peaks. Sequencing and data analysis were carried out by the Center for Medical Genomics at Indiana University School of Medicine.

Unique open chromatin regions for multiple samples were identified as described for H3K9ac ChIP-seq analysis. Reads falling into each unique region were evaluated using pyDNase (v 0.3.0) [29] for different samples.

2.5. Data integration and transcription factor motif analysis

Promoter-proximal regions were defined as -3 kb to +2 kb relative to the transcription start site (TSS), and gene bodies from the +2 kb to the end of the gene annotation. Intergenic regions were DNA sequences outside of this range. Average signals around transcription start sites were plotted using deeptools [30]. Transcription factor motif enrichment analysis was performed using Homer (v 4.10.4) [31] on ATAC-seq unique regions with notable changes in H3K9ac signals and significant gene expression differentiation at the cutoff of false discovery rate (FDR) less than 0.05.

The alignment from RNA-seq, ChIP-seq, and ATAC-seq were converted to bigwig format to generate tracks for UCSC genome browser [25] using BEDTools [20] and bedGraphToBigWig [32].

Sequencing experiments for RNA-seq, ChIP-seq, and ATAC-seq, sequence alignment, RNA-seq and ChIP-seq data processing were carried out by the Center for Medical Genomics at Indiana University School of Medicine. Other bioinformatics analyses were performed by the Collaborative Core for Cancer Bioinformatics (C^3B) shared by Indiana University Simon Comprehensive Cancer Center and Purdue University Center for Cancer Research.

2.6. Primary mouse hepatocyte isolation and adenovirus transduction

Primary hepatocytes were isolated from C57BL/6J mice using a two-step collagenase procedure followed by Percoll gradient centrifugation to separate hepatocytes from non-parenchymal cells, as previously described [33, 34]. Cell viability (>80%) was assessed by trypan blue staining exclusion. Cells were seeded at a density of $6x10^5$ cells per well in 6-well plates, and incubated in a humidified 5% CO_2 incubator at 37°C, in DMEM containing 5 mM glucose, 10% FBS and 100 IU/ml penicillin/100 µg/ml streptomycin, 1 nM dexamethasone. Cells were allowed to attach for 4 hours, and medium was then replaced with fresh medium. Primary hepatocytes were infected with an adenovirus expressing human SREBP-1c [34], rat glucokinase [35] or a control vector without expression cassette (Ad.Null) [35]. Cells were washed twice with PBS prior to harvesting.

2.7. Western blotting

Cells were lysed in RIPA buffer (Thermo Scientific, Waltham, MA) containing protease and phosphatase inhibitors (Roche, Indianapolis, IN), as described previously [33, 34]. Protein concentration was determined by BCA protein assay (Thermo Scientific). Proteins were separated in SDS-polyacrylamide Criterion gels (Bio-Rad, Hercules, CA) and transferred to 0.2-um polyvinylidene difluoride membranes (Bio-Rad). Primary antibodies were used in overnight incubations at 4°C, and secondary antibodies were added for 1 hour at room temperature. Antibodies were purchased from Cell Signaling Technology (Danvers, MA) [H2AK5ac (#2576); H2A (#2578); H2BK5ac (#2574); H2B (#8135); H3K9ac (#9649); H3 (#4499); H4K8ac (#2594);

H4 (#2935)], Thermo Scientific [SREBP-1 (MS-1207)], and Santa Cruz Biotechnology (Dallas, TX) [glucokinase (sc-7908)]. Blots were developed with ECL Western Blotting Substrate (Pierce, Rockford, lL) and exposed to ECL film (GE Healthcare, Piscataway, NJ). Bands were quantified by densitometry using ImageJ v1.48s, and results were normalized to control protein, as specified in the figure legends.

2.8. Statistical analysis

A *P* value of less than 0.05 was considered statistically significant. The Pearson correlation coefficient is shown for associations between high-throughput sequencing data sets. Data in Figure 1 are presented as the arithmetic mean ± standard deviation, and *P* values were calculated using unpaired two-tailed Student's *t*-tests. Boxplot analysis was carried out with SPSS v.26.

3. Results

3.1. Glycolysis flux influences global levels of histone acetylation at specific residues

Intermediates of cell metabolism are used as cofactors for histone-modifying enzymes to link metabolic information with transcriptional control of gene programs [4]. To investigate the connection between changes in glycolysis and histone acetylation, we overexpressed the first enzyme in the glycolysis pathway, glucokinase. Previous studies have shown that increasing glycolysis flux through glucokinase overexpression activates the entire glycolysis and lipogenesis program, including L-pyruvate kinase (*Pklr*), fatty acid synthase (*Fasn*), and stearoyl-CoA desaturase 1 (*Scd1*) [36]. Glucokinase expression in primary mouse hepatocytes was sufficient to increase acetylation of H3 at Lys9 (H3K9ac), while other histone lysine residues remained unchanged (Fig. 1A). Likewise, increased acetylation of H3K9 and H2BK5 (but not of H2AK5 and H4K8), was observed in primary mouse hepatocytes overexpressing the transcription factor Sterol Regulatory Element Binding Protein 1c (SREBP1c) (Fig. 1B), a master regulator of the glycolysis and lipogenesis pathways in the fed state [37]. These data provide evidence that H3K9 acetylation is responsive to changes in signals originating from metabolic pathways.

3.2. High-fat diet feeding induces changes in H3K9ac liver profiles

We then questioned whether genome-wide H3K9ac liver profiles would be altered by diet. Mice were fed a high-fat (HFD, 60% kcal fat) or a control diet (CD, 18% kcal fat) for 10 weeks. Sonicated chromatin from liver tissue was immunoprecipitated with an antibody to H3K9ac, and DNA was subjected to massive parallel sequencing. A total of 57,571 peaks were identified after filtering. Principal component analysis of peak density reads (signal) showed a high level of correlation among the three biological replicates of each group (Supplementary Fig. 2). Given the connection between H3K9ac and gene activity [13-16], we first assessed whether RNAPII position correlates with H3K9ac peaks in CD liver. RNAPII ChIP-seq data identified a

total of 20,525 peaks in liver tissue. Of these, 4,312 peaks were located in the promoter-proximal region (3,693 genes), and 15,009 peaks (5,368 genes) were in gene bodies. RNAPII peaks in promoter-proximal regions strongly correlated with H3K9ac peaks (73.2%), while in gene bodies this percentage was lower (36.2%) (Fig. 2A). Furthermore, the signals of H3K9ac peaks overlapping with RNAPII in the promoter-proximal region were stronger than those that were not associated with RNAPII (Fig. 2B). Thus, presence of RNAPII in the promoter-proximal region is associated with H3K9ac, in agreement with previous reports on the role of this histone mark in regulating transcription elongation and gene expression [13-16].

We then focused on differences between diets. A total of 1,383 H3K9ac peaks showed significantly different signals between diets (False Discovery Rate <0.05), of which 495 increased and 888 decreased in the HFD relative to the CD group (Fig. 3A). Gene enrichment in molecular networks was evaluated with DAVID [38]. KEGG Pathway Analysis showed enrichment in categories relevant to the underlying pathology, including 'metabolic pathways', 'drug/xenobiotic metabolism-P450', 'glutathione metabolism', and 'circadian rhythms' (Fig. 3B, 3C). Several of these genes, including *Igfbp1* (Insulin Like Growth Factor Binding Protein 1), *Serpina12* (Serpin Family A Member 12), *Clock* (Circadian Locomotor Output Cycles Kaput), *Rgs16* (Regulator of G-protein signaling 16), and *Lpin1* (Lipin 1), have previously been associated with metabolic syndrome and NAFLD [39-43]. Furthermore, based on Gene Ontology, peaks with differential signals were located in genes within the functional categories 'transport', 'transcription from RNA pol II promoter', 'oxidation-reduction process', and 'lipid metabolic process' (Supplementary Fig. 2). For instance, the 'circadian rhythm' and 'negative regulation of transcription from RNAPII' include key transcription factors or co-factors that regulate metabolism, including *Foxp1*, *Zbtb16*, *Arntl*, *Clock*, *Per1/2* [44, 45]. Overall, H3K9ac peaks with significant signal differences in the HFD group were located in genes affecting metabolism of carbohydrates and lipids, and several of these genes were transcription factors or cofactors.

3.3. H3K9ac peaks are located in areas of open chromatin

H3K9ac peak distribution across the genome showed preference for promoter-proximal regions (~47%), followed by intronic (~33%) and distal intergenic regions (18%) (Fig. 4A). This distribution was expected, given that H3K9ac is associated with regions with regulatory function. Both groups displayed a similar distribution of peaks, which indicates that the high-fat diet does not induce massive changes in histone marks, only affecting specific loci.

H3K9ac is associated with transcription activity, and peaks were expected to overlap with areas of open chromatin. To identify these areas, ATAC-seq was carried out in control diet-fed livers. A total of 75,443 regions were identified (Fig. 4B and Supplementary Fig. 3). Approximately 20% of peaks were located in promoter-proximal regions, 41% in gene bodies, and 39% in intergenic areas. H3K9ac peaks whose centers fell on the promoter-proximal areas were strongly linked to open chromatin ($r=0.64$; $p<1.0E-304$) (Fig. 4C). H3K9ac peaks falling in non-promoter regions were also significantly associated with open chromatin, albeit the degree of correlation was lower ($r=0.42$; $p<1.0E-304$) (Fig. 4C).

3.4. Changes in H3K9ac peaks are positively associated with changes in transcript levels

To investigate the relationship between H3K9ac signal and gene expression levels transcript abundance was determined by RNA-seq (polyA). A total of 12,334 genes were found expressed in liver, and the two groups showed distinct expression profiles (Supplementary Fig. 4). Of all genes expressed, 87.1% (10,475 genes) had H3K9ac in their promoter-proximal regions. We found that 1,253 genes were differentially expressed (DE) between the HFD and the control diet (626 up, 627 down) (Fig. 5A), and changes in gene expression were linked to changes in H3K9ac signal (Fig. 5B), with a stronger association in promoter-proximal regions ($r=0.38$, $p<1.0E-304$) than gene bodies ($r=0.27$, $p=1.2E-110$).

Remarkably, even though the large majority of DE genes had H3K9ac peaks in the promoter-proximal region (1,059 out of 1,253 genes, or 84.5%), only 218 DE genes (17.4%) had

significant changes in H3K9ac signal in the promoter-proximal region (Fig. 6 and Fig. 7A, Supplementary Table 1). For these genes, there was a strong positive correlation between changes in mRNA and H3K9ac (r=0.83, p=7.8E-56) (Fig. 7B), implying that this histone modification has a role in regulating transcription. Histone acetylation marks on promoter-proximal regions are predictive of RNA Polymerase II (RNAPII) recruitment and elongation [46, 47]. H3K9ac has a characteristic bimodal distribution, with a peak upstream of the TSS, a second peak downstream from the TSS, and depletion on the TSS [15, 46, 48, 49]. To gather more specific information around the TSS, we generated footprints within 5 kb windows of the TSS. In liver, H3K9ac signal intensity showed a bimodal distribution, with a higher peak downstream from the TSS (Fig. 7C). No differences between diets were discernible when all H3K9ac peaks were considered (i.e., peaks with and without significant signal changes between diet groups; Fig. 7C, top). However, when considering only H3K9ac peaks with signal changes in the HFD relative to the CD group, a clear shift in signal intensity was observed downstream of the TSS (Fig. 7C, middle for increased signal and bottom for decreased signal), suggesting that this modification may be important for RNAPII transcription elongation decisions. Furthermore, H3K9ac peak centers of the 218 DE genes displayed a dense distribution close to the TSS, with a median of -193.5 and +435.5 nt, for upstream and downstream location, respectively (Fig. 7D). About 82.1% of the H3K9ac peaks were located downstream from the TSS, while only 17.9% were in the upstream region.

3.5. Transcription factor binding motifs in open chromatin regions

GO Term analysis clustered the subset of DE genes associated with H3K9ac signal changes in four main categories: circadian rhythms, oxidation-reduction, cholesterol biosynthesis, and cell proliferation (Fig. 8A). We sought to determine whether this subgroup of DE genes was enriched for specific transcription factor binding sites. Open chromatin regions were mostly enriched with binding motifs for retinoic acid-related orphan receptors (RORα,

RORγ, RORγτ), and several members of the activator protein 1 (AP-1) family of transcription factors (FRA1, FRA2, FOS, ATF3). Binding sites for GATA factors, paired box (PAX), and Wilms Tumor (WT1), all of which have relevant roles to liver development and function, were also enriched (Fig. 8B). Consistent with these data, Ingenuity Pathway Analysis [50] predicted that FOS, RORα, RORγ, and ATF3 are Upstream Regulators driving expression of genes represented in these categories (Fig. 8C).

4. Discussion

Aberrant transcriptional control of gene expression is central to the pathophysiology of metabolic diseases. Hundreds of genes become dysregulated and their gene products are abnormally expressed, leading to cellular dysfunction. Ongoing studies are providing evidence that epigenetic changes contribute to metabolic disease development, and histone modifications are at the center stage, bringing specific factors to chromatin, modifying chromatin dynamics, and influencing RNAPII activity as well as overall transcriptional output. Nevertheless, the specific histone modifications that play a role, as well as the molecular mechanisms that connect these modifications with fluctuations on gene expression, are not well understood [4, 5]. Acetylation of histone 3 at lysine 9 (H3K9ac) is a mark of gene activity, and it has been shown to be tightly involved in modifying gene expression to regulate the cell cycle, proliferation, as well as apoptosis [51-53]. Furthermore, the imbalance of H3K9 acetylation and deacetylation is a contributor of liver tumorigenesis, and increases the risk of colorectal cancer due to aging [54-56]. Thus, changes in this histone modification may shed information on novel pathways linking transcription control with metabolic dysfunction.

Given the role of glycolysis in providing acetyl-CoA that is used as substrate for histone acetylation, as well as in regulating lysine acetylase/deacetylase activity [6-12], we questioned whether increasing this pathway would influence global levels of histone H3K9 acetylation. We provide proof that increasing glycolysis flux either by overexpression of glucokinase or SREBP-1, induces acetylation of H3K9. These data support the concept that metabolites serve as signals to connect metabolic status with transcription activity. Concurring with these data, we identified hundreds of loci with H3K9ac changes in a mouse model of diet-induced obesity and insulin resistance. KEGG and Gene Ontology analysis showed that peaks were located in genes relevant to the pathophysiology induced by a high fat diet, primarily with a role in metabolism.

In agreement with its known role in regulating transcription, most H3K9ac peaks (>46%) were located in the promoter-proximal region. RNAPII located in this area strongly correlated with H3K9ac (>70% of peaks overlapping with H3K9ac), consistent with a role of H3K9ac in the regulation of transcription. Nonetheless, a significant fraction (~33%) of H3K9ac peaks were located in gene bodies. RNAPII overlapped to a lesser extent with intragenic H3K9ac. These data suggest that H3K9ac intragenic peaks may have factor-recruitment function, in addition to influence transcription of gene-embedded miRNA, or enhancer RNA levels (a measure of the activity of enhancer elements), which altogether have an impact on gene expression. In fact, recent studies in rodents have shown diet-induced alterations in H3K4me1, a histone mark linked to enhancer activity [57]. Regardless of their location in the genome, and consistent with a regulatory function, H3K9ac peaks overlapped with regions of open chromatin. Furthermore, promoter-proximal H3K9ac peaks were more strongly correlated with gene expression than those located in the gene body, in agreement with the link between this histone modification and transcription activity.

The RNAPII transcription cycle includes pre-initiation complex formation and initiation, elongation, and termination. RNAPII is recruited to promoters and initiates transcription, stalling after having transcribed approximately 20–60 nucleotides. RNAPII pausing limits the frequency of transcription initiation, a mechanism that has been termed ‘pause-initiation limit’ [58]. Recent data have shown that the mRNA level of genes does not necessarily correlate with the rate of RNAPII recruitment to the promoter, but instead, is a function of the rate of RNAPII elongation and degradation [59-61]. The transition from initiation to elongation is a rate-limiting step, requiring specific signals to release RNAPII from its paused state and engage in transcription elongation by recruitment and activation of the Super Elongation Complex (SEC). Acetylation of H3K9 is necessary for recruitment of the SEC to phosphorylate and activate RNAPII, and for the transition from transcription pause-release to transcript elongation [16]. Previous studies have shown that only a fraction of genes responding to environmental cues, do so by regulating

transcription elongation via H3K9ac and RNAPII pause-release [16, 17]. Furthermore, variances of acetylation marks located downstream of the TSS are more important for RNAPII elongation and mRNA prediction than those upstream [47]. Of all differentially expressed genes of our study, only ~17% have changes in H3K9ac signal in the promoter-proximal region, and approximately 80% of these H3K9ac peaks were located downstream from the TSS. Thus, in response to a high-fat diet, it is likely that changes in expression of a fraction of genes occurs by controlling the rate of elongation through H3K9ac and RNAPII pause-release.

Gene Ontology analysis showed that this group of genes is involved in circadian rhythm regulation, oxidation-reduction reactions, cholesterol transport and biosynthesis, and cellular proliferation. Not surprisingly, motif analysis in the promoter-proximal region showed strong binding site enrichment for retinoic acid-related orphan receptors (ROR), central regulators of inflammation, circadian rhythms and metabolism homeostasis [62, 63]. RORs directly regulate the circadian clock gene *Bmal1* [64], which in turn, controls expression of downstream targets in part through histone acetylation. In fact, CLOCK has intrinsic acetylase activity and has been reported to acetylate histone H3 at lysines 9 and 14, while its partner BMAL1, enhances the histone acetyl transferase activity [65]. Binding site enrichment also included the AP-1 (activator protein 1) family of transcription factors (including FRA1, FRA2, FOS, and ATF3), crucial regulators of inflammation in liver [66, 67], hepatic detoxification (in particular through glutathione S-transferases [68]), lipid metabolism [69], and cellular proliferation [70]. The overall level of expression of the 218 genes identified in the present study is likely to be the result of a combination of changes in transcription factor recruitment to the promoter and transcript initiation, as well as H3K9ac-mediated regulation of transcription elongation, and future studies will be necessary to assess the contribution of either mechanism to dysregulation of gene expression.

It should be noted that the changes in gene expression in the HFD group are likely to be triggered by the combination of high fat (60% kcal), sucrose (7% kcal) and cholesterol (from lard

and casein), all of which are macronutrients present in the diet. The two diets used in this study were not matched, and therefore, some changes in gene expression could also be due to differences such as fiber content, or phytoestrogens from plant sources present in the chow diet and absent in the purified high-fat diet. Yet, the majority of gene expression changes in the HFD group affected metabolic pathways that could be predicted based on the macronutrient content of the diet.

Recent studies analyzing transcriptome profiles in NAFLD/NASH patients and control individuals have identified transcriptional alterations in functional categories such as insulin signaling, fatty acid and cholesterol metabolism, bile metabolism, and inflammatory signaling [71-74]. These pathways were also significantly affected in the liver of mice fed a HFD in our study. Although no mouse model recapitulates exactly the human condition, and there are differences between the mouse and the human NAFLD transcriptome [75, 76], the metabolic pathways affected are similar, and it is likely that the molecular mechanisms that regulate them are conserved. Thus, future H3K9ac analyses in samples from NAFLD patients will indicate whether parallel observations to those seen in mice take place in human liver.

In conclusion, our data suggest that H3K9ac is strongly linked to changes in gene expression of a fraction of genes that are differentially expressed by exposure to a HFD. Given that the majority of H3K9ac peaks are located downstream from the TSS, this group of genes may be regulated by a common mechanism of control at the transcription elongation step. Future studies will be necessary to identify the specific mechanism/s, which are likely to involve RNAPII stalling/elongation.

Acknowledgments

We thank the staff of the Center for Medical Genomics at Indiana University School of Medicine for high-throughput sequencing analysis, and Dr. Yongyong Hou for technical help. This work was supported by a Biomedical Research Grant (Indiana University School of Medicine), an Indiana University Collaborative Research Grant, and by the Indiana Clinical and Translational Sciences Institute, funded in part by grant #UL1TR002529 from the National Institutes of Health, National Center for Advancing Translational Sciences, Clinical and Translational Sciences Award. High-fat diet studies were partially supported by National Institutes of Health grant #K01DK102492 to AL. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contribution statement

Authors have contributed to the study as follows: NM, conceptualization and design of the study; XC, YW, AC, AL, XCD, NM, acquisition of data; SL, AC, AL, YL, YW, JW, NM, analysis and interpretation of data; NM, drafting the article and intellectual content; all authors, final approval of the version.

Declaration of competing interest

All authors have read and approved the manuscript.

Figure legends

Figure 1. Increasing glycolysis flux enhances H3K9 acetylation in primary mouse hepatocytes. (A) Primary hepatocytes were transduced with an adenoviral vector expressing glucokinase (Ad.GK) or with a control vector without expression cassette (Ad.Null) at MOI 30, and harvested 36 hours later. Increased H3K9ac, but not other histone lysines, was observed; **(B)** Primary hepatocytes were transduced with an adenovirus expressing the mature (nuclear) form of SREBP1c (Ad.SR1) or Ad.Null at MOI 3, and harvested 72 hours later. Global levels of H3K9ac and H2BK5ac increased. *$p<0.05$, n=3.

Figure 2. RNAPII peaks correlate with H3K9ac in liver. (A) RNAPII [24] and H3K9ac ChIP-seq data in the CD group were combined. The overlapping region in the Venn diagram represents the number of peaks in RNAPII data set overlapping with H3K9ac in the CD group. Of all RNAPII peaks located in the promoter-proximal region, 73.2% overlapped with H3K9ac peaks, while 36.2% of RNAPII located in gene bodies correlated with H3K9ac peaks; **(B)** Signal distribution among H3K9ac peaks overlapping versus not overlapping with RNAPII.

Figure 3. Enriched pathway terms for H3K9ac peaks with differential signal between high-fat and control diet-fed. (A) Percentage of H3K9ac peaks with and without significant changes between diets; **(B)** KEGG pathway analysis of differentially bound H3K9ac peaks, FDR<0.05 (n=3); **(C)** Selected examples of genes enriched in KEGG categories; C: control diet, HF: high-fat diet.

Figure 4. H3K9ac peaks overlap with open chromatin regions. (A) H3K9ac distribution of all identified peaks in the CD and HFD groups; **(B)** ATAC-seq was carried out using two biological replicates from the control diet group. Distribution of open regions across the genome; **(C)** Strong correlation between open chromatin regions and H3K9ac peaks in promoter-proximal

areas (-3 kb to +2kb from TSS), and non-promoter regions (gene bodies and intergenic regions).

Figure 5. Correlation between mRNA and H3K9ac data sets. (A) Differential expression (DE) between CD and HFD groups based on RNA-seq data (n=4), FDR<0.05; **(B)** Changes in gene expression are moderately associated with changes in H3K9ac signal in promoter-proximal regions and gene bodies.

Figure 6. Gene expression, H3K9ac, and chromatin accessibility. Genome browser tracks displaying changes in gene expression (top) and H3K9ac (middle) in HFD vs CD groups. Open chromatin areas (ATAC-seq) are shown at the bottom. Promoter-proximal regions are highlighted in yellow. **(A)** *Npas2* (Neuronal PAS Domain Protein 2). The changes in H3K9ac signal did not affect the neighboring gene *Rpl31*, indicating that the changes in H3K9ac induced by a high-fat diet affected only select loci; **(B)** *Per3* (Period Circadian Clock 3).

Figure 7. H3K9ac peak signal distribution around the TSS. (A) Venn diagram showing overlap of peaks in the promoter-proximal region with DE genes; **(B)** Correlation between differentially expressed genes and differential H3K9ac signal in the promoter-proximal region (DE and H3K9ac signal changes at FDR<0.05); **(C)** Top: Overall peak distribution for CD and HFD; Middle: Genes with higher H3K9ac in the high-fat diet group; Bottom: Genes with lower H3K9ac in the high-fat diet group; **(D)** Boxplot of H3K9ac distribution upstream and downstream from the TSS for H3K9ac peaks associated with DE genes.

Figure 8. Transcription factor motif enrichment in promoter-proximal regions. (A) GO Terms (p<0.01) of DE genes with H3K9ac peaks with significantly different signal between the HFD and CD groups; as a reference p=0.05 is shown; **(B)** Transcription factor motifs enriched in

the promoter regions of DE genes that also had H3K9ac peaks with significantly different signal between the HFD and CD groups, and which overlapped with regions of open chromatin; q-value (Benjamini) <0.05; **(C)** Genes targeted by Upstream Regulators (IPA analysis).

References

[1] Z.M. Younossi, A.B. Koenig, D. Abdelatif, Y. Fazel, L. Henry, M. Wymer, Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes, Hepatology, 64 (2016) 73-84.

[2] S. Lallukka, H. Yki-Jarvinen, Non-alcoholic fatty liver disease and risk of type 2 diabetes, Best Pract Res Clin Endocrinol Metab, 30 (2016) 385-395.

[3] G. Targher, C.P. Day, E. Bonora, Risk of cardiovascular disease in patients with nonalcoholic fatty liver disease, N Engl J Med, 363 (2010) 1341-1350.

[4] S.T. Keating, A. El-Osta, Epigenetics and metabolism, Circulation research, 116 (2015) 715-736.

[5] D.E. Schones, A. Leung, R. Natarajan, Chromatin Modifications Associated With Diabetes and Obesity, Arterioscler Thromb Vasc Biol, 35 (2015) 1557-1561.

[6] K.E. Wellen, G. Hatzivassiliou, U.M. Sachdeva, T.V. Bui, J.R. Cross, C.B. Thompson, ATP-citrate lyase links cellular metabolism to histone acetylation, Science, 324 (2009) 1076-1080.

[7] L. Cai, B.M. Sutter, B. Li, B.P. Tu, Acetyl-CoA induces cell growth and proliferation by promoting the acetylation of histones at growth genes, Mol Cell, 42 (2011) 426-437.

[8] A. Moussaieff, M. Rouleau, D. Kitsberg, M. Cohen, G. Levy, D. Barasch, A. Nemirovski, S. Shen-Orr, I. Laevsky, M. Amit, D. Bomze, B. Elena-Herrmann, T. Scherf, M. Nissim-Rafinia, S. Kempa, J. Itskovitz-Eldor, E. Meshorer, D. Aberdam, Y. Nahmias, Glycolysis-mediated changes in acetyl-CoA and histone acetylation control the early differentiation of embryonic stem cells, Cell Metab, 21 (2015) 392-402.

[9] H. Takahashi, J.M. McCaffery, R.A. Irizarry, J.D. Boeke, Nucleocytosolic acetyl-coenzyme a synthetase is required for histone acetylation and global transcription, Mol Cell, 23 (2006) 207-217.

[10] S. Zhao, A. Torres, R.A. Henry, S. Trefely, M. Wallace, J.V. Lee, A. Carrer, A. Sengupta, S.L. Campbell, Y.M. Kuo, A.J. Frey, N. Meurs, J.M. Viola, I.A. Blair, A.M. Weljie, C.M. Metallo,

N.W. Snyder, A.J. Andrews, K.E. Wellen, ATP-Citrate Lyase Controls a Glucose-to-Acetate Metabolic Switch, Cell Rep, 17 (2016) 1037-1052.

[11] M. Thangaraju, E. Gopal, P.M. Martin, S. Ananth, S.B. Smith, P.D. Prasad, E. Sterneck, V. Ganapathy, SLC5A8 triggers tumor cell apoptosis through pyruvate-dependent inhibition of histone deacetylases, Cancer Res, 66 (2006) 11560-11564.

[12] T. Latham, L. Mackay, D. Sproul, M. Karim, J. Culley, D.J. Harrison, L. Hayward, P. Langridge-Smith, N. Gilbert, B.H. Ramsahoye, Lactate, a product of glycolytic metabolism, inhibits histone deacetylase activity and promotes changes in gene expression, Nucleic Acids Res, 40 (2012) 4794-4803.

[13] G. Liang, J.C. Lin, V. Wei, C. Yoo, J.C. Cheng, C.T. Nguyen, D.J. Weisenberger, G. Egger, D. Takai, F.A. Gonzales, P.A. Jones, Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome, Proc Natl Acad Sci U S A, 101 (2004) 7357-7362.

[14] B.E. Bernstein, M. Kamal, K. Lindblad-Toh, S. Bekiranov, D.K. Bailey, D.J. Huebert, S. McMahon, E.K. Karlsson, E.J. Kulbokas, 3rd, T.R. Gingeras, S.L. Schreiber, E.S. Lander, Genomic maps and comparative analysis of histone modifications in human and mouse, Cell, 120 (2005) 169-181.

[15] A. Kratz, E. Arner, R. Saito, A. Kubosaki, J. Kawai, H. Suzuki, P. Carninci, T. Arakawa, M. Tomita, Y. Hayashizaki, C.O. Daub, Core promoter structure and genomic context reflect histone 3 lysine 9 acetylation patterns, BMC genomics, 11 (2010) 257.

[16] L.A. Gates, J. Shi, A.D. Rohira, Q. Feng, B. Zhu, M.T. Bedford, C.A. Sagum, S.Y. Jung, J. Qin, M.J. Tsai, S.Y. Tsai, W. Li, C.E. Foulds, B.W. O'Malley, Acetylation on histone H3 lysine 9 mediates a switch from transcription initiation to elongation, J Biol Chem, 292 (2017) 14456-14472.

[17] J.P. Etchegaray, L. Zhong, C. Li, T. Henriques, E. Ablondi, T. Nakadai, C. Van Rechem, C. Ferrer, K.N. Ross, J.E. Choi, A. Samarakkody, F. Ji, A. Chang, R.I. Sadreyev, S. Ramaswamy,

S. Nechaev, J.R. Whetstine, R.G. Roeder, K. Adelman, A. Goren, R. Mostoslavsky, The Histone Deacetylase SIRT6 Restrains Transcription Elongation via Promoter-Proximal Pausing, Mol Cell, 75 (2019) 683-699 e687.

[18] M.R. Marasco, A.M. Conteh, C.A. Reissaus, J.E.t. Cupit, E.M. Appleman, R.G. Mirmira, A.K. Linnemann, Interleukin-6 Reduces beta-Cell Oxidative Stress by Linking Autophagy With the Antioxidant Response, Diabetes, 67 (2018) 1576-1588.

[19] X. Pan, Y. Zhang, H.G. Kim, S. Liangpunsakul, X.C. Dong, FOXO transcription factors protect against the diet-induced fatty liver disease, Scientific reports, 7 (2017) 44597.

[20] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, Bioinformatics, 26 (2010) 841-842.

[21] Y. Liao, G.K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, Bioinformatics, 30 (2014) 923-930.

[22] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics, 26 (2010) 139-140.

[23] D.J. McCarthy, Y. Chen, G.K. Smyth, Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation, Nucleic Acids Res, 40 (2012) 4288-4297.

[24] A. Kosters, D. Sun, H. Wu, F. Tian, J.C. Felix, W. Li, S.J. Karpen, Sexually dimorphic genome-wide binding of retinoid X receptor alpha (RXRalpha) determines male-female differences in the expression of hepatic lipid processing genes in mice, PloS one, 8 (2013) e71538.

[25] W.J. Kent, C.W. Sugnet, T.S. Furey, K.M. Roskin, T.H. Pringle, A.M. Zahler, D. Haussler, The human genome browser at UCSC, Genome Res, 12 (2002) 996-1006.

[26] R.C. Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, (2020).

[27] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T.R. Gingeras, STAR: ultrafast universal RNA-seq aligner, Bioinformatics, 29 (2013) 15-21.

[28] M.R. Breese, Y. Liu, NGSUtils: a software suite for analyzing and manipulating next-generation sequencing datasets, Bioinformatics, 29 (2013) 494-496.

[29] J. Piper, M.C. Elze, P. Cauchy, P.N. Cockerill, C. Bonifer, S. Ott, Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data, Nucleic Acids Res, 41 (2013) e201.

[30] F. Ramirez, F. Dundar, S. Diehl, B.A. Gruning, T. Manke, deepTools: a flexible platform for exploring deep-sequencing data, Nucleic Acids Res, 42 (2014) W187-191.

[31] S. Heinz, C. Benner, N. Spann, E. Bertolino, Y.C. Lin, P. Laslo, J.X. Cheng, C. Murre, H. Singh, C.K. Glass, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities, Mol Cell, 38 (2010) 576-589.

[32] W.J. Kent, A.S. Zweig, G. Barber, A.S. Hinrichs, D. Karolchik, BigWig and BigBed: enabling browsing of large distributed datasets, Bioinformatics, 26 (2010) 2204-2207.

[33] S. Surendran, V.N. Jideonwo, C. Merchun, M. Ahn, J. Murray, J. Ryan, K.W. Dunn, J. Kota, N. Morral, Gene targets of mouse miR-709: regulation of distinct pools, Scientific reports, 6 (2016) 18958.

[34] V. Jideonwo, Y. Hou, M. Ahn, S. Surendran, N. Morral, Impact of silencing hepatic SREBP-1 on insulin signaling, PloS one, 13 (2018) e0196704.

[35] N. Morral, R. McEvoy, H. Dong, M. Meseck, J. Altomonte, S. Thung, S.L. Woo, Adenovirus-mediated expression of glucokinase in the liver as an adjuvant treatment for type 1 diabetes, Hum Gene Ther, 13 (2002) 1561-1570.

[36] N. Morral, H.J. Edenberg, S.R. Witting, J. Altomonte, T. Chu, M. Brown, Effects of glucose metabolism on the regulation of genes of fatty acid synthesis and triglyceride secretion in the liver, J Lipid Res, 48 (2007) 1499-1510.

[37] R. Ruiz, V. Jideonwo, M. Ahn, S. Surendran, V.S. Tagliabracci, Y. Hou, A. Gamble, J. Kerner, J.M. Irimia-Dominguez, M.A. Puchowicz, A. Depaoli-Roach, C. Hoppel, P. Roach, N. Morral, Sterol Regulatory Element-binding Protein-1 (SREBP-1) Is Required to Regulate

Glycogen Synthesis and Gluconeogenic Gene Expression in Mouse Liver, J Biol Chem, 289 (2014) 5510-5517.

[38] W. Huang da, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, Nat Protoc, 4 (2009) 44-57.

[39] J.T. Heiker, N. Kloting, P. Kovacs, E.B. Kuettner, N. Strater, S. Schultz, M. Kern, M. Stumvoll, M. Bluher, A.G. Beck-Sickinger, Vaspin inhibits kallikrein 7 by serpin mechanism, Cell Mol Life Sci, 70 (2013) 2569-2583.

[40] T. Reinehr, M. Kleber, A.M. Toschke, J. Woelfle, C.L. Roth, Longitudinal association between IGFBP-1 levels and parameters of the metabolic syndrome in obese children before and after weight loss, Int J Pediatr Obes, 6 (2011) 236-243.

[41] K. Kajimoto, E. Suemitsu, Y. Sato, Y. Sakurai, H. Harashima, Liver-Specific Silencing of Lipin1 Reduces Fat Mass as Well as Hepatic Triglyceride Biosynthesis in Mice, Biol Pharm Bull, 39 (2016) 1653-1661.

[42] A. Shetty, J.W. Hsu, P.P. Manka, W.K. Syn, Role of the Circadian Clock in the Metabolic Syndrome and Nonalcoholic Fatty Liver Disease, Dig Dis Sci, (2018).

[43] V. Pashkov, J. Huang, V.K. Parameswara, W. Kedzierski, D.M. Kurrasch, G.G. Tall, V. Esser, R.D. Gerard, K. Uyeda, H.C. Towle, T.M. Wilkie, Regulator of G protein signaling (RGS16) inhibits hepatic fatty acid oxidation in a carbohydrate response element-binding protein (ChREBP)-dependent manner, J Biol Chem, 286 (2011) 15116-15125.

[44] B. Marcheva, K.M. Ramsey, E.D. Buhr, Y. Kobayashi, H. Su, C.H. Ko, G. Ivanova, C. Omura, S. Mo, M.H. Vitaterna, J.P. Lopez, L.H. Philipson, C.A. Bradfield, S.D. Crosby, L. JeBailey, X. Wang, J.S. Takahashi, J. Bass, Disruption of the clock components CLOCK and BMAL1 leads to hypoinsulinaemia and diabetes, Nature, 466 (2010) 627-631.

[45] S. Panda, Circadian physiology of metabolism, Science, 354 (2016) 1008-1015.

[46] Z. Wang, C. Zang, J.A. Rosenfeld, D.E. Schones, A. Barski, S. Cuddapah, K. Cui, T.Y. Roh, W. Peng, M.Q. Zhang, K. Zhao, Combinatorial patterns of histone acetylations and methylations in the human genome, Nat Genet, 40 (2008) 897-903.

[47] Y. Chen, M. Jorgensen, R. Kolde, X. Zhao, B. Parker, E. Valen, J. Wen, A. Sandelin, Prediction of RNA Polymerase II recruitment, elongation and stalling from histone modification data, BMC genomics, 12 (2011) 544.

[48] T.Y. Roh, S. Cuddapah, K. Zhao, Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping, Genes Dev, 19 (2005) 542-552.

[49] C.M. Koch, R.M. Andrews, P. Flicek, S.C. Dillon, U. Karaoz, G.K. Clelland, S. Wilcox, D.M. Beare, J.C. Fowler, P. Couttet, K.D. James, G.C. Lefebvre, A.W. Bruce, O.M. Dovey, P.D. Ellis, P. Dhami, C.F. Langford, Z. Weng, E. Birney, N.P. Carter, D. Vetrie, I. Dunham, The landscape of histone modifications across 1% of the human genome in five human cell lines, Genome Res, 17 (2007) 691-707.

[50] A. Kramer, J. Green, J. Pollard, Jr., S. Tugendreich, Causal analysis approaches in Ingenuity Pathway Analysis, Bioinformatics, 30 (2014) 523-530.

[51] M. Choudhury, P.H. Park, D. Jackson, S.D. Shukla, Evidence for the role of oxidative stress in the acetylation of histone H3 by ethanol in rat hepatocytes, Alcohol, 44 (2010) 531-540.

[52] R. Ding, Y. Jin, X. Liu, Z. Zhu, Y. Zhang, T. Wang, Y. Xu, H3K9 acetylation change patterns in rats after exposure to traffic-related air pollution, Environ Toxicol Pharmacol, 42 (2016) 170-175.

[53] D. Sancak, S. Ozden, Global histone modifications in Fumonisin B1 exposure in rat kidney epithelial cells, Toxicol In Vitro, 29 (2015) 1809-1815.

[54] H. Ji, Y. Zhou, X. Zhuang, Y. Zhu, Z. Wu, Y. Lu, S. Li, Y. Zeng, Q.R. Lu, Y. Huo, Y. Shi, H. Bu, HDAC3 Deficiency Promotes Liver Cancer through a Defect in H3K9ac/H3K9me3 Transition, Cancer Res, 79 (2019) 3676-3688.

[55] P. Tian, Y. Zhu, C. Zhang, X. Guo, P. Zhang, H. Xue, Ras-ERK1/2 signaling contributes to the development of colorectal cancer via regulating H3K9ac, BMC Cancer, 18 (2018) 1286.

[56] I.J. Goossens-Beumer, A. Benard, A.Q. van Hoesel, E.C. Zeestraten, H. Putter, S. Bohringer, G.J. Liefers, H. Morreau, C.J. van de Velde, P.J. Kuppen, Age-dependent clinical prognostic value of histone modifications in colorectal cancer, Transl Res, 165 (2015) 578-588.

[57] A. Leung, B.W. Parks, J. Du, C. Trac, R. Setten, Y. Chen, K. Brown, A.J. Lusis, R. Natarajan, D.E. Schones, Open chromatin profiling in mice livers reveals unique chromatin variations induced by high fat diet, J Biol Chem, 289 (2014) 23557-23567.

[58] S. Gressel, B. Schwalb, P. Cramer, The pause-initiation limit restricts transcription activation in human cells, Nat Commun, 10 (2019) 3603.

[59] L. Core, K. Adelman, Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation, Genes Dev, 33 (2019) 960-982.

[60] M.G. Guenther, S.S. Levine, L.A. Boyer, R. Jaenisch, R.A. Young, A chromatin landmark and transcription initiation at most promoters in human cells, Cell, 130 (2007) 77-88.

[61] L.J. Core, J.J. Waterfall, J.T. Lis, Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters, Science, 322 (2008) 1845-1848.

[62] T. Wada, H.S. Kang, M. Angers, H. Gong, S. Bhatia, S. Khadem, S. Ren, E. Ellis, S.C. Strom, A.M. Jetten, W. Xie, Identification of oxysterol 7alpha-hydroxylase (Cyp7b1) as a novel retinoid-related orphan receptor alpha (RORalpha) (NR1F1) target gene and a functional cross-talk between RORalpha and liver X receptor (NR1H3), Mol Pharmacol, 73 (2008) 891-899.

[63] T. Wada, H.S. Kang, A.M. Jetten, W. Xie, The emerging role of nuclear receptor RORalpha and its crosstalk with LXR in xeno- and endobiotic gene regulation, Exp Biol Med (Maywood), 233 (2008) 1191-1201.

[64] J.S. Takahashi, Transcriptional architecture of the mammalian circadian clock, Nat Rev Genet, 18 (2017) 164-179.

[65] M. Doi, J. Hirayama, P. Sassone-Corsi, Circadian regulator CLOCK is a histone acetyltransferase, Cell, 125 (2006) 497-508.

[66] Q. Zhu, H. Wang, B. Jiang, X. Ni, L. Jiang, C. Li, X. Wang, F. Zhang, B. Ke, L. Lu, Loss of ATF3 exacerbates liver damage through the activation of mTOR/p70S6K/ HIF-1alpha signaling pathway in liver inflammatory injury, Cell Death Dis, 9 (2018) 910.

[67] L. Bakiri, R. Hamacher, O. Grana, A. Guio-Carrion, R. Campos-Olivas, L. Martinez, H.P. Dienes, M.K. Thomsen, S.C. Hasenfuss, E.F. Wagner, Liver carcinogenesis by FOS-dependent inflammation and cholesterol dysregulation, J Exp Med, 214 (2017) 1387-1409.

[68] S.C. Hasenfuss, L. Bakiri, M.K. Thomsen, R. Hamacher, E.F. Wagner, Activator Protein 1 transcription factor Fos-related antigen 1 (Fra-1) is dispensable for murine liver fibrosis, but modulates xenobiotic metabolism, Hepatology, 59 (2014) 261-273.

[69] S.C. Hasenfuss, L. Bakiri, M.K. Thomsen, E.G. Williams, J. Auwerx, E.F. Wagner, Regulation of steatohepatitis and PPARgamma signaling by distinct AP-1 dimers, Cell Metab, 19 (2014) 84-95.

[70] X. Jiang, H. Xie, Y. Dou, J. Yuan, D. Zeng, S. Xiao, Expression and function of FRA1 protein in tumors, Mol Biol Rep, 47 (2020) 737-752.

[71] M.P. Suppli, K.T.G. Rigbolt, S.S. Veidal, S. Heeboll, P.L. Eriksen, M. Demant, J.I. Bagger, J.C. Nielsen, D. Oro, S.W. Thrane, A. Lund, C. Strandberg, M.J. Konig, T. Vilsboll, N. Vrang, K.L. Thomsen, H. Gronbaek, J. Jelsing, H.H. Hansen, F.K. Knop, Hepatic transcriptome signatures in patients with varying degrees of nonalcoholic fatty liver disease compared with healthy normal-weight individuals, Am J Physiol Gastrointest Liver Physiol, 316 (2019) G462-G472.

[72] Z.M. Younossi, F. Gorreta, J.P. Ong, K. Schlauch, L. Del Giacco, H. Elariny, A. Van Meter, A. Younoszai, Z. Goodman, A. Baranova, A. Christensen, G. Grant, V. Chandhoke, Hepatic gene expression in patients with obesity-related non-alcoholic steatohepatitis, Liver international : official journal of the International Association for the Study of the Liver, 25 (2005) 760-771.

[73] B.M. Arendt, E.M. Comelli, D.W. Ma, W. Lou, A. Teterina, T. Kim, S.K. Fung, D.K. Wong, I. McGilvray, S.E. Fischer, J.P. Allard, Altered hepatic gene expression in nonalcoholic fatty liver disease is associated with lower hepatic n-3 and n-6 polyunsaturated fatty acids, Hepatology, 61 (2015) 1565-1578.

[74] M. Ryaboshapkina, M. Hammar, Human hepatic gene expression signature of non-alcoholic fatty liver disease progression, a meta-analysis, Scientific reports, 7 (2017) 12361.

[75] A. Teufel, T. Itzel, W. Erhart, M. Brosch, X.Y. Wang, Y.O. Kim, W. von Schonfels, A. Herrmann, S. Bruckner, F. Stickel, J.F. Dufour, T. Chavakis, C. Hellerbrand, R. Spang, T. Maass, T. Becker, S. Schreiber, C. Schafmayer, D. Schuppan, J. Hampe, Comparison of Gene Expression Patterns Between Mouse Models of Nonalcoholic Fatty Liver Disease and Liver Tissues From Patients, Gastroenterology, 151 (2016) 513-525 e510.

[76] C. Jiang, P. Li, X. Ruan, Y. Ma, K. Kawai, H. Suemizu, H. Cao, Comparative Transcriptomics Analyses in Livers of Mice, Humans, and Humanized Mice Define Human-Specific Gene Networks, Cells, 9 (2020).

Declaration of interests

☐**X** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Author contribution statement

Authors have contributed to the study as follows: NM, conceptualization and design of the study; XC, YW, AC, AL, XCD, NM, acquisition of data; SL, AC, AL, YL, YW, JW, NM, analysis and interpretation of data; NM, drafting the article and intellectual content; all authors, final approval of the version.

Highlights

- Mechanisms leading to gene dysregulation in fatty liver disease are not understood
- Histone H3K9 acetylation overlaps with presence of RNA polymerase II in the liver
- In high-fat diet feeding, changes in gene expression are linked to H3K9 acetylation
- Dysregulated gene expression may be attributable to transcription elongation