G-quadruplex forming regions in *GCK* and *TM6SF2* are targets for differential DNA methylation in metabolic disease and hepatocellular carcinoma patients

Angelika Lahnsteiner[1, 2*], Victoria Ellmer[1], Anna Oberlercher[1], Zita Liutkeviciute[1], Esther Schönauer[3], Bernhard Paulweber[4], Elmar Aigner[4,5], Angela Risch[1,2]

[1]Division of Cancer (Epi-)Genetics, Department of Biosciences and Medical Biology, Center for Tumor Biology and Immunology (CTBI), Paris Lodron University Salzburg, Austria

[2]Cancer Cluster Salzburg, Austria

[3]Division of Structural Biology, Department of Biosciences and Medical Biology, Paris Lodron University Salzburg, Austria

[4]First Department of Medicine, University Clinic Salzburg, Austria

[5]Paracelsus Medical University Salzburg, Austria

*Corresponding author

angelika.lahnsteiner@plus.ac.at

University of Salzburg

Hellbrunnerstraße 34

5020 Salzburg

Austria

Abstract

The alarming increase in global rates of metabolic diseases (MetDs) and their association with cancer risk renders them a considerable burden on our society. The interplay of environmental and genetic factors in causing MetDs may be reflected in DNA methylation patterns, particularly at non-canonical (non-B) DNA structures, such as G-quadruplexes (G4s) or R-loops. To gain insight into the mechanisms of MetD progression, we focused on DNA methylation and functional analyses on intragenic regions of two MetD risk genes, the glucokinase (*GCK*) exon 7 and the transmembrane 6 superfamily 2 (*TM6SF2*) intron 2-exon 3 boundary, which harbor non-B DNA motifs for G4s and R-loops.

Pyrosequencing of 148 blood samples from a nested cohort study revealed significant differential methylation in *GCK* and *TM6SF2* in MetD patients versus healthy controls. Furthermore, these regions harbor hypervariable and differentially methylated CpGs also in hepatocellular carcinoma versus normal tissue samples from The Cancer Genome Atlas (TCGA). Permanganate/S1 nuclease footprinting with direct adapter ligation (PDAL-Seq), native polyacrylamide DNA gel electrophoresis and circular dichroism (CD) spectroscopy revealed the formation of G4 structures in these regions and demonstrated that their topology and stability is affected by DNA methylation. Detailed analyses including histone marks, chromatin conformation capture data, and luciferase reporter assays, highlighted the cell-type specific regulatory function of the target regions. Based on our analyses, we hypothesize that changes in DNA methylation lead to topological changes, especially in *GCK* exon 7, and cause the activation of alternative regulatory elements or potentially play a role in alternative splicing.

Our analyses provide a new view on the mechanisms underlying the progression of MetDs and their link to hepatocellular carcinomas, unveiling non-B DNA structures as important key players already in early disease stages.

Key words: Metabolic disease, hepatocellular carcinoma, DNA methylation, G-quadruplex, PDAL-Seq, alternative promoter

52 **1. Introduction**

53 Overweight and obesity represent a major health issue of the 21st century. Their worldwide rate
54 tripled from 1975 to 2016[1], affecting approximately 39% of the global population[2]. Alarmingly,
55 childhood obesity also increased from 0.7-0.9% in 1975 to 5.6-7.8% in 2016[3]. Caloric imbalance causing
56 obesity is related to dietary habits, physical inactivity, stress, sleep patterns, social and educational
57 status, and endocrine disorders[4]. Excessive fat accumulation increases the risk of developing metabolic
58 diseases (MetDs) such as the metabolic syndrome (MetS), type 2 diabetes (T2D), metabolic
59 dysfunction-associated fatty liver disease (MAFLD)[5,6], and cardiovascular disease (CVD)[7]. Furthermore,
60 MetD patients have an elevated risk for gastric[8], pancreatic[9], hepatocellular[10], colorectal[11], breast[12]
61 and lung cancer[13,14].

62 MAFLD is the most common form of chronic liver disease in industrialized societies[15] with a
63 worldwide prevalence of 25%[16], and is strongly associated with T2D[17]. One hallmark of MAFLD is the
64 development of large cytoplasmic lipid droplets in hepatocytes with the accumulation of proteins
65 involved in lipid homeostasis on their surface. The key disease mechanism in MAFLD is very likely
66 related to peripheral insulin resistance[17,18]. In addition, fatty acid β-oxidation is impaired in MAFLD
67 owing to mitochondrial respiratory chain dysfunction, which, in combination with the accumulation of
68 incompletely metabolized fatty acids, contributes to oxidative and lipotoxic stress in hepatocytes. This
69 facilitates the progression from simple steatosis to advanced stages of MAFLD which can further
70 progress to liver fibrosis, cirrhosis, and to hepatocellular carcinoma (HCC)[19]. Currently, we do not fully
71 understand the molecular mechanisms leading to the development and progression of MAFLD, but
72 they clearly involve a complex interplay between genetic predisposition and environmental factors
73 influencing epigenetic processes, such as DNA methylation at cytosines (5-methylcytosine, 5mC).
74 Studies investigating 5mC have identified differentially methylated positions (DMPs) associated with
75 hepatic fat or key signaling pathways regulating glycolysis and *de novo* lipogenesis[20-24]. However, a
76 mechanistic explanation for the impact of those detected DMPs on cellular processes is often lacking.
77 Particularly, if these DMPs are located in intergenic regions or in gene bodies without overlapping
78 regulatory elements, interpreting their biological functions remains difficult.

79 New insights arise from the observation that G-rich sequences, with the ability to form non-
80 canonical (non-B) DNA structures called G-quadruplexes (G4s), are sites of methylation instability[25],
81 and actively influence DNA methylation[26]. G4s are formed via the consensus motif $G_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}$
82 and can adopt intra- or intermolecular planar G4 tetrads held together by Hoogsteen hydrogen bonds
83 with a 1-7 bp long linker of any base (N) and stabilized by a monovalent cation, *in vivo* typically
84 potassium[27-29]. These structures are found in regulatory regions such as promoters[30,31], enhancers[32], or
85 near splice sites[33], whereas they are less often detected in exons and on the template strand[34]. They
86 often co-occur with RNA:DNA hybrids (R-loops), formed during the transcription of long non-coding

87 (lncRNAs) on the displaced strand and each stabilizing the other[35]. G4s act as binding hubs for several
88 transcription factors (TFs)[36,37] and DNA methyltransferases (DNMTs)[26,38,39], which is underlined by their
89 manifold biological roles extensively reviewed elsewhere[40-42].

90 G4s are fairly sensitive to changes in ion concentrations like potassium[43], the pH[44], or glucose
91 concentrations[45]. Moreover, DNA methylation influences the thermodynamic properties of the DNA
92 duplex [46] and therefore, methylation of CpGs within a G4-forming sequence (G4-CpGs) possibly triggers
93 its rearrangement from a DNA duplex to a non-B DNA structure. This is supported by the observation
94 of enhanced G4 formation due to DNA methylation within the first exon of the human reverse
95 transcriptase (hTERT) gene, which in turn, leads to repressed gene transcription [47]. Furthermore, DNA
96 methylation leads to increased stability of G4 structures in the *BCL2*[48], *VEGF*[49] and *FMR1*[50], whereas
97 decreased stability was detected for the *MEST* G4[51] in *in vitro* assays. Even changes in the G4 topology
98 from parallel to antiparallel in *c-KIT,* or a complete loss of G4 formation in *HRAS* can be caused by DNA
99 methylation[52]. Genome-wide approaches revealed that methylation of CpGs within G4s (G4-CpGs) is
100 significantly decreased compared to CpGs outside of G4s, especially among those CpGs located in CpG
101 islands (CGIs)[26,53]. Furthermore, it has been shown, that DNMT1 preferentially binds to G4 structures
102 rather than to hemi-methylated B-DNA, which results in the inhibition of the DNMT1 and local
103 hypomethylation at the G4 structure[26]. There are still contradictory discussions, since the mentioned
104 studies mostly have shown that G4-CpGs are targets of hypomethylation, while a recent work
105 investigating DNA methylation of aging clocks suggests a general DNA methylation instability at G4-
106 CpGs, rather than a particular hypo- or hypermethylation of these G4-CpGs[25]. Furthermore, whereas
107 studies investigating individual G4 motifs have shown that DNA methylation can also stabilize G4s,
108 genome-wide analysis demonstrated that the stability of G4s with a maximum distance of 50
109 nucleotides to the CGI is associated with hypomethylation and in cases where the G4 containing CGI is
110 located in closed and hypermethylated chromatin, the G4 stability was lowest[54]. Most analyses focused
111 on promoters and on regions overlapping CGIs, but a considerable number of regulatory regions, like
112 alternative promoters[55,56] or enhancers[57], occur outside of CGIs where different regulatory
113 mechanisms probably apply. Selective DNA methylation loss in CG-poor regions can make chromatin
114 accessible for alternative transcription, particularly in HCC[58], and we have previously associated
115 differential methylation with alterations in promoter activation[59]. This indicates that we are still in the
116 early stages of understanding the mechanisms involved, and that different genetic locations probably
117 underlie distinct mechanisms. Therefore, alterations in the G4 landscape could alter transcription
118 factor binding and activate alternative regulatory elements or alternative splicing events, causing
119 aberrant gene expression patterns that contribute to MAFLD and may even promote the development
120 of more severe diseases such as hepatocellular carcinoma.

121 Patients suffering from metabolic diseases are at a higher risk of developing cancer during their
122 lifetime. Therefore, it is necessary to understand which processes are involved and DNA methylation
123 at G4s could be one of these mechanisms, probably already arising in a noncancerous but dysregulated
124 metabolic stage. Hence, we aimed to analyze DNA methylation and G4 formation in MetD patients and
125 in HCC in two intragenic regions outside of CGIs with annotated regulatory elements, as well as the
126 occurrence of putative G4 forming motifs and lncRNAs in the key metabolism genes glucokinase (*GCK*)
127 and transmembrane 6 superfamily 2 (*TM6SF2*). *GCK* deficiency has been linked with T2D[60] and
128 neonatal diabetes[61], while a common *GCK* SNP has been associated with hyperglycemia[62] and several
129 alternative *GCK* promoters are annotated for the gene body and differential methylation of the *GCK*
130 CGI spanning exons 9-10 had been linked with CVD[63]. The intron 2-exon 3 boundary in *TM6SF2* overlaps
131 with a prognostic lncRNA for cancer[64,65], it shows many different transcripts and rs58542926 C>T, the
132 SNP prognostic for MAFLD[66] and HCC[67], is located downstream of the annotated regulatory region.
133 rs58542926 C>T is routinely checked in the clinics and potentially could have an impact on the
134 methylation of the next regulatory element.
135 Herein, we report differential DNA methylation in exon 7 of *GCK* and within the intron 2-exon
136 3 boundary of *TM6SF2* in blood samples from a MetD patient cohort in G4 forming regions, which are
137 involved in the activation of alternative regulatory elements and possibly also in alternative splicing.
138 These new insights emphasize the critical need to understand the interplay between aberrant DNA
139 methylation and non-B DNA structure formation occurring already at an early disease stage.
140

141 ## 2. Results

142 **Cohort demographics and description**

143 To investigate DNA methylation in putative intragenic G4-forming regions in *GCK* and *TM6SF2*,
144 we performed locus-specific 5mC analysis on the Qiagen Pyromark Q24 Platform for 148 blood donors
145 which were classified as lean-healthy (LH), obese-healthy (OH), lean-MAFLD (LM), obese-MAFLD (OM),
146 and advanced MetD (aMetD, including MAFLD and poorly controlled T2DM patients) according to their
147 metabolic parameters, as described in the *Methods* section. The mean values of the clinical parameters
148 are provided in Table 1 and Supplement Fig. 1. The age of our study cohort ranged from 42-77 years,
149 albeit the OM and aMetD groups showed significantly elevated age compared to the LH group (LH vs.
150 OM p=.0152, LH vs aMetD p= 1.53e-05, Student´s t-test, and Supplement Table 1). This trend was not
151 unexpected since MAFLD and T2D are known as aging-related diseases occurring at a higher frequency
152 in the elderly (reviewed in[68]). Fasting blood glucose, insulin, and aspartate transferase (AST) levels were
153 significantly increased in OH, OM and aMetD, but not in LM compared to LH (Students t-test,
154 Supplement Table 1). No significant change was detected in total cholesterol levels in any group
155 compared to healthy controls. High-density lipoprotein (HDL) was significantly decreased in OM and
156 aMetD, and low-density lipoprotein (LDL) was only decreased in aMetD, whereas total triglycerides
157 were increased in LM, OM, and aMetD. The detailed p-values are shown in the Supplement Table 1.

158 **Table 1. Study characteristics.** Data are expressed as means ± standard deviation (SD)
159

	lean-healthy (LH) n=24		obese-healthy (OH) n=31		lean-MAFLD (LM) n=27		obese-MAFLD (OM) n=32		advanced MetD (aMetD) n=34	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Gender [f/m]	14/10 (58% f)		21/10 (68% f)		12/15 (44% f)		10/22 (31% f)		15/19 (44% f)	
Age [y]	52	8	53	6	54	4	57	3	63	8
BMI [kg/m2]	21.9	2.0	33.1	3.1	23.5	1.3	33.4	3.4	30.1	3.9
HbA_{1c} [%]	5.3	0.2	5.5	0.3	5.4	0.3	5.8	0.5	7.3	1.4
HOMA	1.4	0.5	2.4	1.1	1.6	0.6	4.9	3.2	11.0	15.3
FIB4	1.2	0.4	0.9	0.2	1.1	0.2	1.1	0.3	3.0	1.6
Fast. Glucose [mg/dL]	89.6	7.0	93.5	6.9	91.3	6.3	105.3	16.1	166.0	54.2
Fast. Glucose [mmol/L]	5.0	0.4	5.2	0.4	5.1	0.3	5.8	0.9	9.2	3.0
Fast. Insulin [mU/L]	6.2	2.4	10.5	4.6	7.2	2.8	18.4	9.3	29.3	47.8
GGT [U/L]	15.6	6.5	22.3	12.1	28.3	20.5	55.5	71.6	210.1	219.6
AST [U/L]	20.4	5.5	22.0	5.5	23.7	5.6	29.3	12.7	60.2	41.7
ALT [U/L]	18.8	6.9	24.2	10.1	22.8	9.1	39.9	26.1	76.4	48.0
HDL-C [mg/dL]	69.8	15.0	63.9	19.3	62.7	16.5	52.1	15.8	48.6	13.7
LDL-C [mg/dL]	149.6	32.5	144.0	35.1	152.9	27.8	157.6	44.1	120.7	45.8
TG [mg/dL]	83.6	21.7	92.9	27.7	121.4	62.6	178.3	139.8	198.9	98.7
TChol [mg/dL]	216.8	38.8	211.0	35.6	221.3	28.7	222.3	45.2	195.2	50.5

160
161 BMI = Body mass index; HOMA = homeostasis model assessment; HbA_{1c} = Glycated haemoglobin A_{1c}; FIB4 = Fibrosis4 score; Fast. Glucose = Fasting glucose; Fast. Insulin =
162 Fasting insulin; GGT = γ-glutamyl transpeptidase; AST = aspartate aminotransferase; ALT = alanine aminotransferase; HDL-c = High density lipoprotein cholesterol; LDL-c = Low
163 density lipoprotein cholesterol; TG = Triglycerides; TChol = total cholesterol;

164 ***GCK* exon 7 and the *TM6SF2* intron 2-exon 3 boundary are differentially methylated in MetD**
165 **patients and are hypervariable in HCC**

166 *GCK* is encoded on the minus strand of chr. 7 and consists of ten exons (Fig. 1A). 14 different
167 transcripts are annotated for *GCK* in the ENSEMBL genome browser[69]. The investigated region is
168 located in exon 7 (chr7:44,187,339-44,187,380; hg19) and is overlapped by an antisense lncRNA
169 LOC105375258 (XR_927223.3) on the plus strand. Genome segmentation by automating chromatin
170 state discovery and characterization (ChromHMM) in the UCSC genome browser predicts a "weak
171 enhancer or open chromatin *cis* regulatory element" in hepatoblastoma cell line HepG2 and a
172 "promoter region including TSS" in the lymphoblast cell line K562 (Fig. 1A, yellow rectangle)[70].
173 Furthermore, several alternative promoters are found in intron 7 and 8, as well as an EGR1 TF binding
174 site (Fig. 1A, green rectangle)[70]. Significant hypomethylation was detected in CpG 1 and 2 in aMetD
175 patients, and in CpG 3 in LM, OM and aMetD compared to LH subjects. No changes were detected in
176 CpG 4 (Fig. 1A lower panel and Supplement Fig. 2A). In particular, a gradual reduction in DNA
177 methylation was detected for CpG 3 with the individual disease stages, and the advanced MetD
178 patients showing the lowest methylation.

179 In line with *GCK* exon 7, we found differential methylation also for the *TM6SF2* intron 2-exon
180 3 boundary (chr19:19,381,224-19,381,297; hg19, Fig. 1B). *TM6SF2* is encoded by ten exons on the
181 minus strand of chr. 19. In total, there are five transcripts annotated, of which two result in a retained
182 intron and one in a skipped exon overlapping our target region[69]. Notably, exon 3-5 are targets of
183 alternative splicing events and exon 3 is annotated as cassette exon (Supplement Fig. 3)[70].
184 Furthermore, *TM6SF2* overlaps with the sense lncRNA ENSG00000267629 (AC138430.4), sharing the
185 same transcription start site. This lncRNA represents a readthrough between *TM6SF2* and *HAPLN4*.
186 ChromHMM predicts a "weak enhancer or open chromatin *cis* regulatory element" in HepG2 and a
187 "promoter region including TSS" in K562 in the UCSC genome browser[70], as well as an EGR1 binding
188 site (Fig. 1B). 5mC analyses revealed that CpG 3-5 were significantly hypermethylated in the advanced
189 MetD group. In contrast, CpG 4 and 5 were significantly hypomethylated in OM compared to LH and
190 CpG 5 was also significantly hypomethylated in LM compared to LH (Student´s t-test, Fig. 1B lower
191 panel and Supplement Fig. 2B).
192

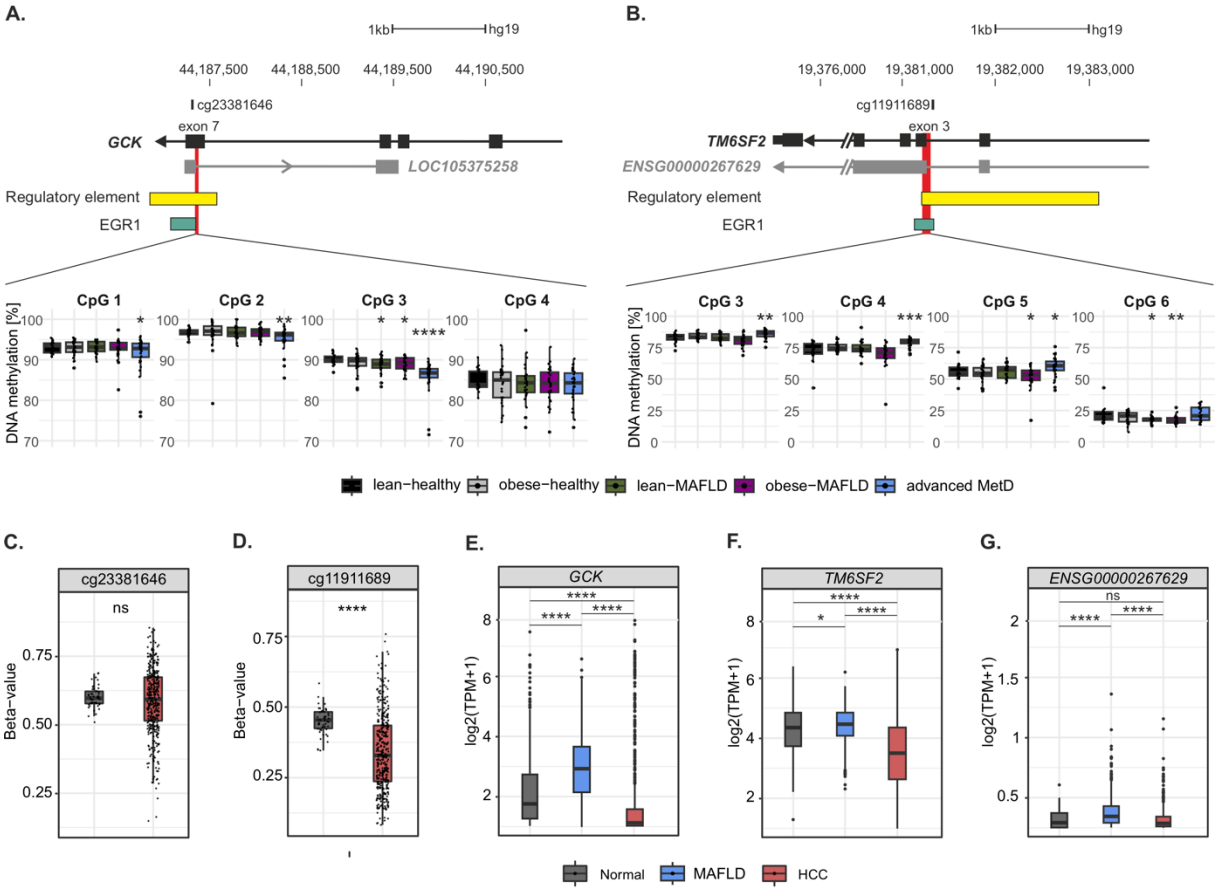


Figure 1. DNA methylation in *GCK* exon 7 and within the *TM6SF2* intron 2-exon 3 boundary. Detailed information about the differentially methylated regions in **A.** *GCK* and **B.** *TM6SF2* including the nearest annotated CpG positions on the Illumina Methylation array. The yellow rectangle indicates the genome segmentation by ChromHMM in UCSC genome browser[70]. The green rectangle represents the EGR1 binding motif. The pyrosequencing region is marked in red. The lower panel shows DNA methylation at the single CpG resolution by metabolic status in **A.** for *GCK* CpG 1-4 and **B.** *TM6SF2* CpG 3-6. DNA methylation data of the nearest annotated CpG on the Illumina Methylation array for **C.** *GCK* cg23381646 and **D.** *TM6SF2* cg11911689 are shown comparing HCC to normal tissue from TCGA[71]. Gene expression for **E.** *GCK*, **F.** *TM6SF2* and **G.** the *TM6SF2*-overlapping lncRNA ENSG00000267629 were obtained from the liver expression atlas (GepLiver)[72]. Pairwise comparison using Students t-test against LH (**A.-B.**) and Wilcoxon test (**C.-G.**), **p*<.05, ** *p*<.01, *** *p*<.001, **** *p*<.0001.

Given that alterations in DNA methylation are also associated with aging[73] and metabolic diseases occur at a higher frequency with increasing age, we analyzed the magnitude of age-relation compared to published aging markers located in *EDARADD* (cg09809672), *IPO8* (cg19722847), *NHLRC1* (cg22736354), *P2RXL1* (cg05442902) and *SCGN* (cg06493994). These five markers were selected based on their consistent identification in three distinct epigenetic clock studies utilizing Illumina Methylation arrays[73-75]. Consistent with published data, we detected a significant positive (*NHLRC1* and *SCGN*) and negative (*IPO8*, *EDARADD* and *P2RXL1*) association between DNA methylation and chronological age in our cohort (Supplement Fig. 4). In addition, the correlation analysis showed a weak but significant age-dependency for *GCK* CpG 3 and *TM6SF2* CpG 1, 3 and strongest for CpG 5, as well as for the mean DNA methylation (mean of CpG 1-6). Therefore, we restricted our analysis to 42–60-year-old subjects.

214 We still observed a significant decreased DNA methylation at *GCK* CpG 3 in all groups compared to
215 lean-healthy subjects, which reached significance even in obese-healthy patients, while this was not
216 the case when analyzing the entire age-spectrum (Supplement Fig. 2C). *TM6SF2* CpG 3, 4 and 6 where
217 still significantly differentially methylated, but the effect was lost in CpG 5 in 42–60-year-old subjects
218 (Supplement Fig. 2D). This points out that aberrant DNA methylation in these regions is only partly
219 caused by aging processes and underlies also a disease-related mechanism.

220 Next, we were curious if the SNP rs58542926 C>T, which causes a p.Glu167Lys (E167K)
221 substitution in *TM6SF2*, could influence the DNA methylation in this specific regulatory region. The
222 SNP is located 1,696 bp away from CpG 4 in exon 6. Although the risk associated variant T is more
223 common in the advanced MetD group (Supplement Fig. 5A), we found no significant difference in DNA
224 methylation by genotype in our study cohort (Supplement Fig. 5B). Therefore, we can exclude that the
225 genetic variant rs58542926 C>T causes the detected differential DNA methylation within *TM6SF2*
226 intron 2-exon 3.

227 Since metabolic diseases are associated with an increased risk for HCC[10], we investigated 5mC
228 of *GCK* and *TM6SF2* in data of *The Cancer Genome Atlas* (TCGA) from HCC compared to normal tissue[71].
229 The nearest annotated CpG for *GCK* covered by the Illumina Methylation array is cg23381646 located
230 45 bp upstream of CpG 3 (Fig. 1A). Although DNA methylation at cg23381646 was not significantly
231 different (Wilcoxon test, Benjamini-Hochberg (BH) adjusted *p*=.58, Fig. 1C), it appeared as highly
232 variable in tumor tissue with DNA methylation beta-values ranging from approximately 0.25-0.8.

233 For *TM6SF2*, TCGA data showed that the nearest annotated CpG cg11911689 (78 bp
234 downstream of CpG 4, Fig. 1B) was significantly hypomethylated in HCC compared to normal tissue
235 (Wilcoxon test, BH adjusted *p*=4.5E-10, Fig. 1D), and again showed a highly variable DNA methylation
236 range in cancer compared to normal control tissue.

237 To explore gene expression of the target genes, we analyzed *GCK* and *TM6SF2* expression in
238 MAFLD and HCC in the liver expression database (GepLiver)[72], since transcriptome data were not
239 available for the MetD cohort due to the lack of liver material. *GCK* expression was increased by 1.73-
240 fold (Wilcoxon test, BH adjusted *p*=2.52E-32) in MAFLD, but decreased expression was detected in HCC
241 compared to normal liver tissue (0.67-fold, Wilcoxon test, BH adjusted *p*=2.73E-32, Fig. 1E). Significant
242 increased expression was also detected for *TM6SF2* in MAFLD (1.04-fold, Wilcoxon test, BH adjusted
243 *p*=.035), which was again decreased in HCC by 0.76-fold (Wilcoxon test, BH adjusted *p*=7.48E-28, Fig.
244 1F). Of note, the expression of the *TM6SF2* overlapping lncRNA ENSG00000267629 was increased in
245 MAFLD by 1.92-fold (Wilcoxon test, BH adjusted *p*= 4.31E-7). Although it remained slightly
246 overexpressed in HCC, it did not reach significance (1.08-fold, Wilcoxon test, BH adjusted *p*= .58, Fig.
247 1G).

248 **Differentially methylated regions detected in *GCK* and *TM6SF2* are able to form non-B DNA**
249 **structures**

250 To investigate if the detected DMPs are located in non-B DNA structures and if they are formed
251 in living cells, we retrieved information about R-loop forming sequences (RLFS)[76] and putative G4
252 motifs (PQS) from pqsfinder[77]. The *GCK* DMR harbors several G4 motifs, but no predicted RLFS, albeit
253 this exon is overlapped by the annotated antisense lncRNA LOC105375258 (Fig. 2A left panel). All four
254 analyzed CpG positions in *GCK* exon 7 are located in a G4 motif in the non-template strand, with the
255 top differentially methylated CpG 3 located in the longest loop of the G4 between the G-run two and
256 three. Notably, the G4 includes a bulge, represented as an interruption in the G-run, between the third
257 and fourth G of G-run two (Sequence motif in Fig. 2A left panel). Furthermore, the downstream exons
258 8-10 harbor both RLFS and PQS (Fig. 2A left panel). DNA methylation of this downstream region was
259 investigated elsewhere and showed also differential methylation[63]. The *TM6SF2* DMR contained a RLFS
260 as well as PQS (Fig. 2A middle panel). The differentially methylated CpGs 4-6 were located directly
261 within the predicted G4 motif in the template strand and CpG 1-3 flanked the G4 motif. As negative
262 control region we have used an intronic region of *PARD3B* without annotated PQS or RLFS (Fig. 2A right
263 panel). *In vitro* G4 formation was verified by analyzing published G4-sequencing (G4-Seq) maps based
264 on a polymerase stop assay in a buffer containing the G4 stabilizing ligand pyridostatin (PDS) in
265 HEK293T cells (Fig. 2A, black track)[78]. Strong signal intensities in *GCK* and *TM6SF2,* but not in *PARD3B*,
266 confirmed the ability of the predicted motifs to form DNA secondary structures. As G4s are not
267 necessarily formed in living cells and their formation varies strongly in different cell types, we analyzed
268 published G4 ChIP-Seq data in HEK293 cells[79]. Especially, *GCK* shows enriched signal intensities
269 compared to its input control over the whole region spanning from exon 7 to exon 10, rather than
270 distinct peaks similar to the G4-Seq signals (Fig. 2A left panel). *TM6SF2* showed a more discrete peak
271 for replicate 1 but not for replicate 2 (Fig. 2A middle panel). Since, we could not infer a clear picture
272 for G4 formation from the antibody-based ChIP-Seq data, particularly for *GCK* exon 7, we aimed to
273 establish an alternative chemical method based on permanganate/S1 nuclease footprinting[80] which in
274 general detects regions of single-stranded DNA (ssDNA). We updated this method by linking it to direct
275 adapter ligation on streptavidin-coated beads, followed by an on-bead PCR published as PDAL-Seq
276 (permanganate/S1 nuclease footprinting with direct adapter ligation and sequencing)[81]. Since PDAL-
277 Seq is a footprinting method, we expected low or absent read frequencies directly at the non-B DNA
278 structure flanked by increased read frequencies at the non-B DNA borders. For HEK293 cells, we did
279 not detect a direct overlap of PDAL-footprints with G4-Seq peaks at *GCK* exon 7 rather than a slight
280 upstream shift, but in MCF7 cells the non-B-footprint was more obvious, highlighting the cell type
281 differences (Fig. 2A left panel, light blue highlighted box). Of note, footprints detected in *GCK* exon 8-
282 10 were more pronounced and overlapped well with the strong G4-Seq signals, while the peaks from

283 the antibody-based G4 ChIP-Seq method were less clear (Fig. 2A, left panel, light red highlighted box).
284 In contrast, the intronic region of *PARD3B* did not contain a G4 motif, nor did it form G4s *in vitro* or in
285 living cells (Fig. 2A right panel).

286 Next, we analyzed EGR1 ChIP-Seq data in K562 from ENCODE[82]. Unfortunately, there were no
287 data available in HEK293 or MCF7 cells. Strong EGR1 signals were detected co-occurring with the G4-
288 Seq peak and the footprinting valley of PDAL-Seq. Currently we do not know, if non-B DNAs are formed
289 in K562 cells. The negative G4 region in *PARD3B* showed also a EGR1 binding site indicating that EGR1
290 does not only target G4 regions.

291 To independently evaluate the G4-forming ability of the target regions, G4 folding was
292 confirmed by native DNA polyacrylamide (PAA) gel electrophoresis (Fig. 2B). The *GCK* and *TM6SF2*
293 target regions and two controls, *MYC* as a positive G4-folding region and an enhancer region of *ESR1*
294 as a G4-negative region, were PCR amplified and separated on a native 10% DNA PAA gel. Strikingly,
295 both *GCK* and *TM6SF2*, as well as *MYC* showed significant shifted bands after the G4 folding procedure
296 with potassium and heat induction (Fig. 2B, labeled with K), and even stronger after G4-stabilization
297 with PDS (Fig. 2B, labeled with PDS), compared to their unfolded controls only containing Tris-buffer
298 (Fig. 2B, labeled with U). This effect was not detected for the negative control region *ESR1*. In
299 agreement with the native gel electrophoresis, circular dichroism (CD) spectroscopy experiments (Fig.
300 2C-F) experiments showed G4 formation for all regions enhanced by increasing K^+ concentrations. *MYC*
301 (Fig. 2C and D) and *TM6SF2* (Fig. 2C and E) showed characteristic parallel G4 structures with a negative
302 peak at 240nm and a positive peak at 260nm. DNA methylation within *MYC* did not change the
303 conformation of the G4 nor did it lead to a loss of the formed structure. For *GCK* we noticed an
304 interesting pattern. While the increased K^+ concentrations led to parallel G4 formation visible due to a
305 shift in the positive peak from 279nm to 267nm, the methylated *GCK* oligo showed a broad positive
306 peak from 263-285nm, which resembled a hairpin structure (Fig. 2D and F). This conformational change
307 was probably caused by the methyl-groups which destabilize the G4, since there were six bases within
308 the bulge followed by one G belonging to G-run two and 16 bases in loop two, which are able to pair
309 (Fig. 2G). In all CD-spectra, the mutated oligos showed a clear shift of the positive peak to around
310 280nm, indicating the loss of G4 formation.
311

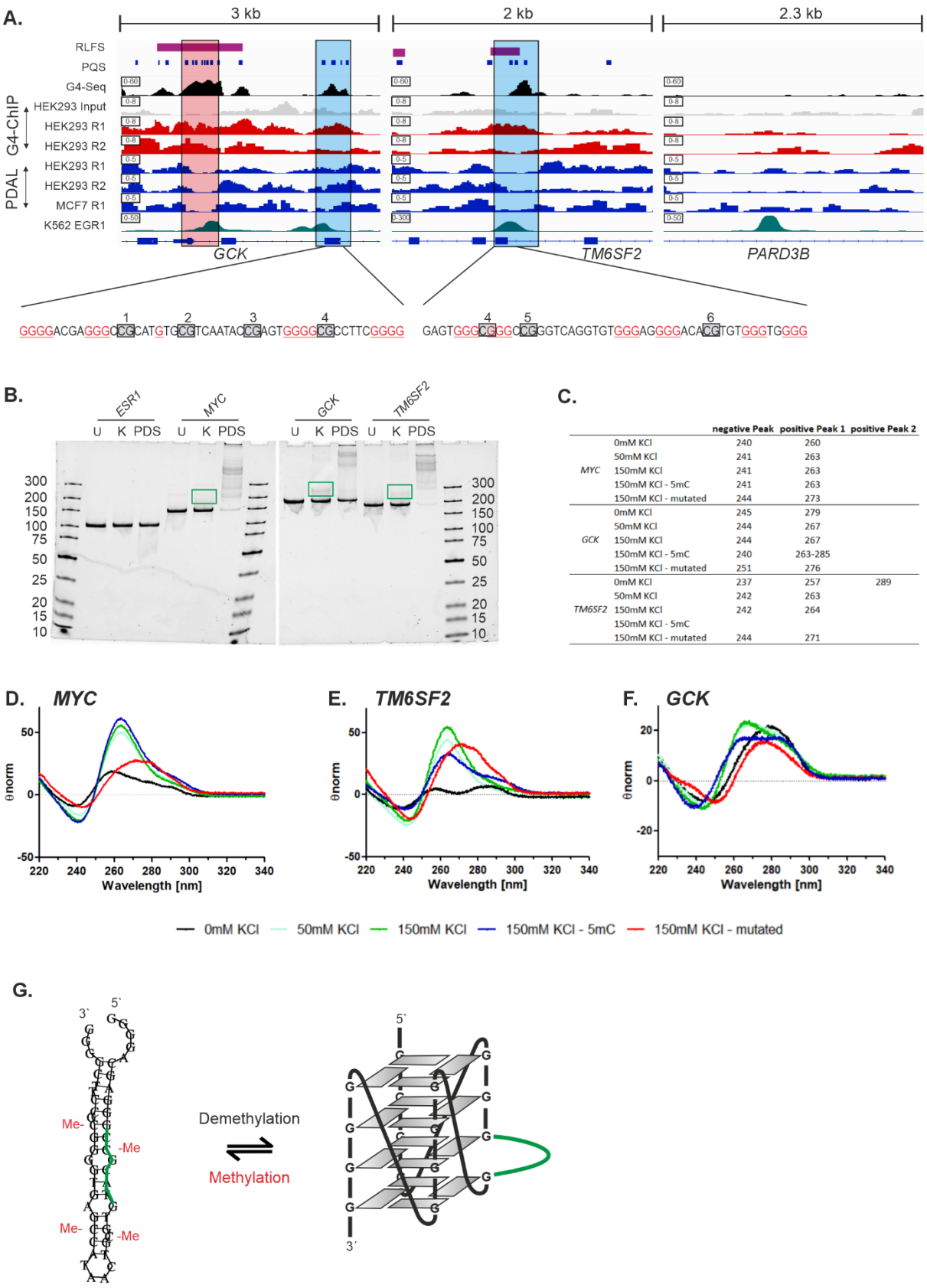


Figure 2. Non-B DNA formation in *GCK* and *TM6SF2*: A. Shows *GCK* (chr7: 44,184,901-44,187,919), *TM6SF2* (chr19: 19,380,358-19,382,362) and the non-G4 forming region in *PARD3B* (chr2:205,780,593-205,782,891). Data from different sources are shown in individual rows: Pink rectangles indicate R-loop forming sequences (RLFS) calculated by Quantitative Model of R-loop Forming Sequence (RLFS) finder (QmRLFS-finder)[76]. Blue rectangles represent putative G4 forming motifs (PQS) obtained from pqsfinder[77], Black track: *in vitro* G4 formation stabilized with pyridostatin (PDS) in HEK293T cells[78]; Grey

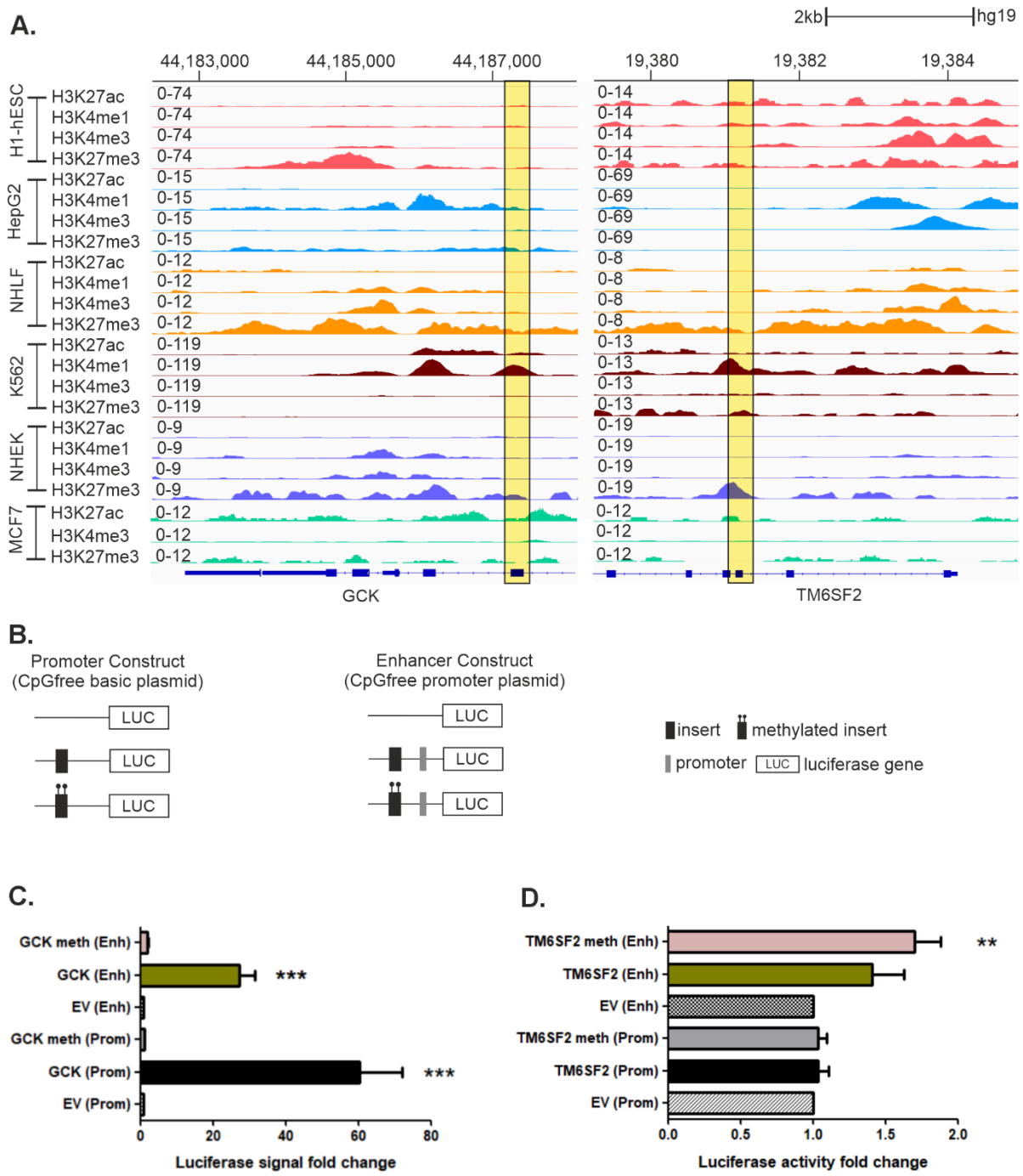
track: ChIP input control[79]; Red tracks: G4-ChIP-Seq tracks in HEK293 cells (2 replicates)[79]; Blue tracks: PDAL-Seq replicates 1 and 2 of HEK293 and MCF7 cells, Olive tracks: transcription factor EGR1 binding in K546 cells[82]. The sequences within *GCK* (antisense G4, sequence is given as reverse complement) and *TM6SF2* (sense G4) forming a G4 structure are detailed below. G-tracks are underlined, grey boxes indicate CpG positions. Blue highlighted areas: identified differentially methylated regions; light red highlighted area: a second regulatory region in *GCK,* which can form G4s. **B.** Native DNA PAA gels identify *in vitro* G4 formation in the G4 control region in *MYC*, *GCK* exon 7 and in the *TM6SF2* intron 2-exon 3 boundary, but not for *ESR1* (negative control). PCR products are loaded as unfolded (50 mM Tris-HCl pH 8.3) (U), folded with 150 mM KCl and heat induction (K), and folded with 150 mM KCl, heat induction and 10 µM PDS (PDS). **C.** positive and negative peak maxima obtained from CD spectra for *MYC*, *GCK* and *TM6SF2* with increasing K+ concentrations, for the methylated or mutated oligos. Individual CD spectra are shown for **D.** *MYC*, **E.** *TM6SF2* and **F.** *GCK*. **G.** Depicts the possible refolding of the methylated hairpin to a parallel G4 structure caused by demethylation at the *GCK* exon 7 region. The green loop indicates the bulge sequence which interrupts G-run two.

Differentially methylated regions in *GCK* and *TM6SF2* harbor regulatory functions and contact neighboring genes

To assess the regulatory potential of the identified DMPs located in G4-forming regions in *GCK* and *TM6SF2*, we analyzed the activating H3K27ac, H3K4me1, H3K4me3 and the repressing H3K27me3 histone marks in H1-hESC (human embryonic stem cell line, red), HepG2 (hepatoblastoma cell line, blue), NHLF (lung fibroblasts, orange), K562 (lymphoblasts, brown), NHEK (normal human epidermal keratinocytes, violet), and MCF7 (breast cancer, green) cell lines in IGV from the ENCODE project[82]. In *GCK* exon 7 H3K4me1 activating histone marks in K562 cells and to a lesser extent also in HepG2 cells were found, whereas NHLF and NHEK cells harbored the repressive H3K27me3 mark (Fig. 3A, left panel). MCF7 cells showed H3K27ac marks near the target region in exon 7. Furthermore, the downstream regions overlapping exon 8-10 showed cell-type specific marks as indicated by H3K4me1 in HepG2 and K562 cells, but H3K27me3 in H1-hESC, NHLF and NHEK cells. For the *TM6SF2* intron 2-exon 3 boundary (Fig. 3A, right panel) we found strong signals for the activating H3K4me1 histone modification in K562 cells, but H3K27me3 signals in NHEK and NHLF cells. HepG2 cells showed H3K4me1 and H3K4me3 marks in the promoter region of *TM6SF2*. In conclusion, the analysis of ENCODE histone modifications revealed cell type specific regulatory elements for both target regions.

To analyze the effect of the identified differentially methylated regions on gene expression, we performed *in vitro* luciferase reporter assays. Therefore, we cloned the regions of interest in two different CpG-free vectors upstream of the luciferase reporter gene and tested for their potential promoter and enhancer functions in a methylated and unmethylated state (Fig. 3B). We detected both, a significant promoter (60.3±11.3-fold, Welch two sample t-test, p=.0004,) and enhancer (27.4±3.9-fold, Welch two sample t-test, p=.0001, Fig. 3C) function for the *GCK* region compared to the empty vector, reflecting well the annotated ChromHMM function in HepG2 in the UCSC genome browser. However, when the same region was methylated prior to the analysis, both functions were lost. In

355 contrast, we did not find a promoter function for the DMR located in *TM6SF2*, but the methylated
356 construct showed a 1.7±0.2-fold (Welch Two Sample t-test *p*=.004, Fig. 3D) elevated enhancer
357 function.
358



359
360 **Figure 3. Regulatory roles of *GCK* and *TM6SF4* target regions. A.** Activating (H3K27ac, H3K4me1 and H3K4me3) and
361 repressing (H3K27me3) histone modifications obtained from ChIP-seq experiments from ENCODE for H1-hESC (stem cells,
362 red), HepG2 (hepatoblastoma, blue), NHLF (healthy lung, orange), K562 (blood, brown), NHEK (normal human epidermal
363 keratinocytes, violet) and MCF7 (breast cancer, green) cell lines [82] in *GCK* (left panel) and *TM6SF2* (right panel). For the hg19
364 genome assembly ENCODE did not provide H3K4me1 data for MCF7 cells. Yellow highlighted areas represent the
365 pyrosequencing regions. Here we showed a cell-type specific pattern of activating and repressing histone marks. **B.** The CpG-
366 free basic vector was used to test for a putative promoter function and the CpG-free promoter (containing a minimal
367 promoter) vector to test for a putative enhancer function of the differentially methylated regions in a luciferase gene reporter

368 assay in **C.** *GCK* exon 7 and **D.** in the *TM6SF2* intron 2-exon 3 boundary, in a methylated and unmethylated state compared
369 to the empty vector (EV). Pairwise comparison against empty vectors (Welch two sample t-test), *p<.05, **p<.01, ***p<.001.
370

371 To predict the potential neighboring targets of the identified *GCK* and *TM6SF2* regulatory
372 regions, we analyzed Hi-C data in HepG2 cells[82] and from HCC patient material[83] using the 3D Genome
373 Browser[84]. *GCK* is located in a topological associated domain (TAD) of approximately 500 kb. Strikingly,
374 within this TAD several genes are known for their association with metabolic diseases and cancer like
375 AE binding protein 1 (*AEBP1*)[85,86], YKT6 v-SNARE homolog (*YKT6*)[87,88], upregulator of cell proliferation
376 (*URGCP*)[89,90] and phosphoglycerate mutase 2 (*PGAM2;* for more details see Supplement Section 4)[91].
377 When comparing the interaction frequencies of HepG2 cells and HCC decreased signal intensities were
378 observed (Fig. 4A, comparison between TAD interaction frequency in HepG2-top and HCC-bottom).
379 By analyzing the potential contact points of the *GCK* DMP, we noticed that this region stays in contact
380 with an enhancer region in *GCK* intron 1 (~chr7:44,157,293-44,158,172; hg38) and a proximal enhancer
381 region between myosin light chain 7 (*MYL7*) and DNA polymerase delta 2 (*POLD2,* Fig. 4B). *TM6SF2* is
382 located in a TAD together with UPF1 RNA helicase and ATPase (*UPF1*) and neurocan (*NCAN*, Fig. 4C),
383 associated with MAFLD[92] and HCC[93]. The interaction analysis showed that the *TM6SF2* DMP contacts a
384 promoter-like region within intron 3 of hyaluronan and proteoglycan link protein 4 (*HPLN4*,
385 ~chr19:19,260,457, hg38) as well as intron 8 of SURP and G-patch domain containing 1 (*SUGP1*,
386 ~chr19:19,280,457, hg38), which was identified as regulator of cholesterol metabolism (Fig. 4D)[94].
387 In summary, both *in silico* data for histone marks and HiC data, as well as the luciferase reporter
388 assay provide evidence that *GCK* exon 7 and the *TM6SF2* intron 2-exon 3 boundary harbor important
389 regulatory elements, which are not only related to metabolic diseases, but also could play roles in
390 tumor formation and prognosis. Furthermore, our analysis points out that these regions contact
391 neighboring genes and that changes in their methylation patterns potentially also affect the
392 interactions with these elements. This strongly indicates that changes in DNA methylation at these
393 regulatory elements are not only associated with MAFLD but could also be linked to the enhanced
394 cancer risk in MetD patients corroborated also by the changes in the HiC data within these TADs.

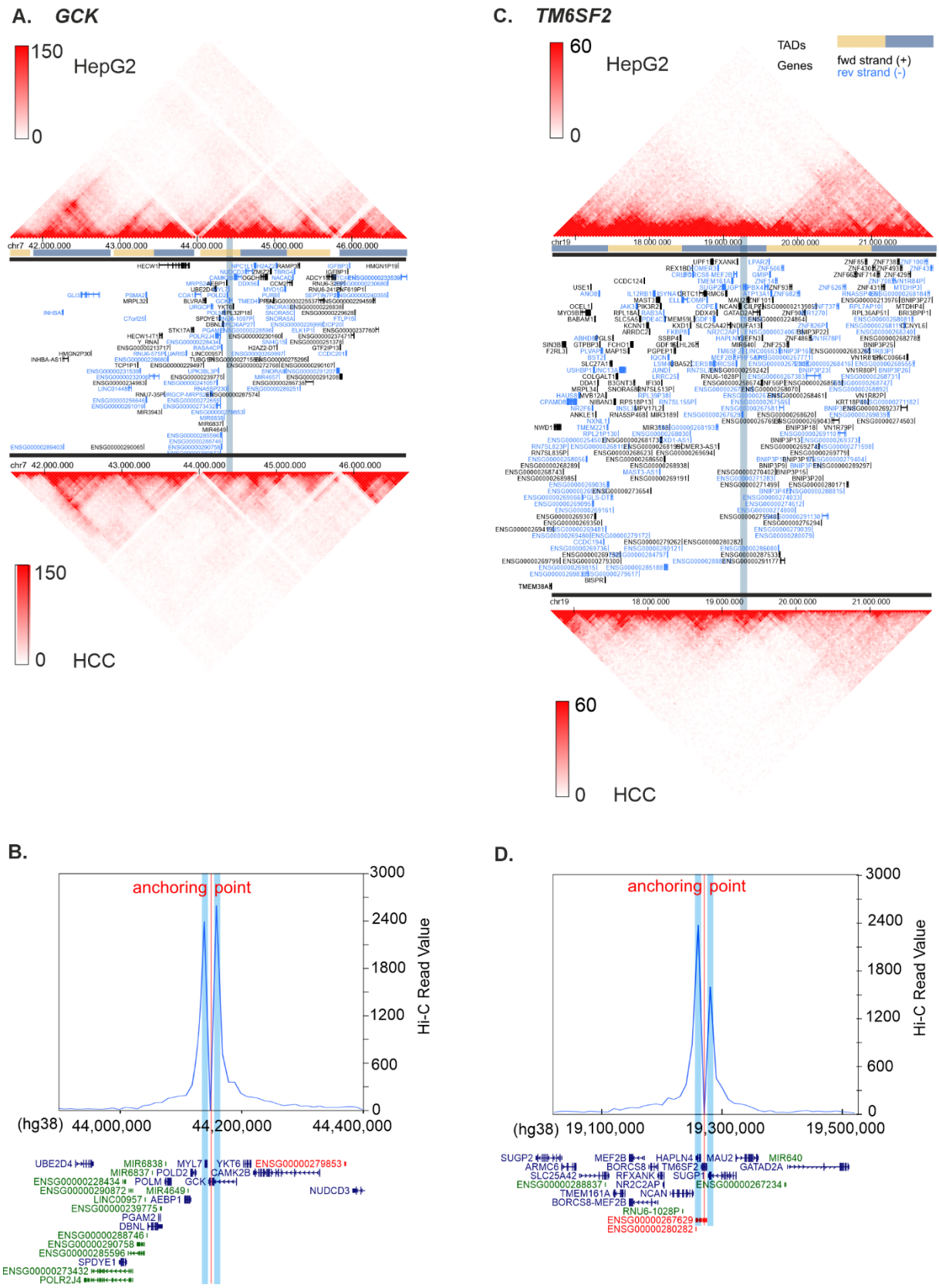


Figure 4. HiC data analysis in HepG2 cells and in HCC tumor material visualized in 3D genome browser[84]. A. TAD formation on chr. 7 around the *GCK* DMP in HepG2 (upper panel)[82] and HCC tumor tissue (lower panel)[83]. **B.** Analysis of potential contact points of the regulatory region (red vertical line) shows interactions (blue highlighted areas) with *GCK* intron 1 and an intergenic region between *MYL7* and *POLD2* near an enhancer region. **C.** TAD formation on chr. 19 around *TM6SF2* in HepG2 (upper panel)[82] and HCC tumor tissue (lower panel)[83]. **D.** Analysis of potential contact points of the regulatory region (red

401 vertical line) shows interactions (blue highlighted areas) with *SUGP1* and *HAPLN4*. Chromosome annotations are given in the
402 newer assembly hg38.
403
404

405 ## 3. Discussion

406 Metabolic diseases often arise due to a complex interplay of genetic and epigenetic processes.
407 Whereas the genetic risk is predefined by the base composition, the epigenetic landscape is flexible
408 and heavily influenced by the individuals´ environment. To assess epigenetic changes in two metabolic
409 disease associated genes, *GCK* and *TM6SF2,* we analyzed 5mC in 148 patients grouped in five different
410 categories: lean- and obese healthy subjects, lean and obese MAFLD, as well as patients suffering from
411 advanced MetDs. We focused on two intragenic regions, annotated as regulatory regions in
412 ChromHMM and predicted as non-B DNA forming sites (G4s and R-loops), which are not overlapping
413 with CGIs. Significant 5mC changes were detected for both target genes. While *GCK* exon 7 was
414 significantly hypomethylated especially at CpG 3 in all groups compared to lean healthy subjects (Fig.
415 1A), we observed a diverse pattern for the *TM6SF2* intron 2-exon 3 boundary. Lean- and obese MAFLD
416 subjects showed hypomethylation, whereas hypermethylation was detected in advanced MetD (Fig.
417 1B). This pattern likely mirrors the elevated blood glucose levels observed also as HbA_{1c}>7% within the
418 advanced MetD group, potentially leading to hypermethylation. Furthermore, it also could indicate a
419 change in blood cell type composition in this group, which also influences DNA methylation patterns[95].

420 Since increasing chronological age is one major risk factor for developing metabolic diseases
421 (reviewed in[68]), our findings of a significant age correlation with DNA methylation within *GCK* and
422 *TM6SF2* were expected. Nevertheless, after limiting the analysis to patients between 42-60-years of
423 age, significant differential methylation was still detected for both genes, pointing towards a combined
424 disease and aging mechanism. In addition, we were able to exclude an influence of the *TM6SF2* disease
425 risk SNP rs58542926 C>T on DNA methylation in this region.

426 Moreover, we confirmed our results in hepatocellular carcinoma versus adjacent normal tissue
427 obtained from *TCGA* data[71]. Although in contrast to cg11911689 (*TM6SF2,* Fig. 1D), differential
428 methylation was not significant in cg23381646 (*GCK,* Fig. 1C), both CpGs showed a highly variable 5mC
429 range, which is a sign for DNA methylation instability as previously observed, e.g. in Alzheimer's
430 disease[96]. Furthermore, DMPs in *TM6SF2* were hypermethylated in advanced MetD patients, whereas
431 the TCGA data from HCC tissue material showed hypomethylation. The observation of opposing
432 directions of differential methylation in blood versus the affected tissue is not unusual. In our latest
433 epigenome-wide association study, a regulatory region upstream of NHL repeat containing 1 (*NHLRC1*)
434 was found to be hypervariable, and hypermethylated in blood of lung cancer patients, but
435 hypomethylated in lung cancer versus normal adjacent tissue [97]. Although blood DNA methylation
436 patterns are not necessarily reflecting the changes in the affected tissues, especially in metabolic
437 diseases where blood glucose and triglyceride or cholesterol levels are constantly elevated when
438 untreated, these patterns could direct us towards the affected target genes.

439 The analysis of gene expression in liver tissue of MAFLD and HCC patients demonstrated their
440 differential expression. In line with the data from the liver expression database (Fig. 1E-G)[72], Ilieva and
441 colleagues found that *GCK* expression was increased by 2.48-fold (false discovery rate (FDR) adjusted
442 *p*=1.08E-05) in MAFLD compared to obese-healthy subjects and by 1.83-fold (FDR adjusted *p*=1.21E-
443 05) in advanced MetD patients compared to lean-healthy subjects in blood[98]. In contrast, *GCK* was
444 downregulated in HCC compared to normal tissue from TCGA (Fig. 1E), which is at least in part
445 explained by the replacement of *GCK* by hexokinase 2 in tumor cells[99]. The same trend was observed
446 for *TM6SF2*, which was overexpressed in MAFLD, but decreased in HCC. Strikingly, the overlapping
447 sense lncRNA ENSG00000267629 was overexpressed by 1.92-fold in the liver expression database in
448 MAFLD compared to normal tissue[72] and even by 3.19-fold (FDR adjusted *p*=.0005898) in MAFLD
449 compared to obese-healthy subjects in the study of Ilieva et al.[98]. Furthermore, the expression level of
450 this lncRNA was prognostic for the survival of intrahepatic cholangiocarcinoma (iCCA) with a hazard
451 ratio (HR) of 1.582 (1.143-2.190, 95% CI, *p*=.006). Patients with a higher survival probability showed
452 lower expression of this lncRNA transcript [65]. Moreover, it was also prognostic for patients suffering
453 from acute myeloid leukemia (AML)[64].
454 Summarizing, DNA methylation of the target regions is not only detected in early stages of a
455 metabolic disease (in *TM6SF2* after age correction even already in obesity) but also in HCC, indicating
456 that DNA methylation changes observed in cancer are in some cases already detectable in a
457 precancerous state which is in keeping with the increased risk for HCC observed in MAFLD patients.
458
459 G4s have emerged as major contributors to DNA methylation regulation and therefore, the
460 detection of DMPs in G4 forming motifs is not new. Nevertheless, the shared occurrence of aberrant
461 DNA methylation in an early disease stage such as obesity or metabolic diseases prior to cancer
462 formation, implies an urgent need to understand their functions. We were able to show G4 formation
463 in the *TM6SF2* intron 2-exon 3 boundary with PDAL-Seq in living cells (Fig. 2A), which adopts a parallel
464 structure according to the CD-spectra (Fig. 2E). Overall, the DNA methylation in normal liver tissue (Fig.
465 2D) and also in several cell lines is low (Supplement Fig. 6A). Therefore, stable G4 formation is most
466 likely occurring in this region. In contrast, *GCK* exon 7 harbors a ssDNA stretch, but slightly downstream
467 of the G4-Seq peak. Since many cell lines such as HepG2 or MCF7 (Supplement Fig. 6B) as well as the
468 normal liver tissue (Fig. 1C) are highly methylated in this region, stable G4 formation is unlikely, rather
469 than a hairpin, as inferred from the CD spectra of the methylated *GCK* oligo. Hence, when
470 demethylation occurs, a parallel structure is formed (Fig. 2F and G). This observation is in line with the
471 structural change in *c-KIT* and *HRAS* caused by DNA methylation[52]. Furthermore, PDAL-Seq revealed
472 the formation of G4 structures also exon 8-10, which has been shown as significantly hypomethylated
473 in patients suffering from coronary heart disease (49.77 ± 6.43%) compared to controls (54.47 ± 7.65%,

474 *p*=.018)[63]. CVDs are a common comorbidity of MAFLD and T2D and therefore this hypomethylation
475 reflects the same pattern than the observed hypomethylation at exon 7 in our cohort.
476
477 When differential methylation occurs within regulatory elements, it inevitably affects gene
478 expression patterns. Both regions exhibit cell-type-specific histone marks (Fig. 3A), luciferase reporter
479 assays demonstrate that their regulatory functions are activated in a DNA methylation-dependent
480 manner (Fig. 3B and C), and HiC data indicate that these elements interact with neighboring genes (Fig.
481 4). Notably, DNA methylation of both regions is highly variable, and exhibit enriched binding for the
482 zinc finger transcription factor EGR1. While it has been reported that G4s drive DNA methylation
483 processes by sequestering DNMT1 to the G4 and cause its inactivation[26], we propose a slightly
484 different, DNMT1-independent mechanism for these regions.
485 EGR1 recognizes a G-rich motif[100] and is predicted as G4 binding protein[101]. It recruits the ten
486 eleven translocase 1 (TET1), leading to DNA demethylation processes adjacent to its binding sites[102,103].
487 Its binding is likely not restricted solely to folded G4s, as the *PARD3B* region, which lacks any non-B
488 DNA structure, also shows EGR1 binding. EGR1 has been shown as important transcription factor
489 regulating inflammatory genes by binding to their enhancers[104]. EGR1 is upregulated in MAFLD, and a
490 its knockdown resulted in a reduced expression of pro-inflammatory genes in hepatocytes[105].
491 Given that the *GCK* exon 7 G4-CpGs are methylated in normal liver tissue, as well as different
492 cell lines, we hypothesize that local demethylation processes initiated by EGR1 binding, facilitate G4
493 formation, thereby allowing transcription factor binding and activation of the regulatory element (Fig.
494 5A). This is supported by experiments showing a ~60-fold increase in promoter activity and a ~27-fold
495 increase in enhancer function when GCK exon 7 is unmethylated (Fig. 3C). The *TM6SF2* intron 2-exon
496 3 boundary shows clear non-B DNA formation, and methylation reduces the stability of its G4 structure
497 (Fig. 2E). DNA methylation at the nearest CpG position is lower in the target region in *TM6SF2* than in
498 *GCK* in both normal tissues and cell lines (Fig. 1D, Supplement Fig. 6), implying that this element is
499 active in a healthy state, at least in a cell-type specific manner. The strong binding of EGR1 to the region
500 suggests, that DNA demethylation is facilitated by EGR1 and TET1 recruitment and a stable parallel G4
501 is formed. However, increased DNA methylation in this region, as seen in advanced MetD patients or
502 in cases of high methylation instability as observed in HCC, could impact the regulatory element's
503 activity. Notably, we analyzed only a rather small part of the UCSC-annotated regulatory element,
504 which may have caused the overall low detected enhancer strength. In line with our study, but
505 investigating long G4-forming regions, Williams et al. found that G4-forming regions harbor promoter
506 and enhancer functions and that the most prominently enriched interacting transcription factor
507 binding site was that of EGR1[32].
508 Alternative promoters play critical roles in hepatocellular and breast cancer, two cancer types
509 highly linked to metabolic diseases and obesity[10,12]. Genes involved in glycolysis and cholesterol

510 metabolism show elevated expression from alternative promoters, which typically exhibit low
511 H3K4me3 levels and hypomethylation in CpGs located in non-CGI promoters[58]. CGIs are
512 underrepresented in alternative promoters, and Nepal et al. propose that activation of these
513 promoters in HCC occurs through demethylation of CpG-poor intragenic regions[58], similar to the
514 investigated G4-forming regions.

515 Additionally, the TM6SF2 intron 2-exon 3 boundary could be subject to alternative splicing,
516 since exon 3 of *TM6SF2* is annotated as a cassette exon, and G4s are enriched at distal splice sites of
517 such exons[33]. Splicing factors like U2AF1, which recognizes the 3' splice site, are among the top
518 candidates associated with G4s[37]. G4 formation regulates alternative splicing of BCL-x, since
519 stabilization of the G4 structure promotes the expression of the pro-apoptotic BCL-xS isoform and
520 triggering apoptosis[106]. However, this requires critical discussion since the G4 motif is found on the
521 template strand, meaning the pre-mRNA cannot not form the G4. Probably the G4 creates an obstacle
522 for the RNA polymerase, causing reduced polymerase speed or stalling. Of note, proteins recognizing
523 methylated DNA can slow down RNA polymerase II, thereby indirectly influencing alternative splicing
524 mechanisms[107]. In line with this, targeted demethylation via TET1 linked to dCas9 causes exon
525 skipping[108]. Aberrant splicing regulation can result in readthroughs to neighboring genes during stress
526 conditions[109], which could explain the overexpressed lncRNA ENSG00000267629 in MAFLD. In
527 summary, the observed differential methylation in MAFLD groups in HCC, in combination with G4
528 formation, suggests a rather complex interplay between transcription and probably also alternative
529 splicing processes (Fig. 5B).
530

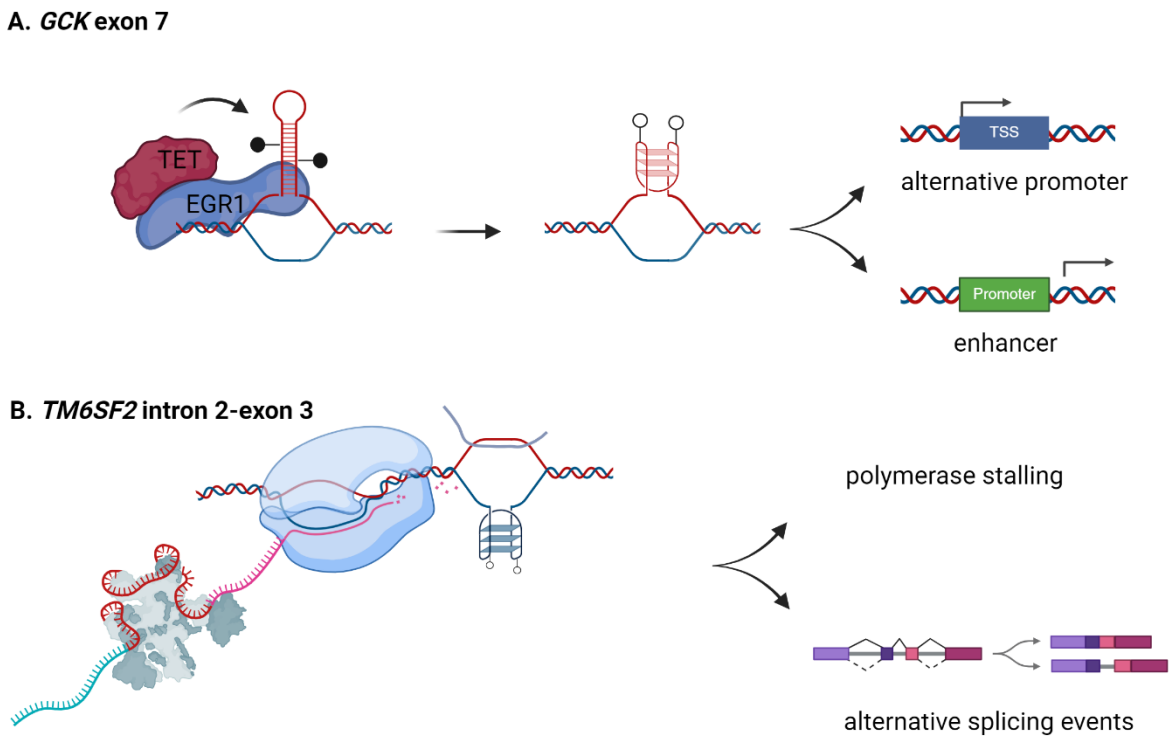


Figure 5: Mechanistic explanation. A. Hypomethylation of *GCK* exon 7 leads to the formation of a G4 structure which in turn activates the alternative regulatory element. **B.** Methylation instability at the cassette exon and G4 forming region in the *TM6SF2* intron 2-exon 3 boundary could cause polymerase slow down or stalling and alternative splicing events leading to a readthrough of *TM6SF2* to *HAPLN4*, forming the long non-coding RNA. Created with BioRender.com.

Conclusion

We were able to show aberrant DNA methylation within the investigated intragenic regions of *GCK* and *TM6SF2* in metabolic diseases as well as in HCC and that these structures are alternative regulatory elements potentially controlled by G4 formation. The link between G4s and cancer is well known. While the number of G4s naturally decreases during cell differentiation[110], G4s are increased in tumor tissue[111] and in the human keratinocytes (HaCaT) cancer cell line versus normal human epidermal keratinocytes NHEK[112]. Even the cell-type specific transcriptome has been shown to be shaped by G4s[113]. Therefore, G4s harbor a great potential as novel therapeutics and there are already G4 ligands which are used in clinical trial for solid tumors with mutations in homologous repair genes or pancreatic cancers (reviewed in[114]). Recently it has been shown that pathologically high glucose concentrations cause the formation of a stable promoter-G4 and in turn result in significantly increased gene expression compared to normal glucose concentrations[45]. We suggest that a similar mechanism also applies for MetD patients since they are exposed constantly to high glucose and high fatty acid concentrations, if not treated properly.

551 Although further research is necessary to get a deeper understanding of the mechanisms
552 involved, we were able to connect detected DMPs in metabolic risk genes to G-quadruplex formation
553 in a patient cohort suffering from metabolism dysfunction-associated fatty liver disease and type II
554 diabetes, for the very first time. Unfortunately, we are currently lacking G4 formation data in human
555 liver biopsies, since this type of material is rare. Although the assessment of anthropometric and blood
556 parameters, abdominal MRI and liver biopsies are the current standard for MAFLD diagnosis, liver
557 biopsies cannot be applied on a broad scale in asymptomatic obese or lean-MAFLD subjects for ethical
558 reasons or low patient acceptance. In conclusion, since metabolic diseases are linked to an elevated
559 cancer risk, and aberrant DNA methylation as well as elevated non-B DNA formation occurs during
560 cancer formation, these mechanisms could point to a common mechanism of metabolic diseases and
561 liver cancer development. Understanding this intricate connection especially in intragenic regions with
562 alternative regulatory roles is paramount in unraveling the complexities of metabolic diseases,
563 potentially leading to G4s as novel targets for metabolic disease treatment as already evaluated for
564 solid cancers.
565

566 ## 4. Methods

567 **Study design**

568 We designed a nested cohort study by selecting 148 blood samples from Caucasian origin collected
569 at the University Hospital Salzburg, Austria, which fit into one of five groups according to the criteria
570 below. The local ethics committee (Ethikkommission des Landes Salzburg No. 415-E/1521/6-2012, and
571 an ethics approval by the Paracelsus Medical University Salzburg No. 01-2022) approved the study and
572 informed consent was obtained from all participants. The study was conducted in accordance with the
573 guidelines of the World Medical Association's Declaration of Helsinki.
574 Alanine transaminase (ALT) and aspartate transaminase (AST) levels ≤40 U/L for women and ≤50
575 U/L for men, with or without increased levels of γ-glutamyl transpeptidase (GGT), were considered as
576 normal liver tests. Routine clinical parameters were determined during sampling at the outpatient
577 clinics of the University Hospital Salzburg, Austria.
578 (1) Lean healthy subjects (LH, n=24): body mass index (BMI) ≤25.0 kg/m^2, normal liver tests,
579 normal ultrasound and no components of MetS according to ATPIII criteria[115]
580 (2) Obese healthy subjects (OH, n=31): metabolic healthy obese (MHO) BMI ≥30.0 kg/m^2, normal
581 ultrasound and normal liver tests, and ≤2 components of MetS (obesity and one additional
582 metabolic syndrome defining criterion allowed and glycated hemoglobin (HbA_{1c})<6.0%;
583 (3) Lean MAFLD subjects (LM, n=27): metabolic unhealthy normal weight (MUNW), BMI ≤25.0
584 kg/m^2, ultrasound evidence of a fatty liver, with or without elevated liver tests, HbA_{1c}<7%;
585 (4) Obese MAFLD subjects (OM, n=32): BMIs ≥30.0 kg/m^2, unequivocal fatty liver ultrasound, with
586 or without elevated liver tests, HbA_{1c}<7%;
587 (5) Advanced MetD (aMetD, n=34): Elevated fibrosis-4 (FIB4) score>2.67 and/or established T2D
588 with elevated HbA_{1c}>7%.

589 **Sequence specific DNA methylation analysis by pyrosequencing**

590 DNA was extracted from 200 µL whole blood with the ReliaPrep Blood gDNA Miniprep System
591 (Promega, Austria) according to the manufacturer´s guidelines. Bisulfite (BS) conversion was
592 performed with 500 ng of genomic DNA (gDNA) input material using a Bisulfite Conversion Kit (Zymo
593 Research, Austria). DNA was eluted with 60 µL M-Elution buffer (~8 ng/µL). PCR was performed in 25
594 µL reaction volume containing 200 µM dNTPs, 1 mM $MgCl_2$, each 0.2 µM forward and reverse primer,
595 0.63 U HotStar Taq polymerase (Qiagen, Austria) and 16 ng (*P2RXL1*, SCGN and *NHLRC1*) or 32 ng
596 (*EDARADD*, *IPO8*, *GCK* and *TM6SF2*) of BS-converted DNA to reach sufficient material for
597 pyrosequencing. The PCR protocol included an initial heating step of 95°C for 15 minutes, followed by
598 45 cycles of 95°C for 30 seconds, 55-64°C primer annealing depending on the target (see Supplement

599 Material & Methods section 1.1) for 30 seconds and 72°C for 15 seconds, as well as a final elongation
600 step at 72°C for 7 minutes. Pyrosequencing was performed using 20 µL PCR product and 7.5 µM (*IPO8*,
601 *GCK*, *TM6SF2, P2RXL1*, and *NHLRC1*) or 9 µM (*SCGN* and *EDARADD*) sequencing primer. Detailed assay
602 design, primer sequences, annealing temperatures and sequencing conditions can be found in the
603 Supplement (Material & Methods section 1.1). Data analysis was performed with Pyromark Q24
604 Advanced software, R version 4.3.1[116] and R studio[117].

605 **rs58542926 C>T single nucleotide polymorphism (SNP) genotyping by pyrosequencing**

606 To infer the genotype of the rs58542926 C>T (E167K) risk variant, we amplified the region in
607 25 µL reaction volume using 200 µM dNTPs, 1 mM $MgCl_2$, each 0.2 µM forward and reverse primer,
608 0.63 U HotStar Taq polymerase (Qiagen, Austria) and 10 ng genomic DNA. The PCR protocol included
609 an initial heating step of 95°C for 15 minutes, followed by 45 cycles of 95°C for 30 seconds, 63°C for 30
610 seconds and 72°C for 15 seconds, as well as a final elongation step at 72°C for 7 minutes. For the
611 pyrosequencing reaction, we used 20 µl PCR product and 7.5 µM sequencing primer. Detailed
612 information is given in the Supplement Material & Methods section 1.1. Data analysis was performed
613 with Pyromark Q24 Advanced software, R version 4.3.1 [116] and R studio [117].

614 **Gene expression data**

615 Expression data for MAFLD and HCC were extracted from the liver expression portal
616 (GepLiver)[72]. For the analysis, we downloaded the datasets for Bulk-01 (normal liver), Bulk-05-09
617 (including MAFLD) and Bulk-19 (TCGA; including HCC) datasets as listed in the GepLiver portal. All
618 statistical analyses were performed using R version 4.3.1[116] and R Studio[117].

619 **Non-B DNA prediction tools**

620 To predict non-B forming sites, we utilized pqsfinder for G4 forming sites[77], QmRLFS-finder for
621 R-loop forming sequences (RLFS)[76], and the *RNAfold* web server of the University of Vienna/Austria for
622 the folding of hairpins[118].

623 **Non-B DNA detection with permanganate/S1 nuclease footprinting with direct adapter ligation and**
624 **sequencing (PDAL-Seq)**

625 PDAL-Seq was performed according to the protocol described in[81]. Further details can be found
626 in the Supplement (Material & Methods Section 1.3). In short, 5-7 x 10^6 cells were washed two times
627 with 1x PBS without Ca^{2+} and Mg^{2+} and incubated in 1 mL low salt buffer (15 mM Tris-HCl pH 7.5, 15
628 mM NaCl, 60 mM KCl, 300 mM sucrose, 0.5 mM EGTA) for 1 min. 670 µL of freshly prepared $KMnO_4$

629 stock solution (40 mM for HEK293, 70mM for MCF7) were added and incubated at 37°C in the dark for
630 80 sec. Subsequently, the reaction was stopped with 2.67 mL stop solution (50 mM EDTA, 700 mM β-
631 mercaptoethanol, 1% SDS) and 79.5 µL proteinase K (20 mg/ml stock solution, Promega, Austria) were
632 added. After overnight incubation at 37°C, DNA was extracted with phenol chloroform isoamylalcohol
633 (PCI, Merck, Austria) and precipitated with 2 M NH_4OAc and 100% ethanol. After resuspending the
634 pellet in 1 mL of 10 mM Tris HCl pH 8.0 buffer, RNA was digested by adding 2.5 µL of 20 mg/mL RNase
635 A (Promega, Austria) for 1 h followed by another PCI extraction. Free DNA ends were blocked with 500
636 U Terminal Transferase (Roche, Austria), 120 µM cordycepin-5´-triphosphate and 5 mM $CoCl_2$ in 500
637 µL total reaction volume, followed by a PCI extraction and two ethanol precipitations to avoid
638 cordycepin carry over. 40 µg DNA were digested with 20 U S1 nuclease (0.5 U/1 µg DNA, Promega,
639 Austria) in a total reaction volume of 600 µL at 37°C for 20 min. The reaction was stopped by adding
640 10 mM EDTA. After PCI extraction, end repair using T4 DNA polymerase (NEB, Austria) was performed
641 for 15 min at 12°C, followed by PCI extraction. 500 nM biotinylated P5 adapter (Merck, Germany) were
642 linked with 2000 U T4 DNA ligase (Promega, Austria) in a final volume of 100 µL at 16°C overnight. To
643 remove unligated P5 adapters, Agencourt SPRI bead cleanup (Beckman Coulter, Austria) was
644 performed with 0.9 volumes of beads and eluted in 132 µL 1x TE buffer (10 mM Tris HCl pH 8.0, 1mM
645 EDTA). Sonication was performed with an Covaris M220 focused-ultrasonicator in 130 µL microTube-
646 AFA Fiber Pre-Slit Snap Cap tubes with the parameters: peak incidence power: 75, duty factor: 10,
647 cycles per burst: 200 and treatment time: 80 sec, to reach a target peak of approximately 400 bp.
648 Fragments were bound to streptavidin coated magnetic beads (Dynabeads® M-280 kilobaseBINDER
649 kit, Thermo Scientific, Austria) and carefully washed with 10 mM Tris-HCl pH 7.5. The beads were then
650 resuspended in 100 µL end repair mix (NEBNext End Repair Module, NEB, Austria) and incubated at
651 20°C for 30 min. After washing with 10 mM Tris-HCl pH 7.5, P7 adapter ligation (1 µM adapters, 400 U
652 T4 DNA ligase) was performed on a wheel tube rotator at 20°C for 4 h. Beads were washed with 10
653 mM Tris-HCl pH 7.5 and resuspended in 100 µL PCR mix and PCR was performed. Detailed information
654 about PCR reaction and primers are provided in the methods paper[81] and in the Supplement (Material
655 and Methods section 1.3). SPRI bead purification removed unbound adapters and primer dimers and
656 the quality of the resulting libraries was checked using an Agilent Fragment Analyzer. Library
657 concentration was determined with a KAPA Library Quantification Kit (Roche, Austria) and sequenced
658 on a NovaSeq platform with approximately 80 million reads. After demultiplexing the raw reads and
659 quality control, reads were aligned to the human reference genome hg19 with the BWA-MEM
660 aligner[119]. Subsequent filtering removed low-quality reads (mapping quality MQ <30) and ENCODE
661 blacklisted regions[82]. BamCoverage files were generated with a 50 bp bin size and normalized to 1×
662 coverage and visualized in IGV[120]. Data analysis was performed using the open-source platform
663 Galaxy[121].

664 **Studying *in vitro* G4 formation with native DNA polyacrylamide gels and circular dichroism (CD)**
665 **spectroscopy**

666 Target regions harboring the DMPs in *GCK* and *TM6SF2*, a positive control region in *MYC*
667 containing a predicted G4 motif[122], and a negative control region within a regulatory region of the
668 estrogen receptor substrate 1 (*ESR1)* without a predicted G4 motif[123] were PCR-amplified. To do so,
669 we prepared a 10 µL PCR reaction containing 200 µM dNTPs, 1 mM $MgCl_2$, each 0.2 µM forward and
670 reverse primer, 0.25 U HotStar Taq polymerase (Qiagen, Austria) and 10 ng genomic DNA. The PCR
671 protocol included an initial heating step of 95°C for 15 minutes, followed by 45 cycles of 95°C for 30
672 seconds, 54-63°C primer annealing depending on the gene for 30 seconds and 72°C for 15 seconds, as
673 well as a final elongation step at 72°C for 7 minutes. The PCR fragments were cleaned-up with ReliaPrep
674 DNA concentration kit (Promega, Austria) and eluted with dH_2O. The amplicons were then diluted to a
675 final concentration of 20 ng/µL in 50 mM Tris-HCl (pH 8.3). To enhance G4 folding, we prepared the
676 same dilutions and added 150 mM KCl or 150 mM KCl and 10 µM pyridostatin (PDS, Merck, Austria) to
677 the Tris-HCl buffer. The fragments were then heated to 95°C and cooled gradually down to 25°C in a
678 thermocycler overnight. DNA fragments were loaded onto a native 10% DNA polyacrylamide (PAA) gel
679 using 1× TBE buffer at 68 V (8 V/cm^2) for 2h. DNA PAA gels were post-stained with GelRed (Biotium,
680 VWR, Austria) for 5 min and visualized with a ChemiDoc station (BioRad, Austria). Detailed information
681 can be found in the Supplement (Material & Methods Section 1.3) as well as the raw gel image in
682 Supplement Fig. 7.

683 Circular dichroism (CD) spectra were obtained from single stranded DNA oligos from the
684 predicted G4 motif as well as in a methylated and mutated form (Supplement Material & Methods
685 Section 1.4). All oligos were reconstituted in 50 mM Tris-HCl (pH 7.5) and ranging from 0 mM, 50 mM
686 to 150 mM KCl in a final volume of 200 µL with an oligo concentration of 5µM. CD spectra were
687 recorded using the Chirascan™ Plus CD Spectrophotometer instrument (Applied Photophysics, United
688 Kingdom) in a 0.5 cm path cuvette at 25°C. Spectra were scanned from 340 to 220 nm., with a step size
689 of 0.2 nm and a bandwidth of 1 nm. Four spectra per sample were averaged. After blank correction,
690 the CD ellipticity θ in millidegrees was normalized using the relation $\theta norm = \theta / (32980 \times c \times l)$, where
691 c is the DNA concentration in mol/L and l is the path length in cm.

692 **TCGA data and ENCODE histone mark and TF binding data analysis**

693 DNA methylation and gene expression data were obtained from tumor versus normal tissue
694 from patients suffering from hepatocellular carcinoma (HCC) via the Shiny Methylation Analysis
695 Resource Tool (SMART), an interactive tool for analyzing *The Cancer Genome Atlas* (TCGA) data[124].

696 *The Encyclopedia of DNA Elements* (ENCODE) data for H3K27ac, H3K4me1 and H3K4me3 from
697 the liver hepatoblastoma cell line HepG2 and the lymphoblast cell line K562[82] were downloaded and

698 visualized using the Integrative Genome Viewer (IGV)[120]. Early growth response 1 (EGR1) chromatin
699 immunoprecipitation and sequencing (ChIP-Seq) data from K562 cells were obtained from the ENCODE
700 database[82].

701 **Analyzing contact points in HiC datasets**

702 Genome organization and chromatin interaction data (HiC) were obtained from HepG2 cells
703 via ENCODE[82], and for HCC material[83], and visualized with 3D Genome Browser[84]. To detect potential
704 interactions with surrounding genes, we used the coordinates chr7: 44,187,355 in hg19
705 (chr7:44,147,756; hg38) for *GCK* exon 7 and chr19:19,381,266 in hg19 (chr19:19,270,457; hg38) for
706 *TM6SF2* intron 2-exon 3 for virtual 4C analysis using HiC data in the 3D Genome Browser[84].

707 **Testing for regulatory functions using the luciferase reporter assay**

708 To verify the regulatory role of the DMPs identified in *GCK* exon 7 and *TM6SF2* intron 2-exon
709 3 boundary, the regions were cloned into pCpG-free vectors (Invivogen, Belgium). The pCpG-free basic
710 vector was used to detect potential promoter activity and the pCpG-free promoter vector contains a
711 minimal promoter to test for enhancer activity of the regions of interest. Since these vectors do not
712 contain any CpG dinucleotides in the backbone, *in vitro* DNA methylation can be performed with the
713 *M.SssI* enzyme, which then selectively methylates only the inserted human *GCK* or *TM6SF2* region and
714 allows to selectively study the effect of 5mC in the insert on luciferase gene expression.
715 Briefly, DNA was isolated from HepG2 cells and the regions were amplified in 50 µL reaction
716 volumes containing 1x GC buffer, 200 µM dNTPs, 0.5 µM forward and reverse primer, and 0.5 U
717 Phusion HSII polymerase (Thermo Scientific, Austria). The PCR protocol contained an initial heating
718 step of 98°C for 30 seconds, followed by 45 cycles of 98°C for 10 seconds, 69°C for 10 seconds and 72°C
719 for 15 seconds, as well as a final elongation step at 72°C for 5 minutes. The PCR primers contained
720 restriction enzyme (RE) cutting sites for *NsiI* (NEB, Austria) and *BamHI*/*HindIII* (NEB, Austria) at the
721 5´ends of the forward and reverse primers, respectively. 500 ng of each, PCR amplicon and plasmid,
722 were double-digested at 37°C for 20 min and cleaned up with a DNA cleanup and concentration system
723 (Promega, Austria). Detailed information can be found in the Supplement (Material & Methods section
724 1.5).
725 The ligation reaction was performed in a 1:5 plasmid to insert copy number ratio using T4 DNA
726 ligase (NEB) at 16°C overnight. 5 µL of the ligated construct were transferred to chemically competent
727 *E.coli* GT115 cells (*pir* mutant strain, Invivogen, Belgium) and spread on lysogeny broth (LB) plates
728 containing zeocin (50 µL/ml; Invivogen, Belgium). Singe colonies were picked and incubated in 3 ml LB-
729 zeocin medium overnight. Plasmid was extracted with PureYield Plasmid Midiprep System (Promega,
730 Austria) and successful cloning was verified by Sanger Sequencing at LGC genomics (Germany). 1 µg of

731 plasmid was incubated with 8 U *M.SssI* (NEB) at 37°C for 4 h for *in vitro* DNA methylation. The reaction
732 was heat deactivated at 65°C for 20 min, followed by plasmid cleanup with ReliaPrep DNA Clean-Up
733 and Concentration kit (Promega, Austria).

734 24 h prior to transfection, 0.15 x 10^5 HepG2 cells per well were seeded in 24 well plates with
735 DMEM low glucose medium + 10% fetal bovine serum without antibiotics. Lipofectamin 3000 (Thermo
736 Scientific, Austria) was used to transfect 500 ng plasmid (renilla luciferase) and 25 ng pGL4 SV40
737 background plasmid (firefly luciferase) and incubated overnight. After cell lysis, chemiluminescent
738 read-out was detected on a Tecan Spark plate reader (Tecan, Austria). Renilla readouts were
739 normalized to the firefly luciferase and finally, data were expressed as fold-change normalized against
740 the empty vector.

741 **Statistical analyses**

742 All statistical analyses were performed using R version 4.3.1 [116] and R Studio [117]. Significant
743 changes in DNA methylation in all groups were detected using analysis of variance (ANOVA) with $p<.05$.
744 For individual group comparisons, Student´s t-test ($p<.05$) was applied. DNA methylation data from
745 TCGA and gene expression data from GepLiver database were analyzed with Wilcoxon rank sum test
746 using Benjamini Hochberg (BH) correction. Welch´s two sample t-tests were performed to calculate
747 significant differences in luciferase reporter assays ($p<.05$).

748 **Data Availability**

749 PDAL-Seq data are available at the Sequence Read Archive (SRA) under the BioProject number
750 PRJNA1088738 (https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA1088738&o=acc_s%3Aa).

751 **Author contribution**

752 A. L. established PDAL Seq, designed assays and conceptualized the study; A. L., V. E., A. O., Z.
753 L. and E. S. performed experiments, B. P. and E. A. collected and grouped the samples and gave
754 substantial advice and intellectual input to the study; A. L. wrote the manuscript, A. L., E.A. and A.R.
755 revised and edited the manuscript. All authors read and approved the final manuscript.

756 **Funding**

757 A.L was supported by a *Starter Grant* awarded by the Austrian Diabetes Association (ÖDG) as
758 well as with the Early Career Grant of the University of Salzburg (Pure ID: 35010166). This research was
759 in part funded by the County of Salzburg, Cancer Cluster Salzburg [grant number 20102-P1601064-
760 FPR].

Acknowledgments

We want to thank the Austrian Diabetes Association (ÖDG) very warmly for their support as well as all patients who contributed their blood samples to the study. Furthermore, we thank Kateryna Makova and Matthias Weissensteiner for plenty of fruitful discussions on PDAL-Seq and their input on G-quadruplex structures.

Conflicts of Interest

A. L., V. E., A. O., Z. L., E. S., B. P., E. A. and A. R. do not have any conflicts of interest.

Abbreviations

5mC	5-methylcytosine
BS	Bisulfite conversion
ChIP-Seq	Chromatin immunoprecipitation and sequencing
DMP	Differentially methylated position
DMR	Differentially methylated region
G4	G-quadruplex
GCK	Glucokinase
HiC/4C	Chromatin conformation capture
MAFLD	Metabolism dysfunction-associated fatty liver disease
MetD	Metabolic disease
MetS	Metabolic syndrome
Non-B DNA	Non-canonical DNA structure
PDAL-Seq	Permanganate/S1 nuclease footprinting with direct adapter ligation
PQS	Putative G4 forming sequence
R-loop	RNA:DNA hybrid
RLFS	R-loop forming sequence
T2D	Type 2 diabetes mellitus
TF	Transcription factor
TM6SF2	Transmembrane 6 superfamily 2

References

1 WHO. *Obesity and overweight, Fact Sheets, https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight*, 2020).

2 Koenen, M., Hill, M. A., Cohen, P. & Sowers, J. R. Obesity, Adipose Tissue and Vascular Dysfunction. *Circ Res* **128**, 951-968 (2021). https://doi.org/10.1161/CIRCRESAHA.121.318093

3 Collaboration, N. C. D. R. F. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128.9 million children, adolescents, and adults. *Lancet* **390**, 2627-2642 (2017). https://doi.org/10.1016/S0140-6736(17)32129-3

4 Samblas, M., Milagro, F. I. & Martinez, A. DNA methylation markers in obesity, metabolic syndrome, and weight loss. *Epigenetics* **14**, 421-444 (2019). https://doi.org/10.1080/15592294.2019.1595297

5 Fabbrini, E., Sullivan, S. & Klein, S. Obesity and nonalcoholic fatty liver disease: biochemical, metabolic, and clinical implications. *Hepatology* **51**, 679-689 (2010). https://doi.org/10.1002/hep.23280

6 Yamamura, S. *et al.* MAFLD identifies patients with significant hepatic fibrosis better than NAFLD. *Liver Int* **40**, 3018-3030 (2020). https://doi.org/10.1111/liv.14675

7 Afshin, A., Reitsma, M. B. & Murray, C. J. L. Health Effects of Overweight and Obesity in 195 Countries. *N Engl J Med* **377**, 1496-1497 (2017). https://doi.org/10.1056/NEJMc1710026

8 Chen, Y. *et al.* Body mass index and risk of gastric cancer: a meta-analysis of a population with more than ten million from 24 prospective studies. *Cancer Epidemiol Biomarkers Prev* **22**, 1395-1408 (2013). https://doi.org/10.1158/1055-9965.EPI-13-0042

9 Genkinger, J. M. *et al.* A pooled analysis of 14 cohort studies of anthropometric factors and pancreatic cancer risk. *Int J Cancer* **129**, 1708-1717 (2011). https://doi.org/10.1002/ijc.25794

10 Wang, Y., Wang, B., Shen, F., Fan, J. & Cao, H. Body mass index and risk of primary liver cancer: a meta-analysis of prospective studies. *Oncologist* **17**, 1461-1468 (2012). https://doi.org/10.1634/theoncologist.2012-0066

11 Ma, Y. *et al.* Obesity and risk of colorectal cancer: a systematic review of prospective studies. *PLoS One* **8**, e53916 (2013). https://doi.org/10.1371/journal.pone.0053916

12 Chlebowski, R. T. *et al.* Breast cancer incidence and mortality by metabolic syndrome and obesity: The Women's Health Initiative. *Cancer* (2024). https://doi.org/10.1002/cncr.35318

13 Carreras-Torres, R. *et al.* Obesity, metabolic factors and risk of different histological types of lung cancer: A Mendelian randomization study. *PLoS One* **12**, e0177875 (2017). https://doi.org/10.1371/journal.pone.0177875

14 Li, M. *et al.* Association of Metabolic Syndrome With Risk of Lung Cancer: A Population-Based Prospective Cohort Study. *Chest* (2023). https://doi.org/10.1016/j.chest.2023.08.003

15 Grasselli, E. *et al.* Models of non-Alcoholic Fatty Liver Disease and Potential Translational Value: the Effects of 3,5-L-diiodothyronine. *Ann Hepatol* **16**, 707-719 (2017). https://doi.org/10.5604/01.3001.0010.2713

16 Younossi, Z. *et al.* Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. *Nat Rev Gastroenterol Hepatol* **15**, 11-20 (2018). https://doi.org/10.1038/nrgastro.2017.109

17 Roden, M. Mechanisms of Disease: hepatic steatosis in type 2 diabetes--pathogenesis and clinical relevance. *Nat Clin Pract Endocrinol Metab* **2**, 335-348 (2006). https://doi.org/10.1038/ncpendmet0190

18 Gaggini, M. *et al.* Non-alcoholic fatty liver disease (NAFLD) and its connection with insulin resistance, dyslipidemia, atherosclerosis and coronary heart disease. *Nutrients* **5**, 1544-1560 (2013). https://doi.org/10.3390/nu5051544

19 Chalasani, N. *et al.* The diagnosis and management of non-alcoholic fatty liver disease: practice Guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association. *Hepatology* **55**, 2005-2023 (2012). https://doi.org/10.1002/hep.25762

20 Ma, J. *et al.* A Peripheral Blood DNA Methylation Signature of Hepatic Fat Reveals a Potential Causal Pathway for Nonalcoholic Fatty Liver Disease. *Diabetes* **68**, 1073-1083 (2019). https://doi.org/10.2337/DB18-1193

21 Gerhard, G. S. *et al.* Differentially methylated loci in NAFLD cirrhosis are associated with key signaling pathways. *Clin Epigenetics* **10**, 93 (2018). https://doi.org/10.1186/s13148-018-0525-9

22 Ahrens, M. *et al.* DNA methylation analysis in nonalcoholic fatty liver disease suggests distinct disease-specific and remodeling signatures after bariatric surgery. *Cell Metab* **18**, 296-302 (2013). https://doi.org/10.1016/j.cmet.2013.07.004

23 Kirchner, H. *et al.* Altered DNA methylation of glycolytic and lipogenic genes in liver from obese and type 2 diabetic patients. *Mol Metab* **5**, 171-183 (2016). https://doi.org/10.1016/j.molmet.2015.12.004

24 Nano, J. *et al.* Epigenome-Wide Association Study Identifies Methylation Sites Associated With Liver Enzymes and Hepatic Steatosis. *Gastroenterology* **153**, 1096-1106 e1092 (2017). https://doi.org/10.1053/j.gastro.2017.06.003

25 Rauchhaus, J., Robinson, J., Monti, L. & Di Antonio, M. G-quadruplexes Mark Sites of Methylation Instability Associated with Ageing and Cancer. *Genes (Basel)* **13** (2022). https://doi.org/10.3390/genes13091665

26 Mao, S. Q. *et al.* DNA G-quadruplex structures mold the DNA methylome. *Nat Struct Mol Biol* **25**, 951-957 (2018). https://doi.org/10.1038/s41594-018-0131-8

27 Williamson, J. R., Raghuraman, M. K. & Cech, T. R. Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell* **59**, 871-880 (1989). https://doi.org/10.1016/0092-8674(89)90610-7

28 Sen, D. & Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **334**, 364-366 (1988). https://doi.org/10.1038/334364a0

29 Haider, S., Parkinson, G. N. & Neidle, S. Crystal structure of the potassium form of an Oxytricha nova G-quadruplex. *J Mol Biol* **320**, 189-200 (2002). https://doi.org/10.1016/S0022-2836(02)00428-X

30 Verma, A. *et al.* Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved cis-regulatory elements in human and related species. *J Med Chem* **51**, 5641-5649 (2008). https://doi.org/10.1021/jm800448a

31 Huppert, J. L. & Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res* **35**, 406-413 (2007). https://doi.org/10.1093/nar/gkl1057

32 Williams, J. D. *et al.* Characterization of long G4-rich enhancer-associated genomic regions engaging in a novel loop:loop 'G4 Kissing' interaction. *Nucleic Acids Res* **48**, 5907-5925 (2020). https://doi.org/10.1093/nar/gkaa357

33 Georgakopoulos-Soares, I. *et al.* Alternative splicing modulation by G-quadruplexes. *Nat Commun* **13**, 2404 (2022). https://doi.org/10.1038/s41467-022-30071-7

34 Guiblet, W. M. *et al.* Selection and thermostability suggest G-quadruplexes are novel functional elements of the human genome. *Genome Res* **31**, 1136-1149 (2021). https://doi.org/10.1101/gr.269589.120

35 Duquette, M. L., Handa, P., Vincent, J. A., Taylor, A. F. & Maizels, N. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev* **18**, 1618-1629 (2004). https://doi.org/10.1101/gad.1200804

36 Raiber, E. A., Kranaster, R., Lam, E., Nikan, M. & Balasubramanian, S. A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. *Nucleic Acids Res* **40**, 1499-1508 (2012). https://doi.org/10.1093/nar/gkr882

37 Spiegel, J. *et al.* G-quadruplexes are transcription factor binding hubs in human chromatin. *Genome Biol* **22**, 117 (2021). https://doi.org/10.1186/s13059-021-02324-z

38 Smith, S. S., Kan, J. L., Baker, D. J., Kaplan, B. E. & Dembek, P. Recognition of unusual DNA structures by human DNA (cytosine-5)methyltransferase. *J Mol Biol* **217**, 39-51 (1991). https://doi.org/10.1016/0022-2836(91)90609-a

39 Chen, X., Mariappan, S. V., Moyzis, R. K., Bradbury, E. M. & Gupta, G. Hairpin induced slippage and hyper-methylation of the fragile X DNA triplets. *J Biomol Struct Dyn* **15**, 745-756 (1998). https://doi.org/10.1080/07391102.1998.10508989

40 Frasson, I., Pirota, V., Richter, S. N. & Doria, F. Multimeric G-quadruplexes: A review on their biological roles and targeting. *Int J Biol Macromol* **204**, 89-102 (2022). https://doi.org/10.1016/j.ijbiomac.2022.01.197

41 Varshney, D., Spiegel, J., Zyner, K., Tannahill, D. & Balasubramanian, S. The regulation and functions of DNA and RNA G-quadruplexes. *Nat Rev Mol Cell Biol* **21**, 459-474 (2020). https://doi.org/10.1038/s41580-020-0236-x

42 Makova, K. D. & Weissensteiner, M. H. Noncanonical DNA structures are drivers of genome evolution. *Trends Genet* **39**, 109-124 (2023). https://doi.org/10.1016/j.tig.2022.11.005

43 Fujii, T., Podbevsek, P., Plavec, J. & Sugimoto, N. Effects of metal ions and cosolutes on G-quadruplex topology. *J Inorg Biochem* **166**, 190-198 (2017). https://doi.org/10.1016/j.jinorgbio.2016.09.001

44 Nishio, M., Tsukakoshi, K. & Ikebukuro, K. G-quadruplex: Flexible conformational changes by cations, pH, crowding and its applications to biosensing. *Biosens Bioelectron* **178**, 113030 (2021). https://doi.org/10.1016/j.bios.2021.113030

45 Guneri, D. *et al.* Structural Insights into Regulation of Insulin Expression Involving i-Motif DNA Structures in the Insulin-Linked Polymorphic Region. *bioRxiv*, 2023.2006.2001.543149 (2023). https://doi.org/10.1101/2023.06.01.543149

46 Perez, A. *et al.* Impact of methylation on the physical properties of DNA. *Biophys J* **102**, 2140-2148 (2012). https://doi.org/10.1016/j.bpj.2012.03.056

47 Li, P. T. *et al.* Expression of the human telomerase reverse transcriptase gene is modulated by quadruplex formation in its first exon due to DNA methylation. *J Biol Chem* **292**, 20859-20870 (2017). https://doi.org/10.1074/jbc.M117.808022

48 Lin, J. *et al.* Stabilization of G-quadruplex DNA by C-5-methyl-cytosine in bcl-2 promoter: implications for epigenetic regulation. *Biochem Biophys Res Commun* **433**, 368-373 (2013). https://doi.org/10.1016/j.bbrc.2012.12.040

49 Yoshida, W. *et al.* Detection of DNA Methylation of G-Quadruplex and i-Motif-Forming Sequences by Measuring the Initial Elongation Efficiency of Polymerase Chain Reaction. *Anal Chem* **88**, 7101-7107 (2016). https://doi.org/10.1021/acs.analchem.6b00982

50 Fry, M. & Loeb, L. A. The fragile X syndrome d(CGG)n nucleotide repeats form a stable tetrahelical structure. *Proc Natl Acad Sci U S A* **91**, 4950-4954 (1994). https://doi.org/10.1073/pnas.91.11.4950

51 Stevens, A. J. *et al.* G-quadruplex structures and CpG methylation cause drop-out of the maternal allele in polymerase chain reaction amplification of the imprinted MEST gene promoter. *PLoS One* **9**, e113955 (2014). https://doi.org/10.1371/journal.pone.0113955

52 Tsukakoshi, K., Saito, S., Yoshida, W., Goto, S. & Ikebukuro, K. CpG Methylation Changes G-Quadruplex Structures Derived from Gene Promoters and Interaction with VEGF and SP1. *Molecules* **23** (2018). https://doi.org/10.3390/molecules23040944

53 Halder, R. *et al.* Guanine quadruplex DNA structure restricts methylation of CpG dinucleotides genome-wide. *Mol Biosyst* **6**, 2439-2447 (2010). https://doi.org/10.1039/c0mb00009d

921 54 Jara-Espejo, M. & Line, S. R. DNA G-quadruplex stability, position and chromatin accessibility
922 are associated with CpG island methylation. *FEBS J* **287**, 483-495 (2020).
923 https://doi.org/10.1111/febs.15065
924 55 Singer, G. A. *et al.* Genome-wide analysis of alternative promoters of human genes using a
925 custom promoter tiling array. *BMC Genomics* **9**, 349 (2008). https://doi.org/10.1186/1471-
926 2164-9-349
927 56 Wang, J., Ungar, L. H., Tseng, H. & Hannenhalli, S. MetaProm: a neural network based meta-
928 predictor for alternative human promoter prediction. *BMC Genomics* **8**, 374 (2007).
929 https://doi.org/10.1186/1471-2164-8-374
930 57 Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature*
931 **507**, 455-461 (2014). https://doi.org/10.1038/nature12787
932 58 Nepal, C. & Andersen, J. B. Alternative promoters in CpG depleted regions are prevalently
933 associated with epigenetic misregulation of liver cancer transcriptomes. *Nat Commun* **14**,
934 2712 (2023). https://doi.org/10.1038/s41467-023-38272-4
935 59 Schamschula, E. *et al.* Disease-related blood-based differential methylation in cystic fibrosis
936 and its representation in lung cancer revealed a regulatory locus in PKP3 in lung epithelial
937 cells. *Epigenetics* **17**, 837-860 (2022). https://doi.org/10.1080/15592294.2021.1959976
938 60 Hattersley, A. T. *et al.* Linkage of type 2 diabetes to the glucokinase gene. *Lancet* **339**, 1307-
939 1310 (1992). https://doi.org/10.1016/0140-6736(92)91958-b
940 61 Njolstad, P. R. *et al.* Neonatal diabetes mellitus due to complete glucokinase deficiency. *N*
941 *Engl J Med* **344**, 1588-1592 (2001). https://doi.org/10.1056/NEJM200105243442104
942 62 Rose, C. S. *et al.* A -30G>A polymorphism of the beta-cell-specific glucokinase promoter
943 associates with hyperglycemia in the general population of whites. *Diabetes* **54**, 3026-3031
944 (2005). https://doi.org/10.2337/diabetes.54.10.3026
945 63 Xu, L. *et al.* GCK gene-body hypomethylation is associated with the risk of coronary heart
946 disease. *Biomed Res Int* **2014**, 151723 (2014). https://doi.org/10.1155/2014/151723
947 64 Pan, J. Q., Zhang, Y. Q., Wang, J. H., Xu, P. & Wang, W. lncRNA co-expression network model
948 for the prognostic analysis of acute myeloid leukemia. *Int J Mol Med* **39**, 663-671 (2017).
949 https://doi.org/10.3892/ijmm.2017.2888
950 65 Zou, W. *et al.* A metabolism-related 4-lncRNA prognostic signature and corresponding
951 mechanisms in intrahepatic cholangiocarcinoma. *BMC Cancer* **21**, 608 (2021).
952 https://doi.org/10.1186/s12885-021-08322-5
953 66 Kozlitina, J. *et al.* Exome-wide association study identifies a TM6SF2 variant that confers
954 susceptibility to nonalcoholic fatty liver disease. *Nat Genet* **46**, 352-356 (2014).
955 https://doi.org/10.1038/ng.2901
956 67 Tang, S. *et al.* Association of TM6SF2 rs58542926 T/C gene polymorphism with hepatocellular
957 carcinoma: a meta-analysis. *BMC Cancer* **19**, 1128 (2019). https://doi.org/10.1186/s12885-
958 019-6173-4
959 68 Bresgen, N., Kovacs, M., Lahnsteiner, A., Felder, T. K. & Rinnerthaler, M. The Janus-Faced
960 Role of Lipid Droplets in Aging: Insights from the Cellular Perspective. *Biomolecules* **13**
961 (2023). https://doi.org/10.3390/biom13060912
962 69 ENSEMBL. *https://www.ensembl.org/index.html*, 2013).
963 70 Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res* **12**, 996-1006 (2002).
964 https://doi.org/10.1101/gr.229102
965 71 Cancer Genome Atlas Research, N. *et al.* The Cancer Genome Atlas Pan-Cancer analysis
966 project. *Nat Genet* **45**, 1113-1120 (2013). https://doi.org/10.1038/ng.2764
967 72 Li, Z. *et al.* GepLiver: an integrative liver expression atlas spanning developmental stages and
968 liver disease phases. *Sci Data* **10**, 376 (2023). https://doi.org/10.1038/s41597-023-02257-1
969 73 Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol* **14**, R115
970 (2013). https://doi.org/10.1186/gb-2013-14-10-r115
971 74 Hannum, G. *et al.* Genome-wide methylation profiles reveal quantitative views of human
972 aging rates. *Mol Cell* **49**, 359-367 (2013). https://doi.org/10.1016/j.molcel.2012.10.016

973 75 Levine, M. E. *et al.* An epigenetic biomarker of aging for lifespan and healthspan. *Aging*
974 *(Albany NY)* **10**, 573-591 (2018). https://doi.org/10.18632/aging.101414
975 76 Jenjaroenpun, P., Wongsurawat, T., Yenamandra, S. P. & Kuznetsov, V. A. QmRLFS-finder: a
976 model, web server and stand-alone tool for prediction and analysis of R-loop forming
977 sequences. *Nucleic Acids Res* **43**, 10081 (2015). https://doi.org/10.1093/nar/gkv974
978 77 Hon, J., Martinek, T., Zendulka, J. & Lexa, M. pqsfinder: an exhaustive and imperfection-
979 tolerant search tool for potential quadruplex-forming sequences in R. *Bioinformatics* **33**,
980 3373-3379 (2017). https://doi.org/10.1093/bioinformatics/btx413
981 78 Marsico, G. *et al.* Whole genome experimental maps of DNA G-quadruplexes in multiple
982 species. *Nucleic Acids Res* **47**, 3862-3874 (2019). https://doi.org/10.1093/nar/gkz179
983 79 Li, C. *et al.* Ligand-induced native G-quadruplex stabilization impairs transcription initiation.
984 *Genome Res* **31**, 1546-1560 (2021). https://doi.org/10.1101/gr.275431.121
985 80 Kouzine, F. *et al.* Permanganate/S1 Nuclease Footprinting Reveals Non-B DNA Structures
986 with Regulatory Potential across a Mammalian Genome. *Cell Syst* **4**, 344-356 e347 (2017).
987 https://doi.org/10.1016/j.cels.2017.01.013
988 81 Lahnsteiner, A. *et al.* In vivo detection of DNA secondary structures using Permanganate/S1
989 Footprinting with Direct Adapter Ligation and Sequencing (PDAL-Seq). *Methods in*
990 *Enzymology* (2024). https://doi.org/https://doi.org/10.1016/bs.mie.2023.12.003.
991 82 ENCODE. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the
992 human genome. *Nature* **489**, 57-74 (2012). https://doi.org/10.1038/nature11247
993 83 Xu, Z. *et al.* Structural variants drive context-dependent oncogene activation in cancer.
994 *Nature* **612**, 564-572 (2022). https://doi.org/10.1038/s41586-022-05504-4
995 84 Wang, Y. *et al.* The 3D Genome Browser: a web-based browser for visualizing 3D genome
996 organization and long-range chromatin interactions. *Genome Biol* **19**, 151 (2018).
997 https://doi.org/10.1186/s13059-018-1519-9
998 85 Lou, Y. *et al.* Potential Regulators Driving the Transition in Nonalcoholic Fatty Liver Disease: a
999 Stage-Based View. *Cell Physiol Biochem* **41**, 239-251 (2017).
1000 https://doi.org/10.1159/000456061
1001 86 Zhang, W. *et al.* Fibroblast-specific adipocyte enhancer binding protein 1 is a potential
1002 pathological trigger and prognostic marker for liver fibrosis independent of etiology. *J Dig Dis*
1003 **24**, 550-561 (2023). https://doi.org/10.1111/1751-2980.13230
1004 87 Ho, C. Y., Lee, J. I., Huang, S. P., Chen, S. C. & Geng, J. H. A Genome-Wide Association Study of
1005 Metabolic Syndrome in the Taiwanese Population. *Nutrients* **16** (2023).
1006 https://doi.org/10.3390/nu16010077
1007 88 Kubickova, K. N. *et al.* Predictive role BLVRA mRNA expression in hepatocellular cancer. *Ann*
1008 *Hepatol* **15**, 881-887 (2016). https://doi.org/10.5604/16652681.1222104
1009 89 Tufan, N. L. *et al.* Hepatitis Bx antigen stimulates expression of a novel cellular gene, URG4,
1010 that promotes hepatocellular growth and survival. *Neoplasia* **4**, 355-368 (2002).
1011 https://doi.org/10.1038/sj.neo.7900241
1012 90 Dodurga, Y., Secme, M. & Lale Satiroglu-Tufan, N. A novel oncogene URG4/URGCP and its
1013 role in cancer. *Gene* **668**, 12-17 (2018). https://doi.org/10.1016/j.gene.2018.05.047
1014 91 Okuda, J. *et al.* Persistent overexpression of phosphoglycerate mutase, a glycolytic enzyme,
1015 modifies energy metabolism and reduces stress resistance of heart in mice. *PLoS One* **8**,
1016 e72173 (2013). https://doi.org/10.1371/journal.pone.0072173
1017 92 Di Costanzo, A. *et al.* Evaluation of Polygenic Determinants of Non-Alcoholic Fatty Liver
1018 Disease (NAFLD) By a Candidate Genes Resequencing Strategy. *Sci Rep* **8**, 3702 (2018).
1019 https://doi.org/10.1038/s41598-018-21939-0
1020 93 Zhou, Y. *et al.* UPF1 inhibits the hepatocellular carcinoma progression by targeting long non-
1021 coding RNA UCA1. *Sci Rep* **9**, 6652 (2019). https://doi.org/10.1038/s41598-019-43148-z
1022 94 Kim, M. J. *et al.* SUGP1 is a novel regulator of cholesterol metabolism. *Hum Mol Genet* **25**,
1023 3106-3116 (2016). https://doi.org/10.1093/hmg/ddw151
1024 95 Loyfer, N. *et al.* A DNA methylation atlas of normal human cell types. *Nature* **613**, 355-364
1025 (2023). https://doi.org/10.1038/s41586-022-05580-6

96 Huo, Z. *et al.* DNA methylation variability in Alzheimer's disease. *Neurobiol Aging* **76**, 35-44 (2019). https://doi.org/10.1016/j.neurobiolaging.2018.12.003

97 Faltus, C. *et al.* Identification of NHLRC1 as a Novel AKT Activator from a Lung Cancer Epigenome-Wide Association Study (EWAS). *Int J Mol Sci* **23** (2022). https://doi.org/10.3390/ijms231810699

98 Ilieva, M. *et al.* Systematic Analysis of Long Non-Coding RNA Genes in Nonalcoholic Fatty Liver Disease. *Noncoding RNA* **8** (2022). https://doi.org/10.3390/ncrna8040056

99 Perrin-Cocon, L. *et al.* A hexokinase isoenzyme switch in human liver cancer cells promotes lipogenesis and enhances innate immunity. *Commun Biol* **4**, 217 (2021). https://doi.org/10.1038/s42003-021-01749-3

100 Cao, X., Mahendran, R., Guy, G. R. & Tan, Y. H. Detection and characterization of cellular EGR-1 binding to its recognition site. *J Biol Chem* **268**, 16949-16957 (1993).

101 Brazda, V. *et al.* The Amino Acid Composition of Quadruplex Binding Proteins Reveals a Shared Motif and Predicts New Potential Quadruplex Interactors. *Molecules* **23** (2018). https://doi.org/10.3390/molecules23092341

102 Sun, Z. *et al.* EGR1 recruits TET1 to shape the brain methylome during development and upon neuronal activity. *Nat Commun* **10**, 3892 (2019). https://doi.org/10.1038/s41467-019-11905-3

103 Yadav, S. *et al.* ERK1/2-EGR1-SRSF10 Axis Mediated Alternative Splicing Plays a Critical Role in Head and Neck Cancer. *Front Cell Dev Biol* **9**, 713661 (2021). https://doi.org/10.3389/fcell.2021.713661

104 Trizzino, M. *et al.* EGR1 is a gatekeeper of inflammatory enhancers in human macrophages. *Sci Adv* **7** (2021). https://doi.org/10.1126/sciadv.aaz8836

105 Guo, Y. *et al.* Zinc finger transcription factor Egf1 promotes non-alcoholic fatty liver disease. *JHEP Rep* **5**, 100724 (2023). https://doi.org/10.1016/j.jhepr.2023.100724

106 Le Senechal, R. *et al.* Alternative splicing of BCL-x is controlled by RBM25 binding to a G-quadruplex in BCL-x pre-mRNA. *Nucleic Acids Res* **51**, 11239-11257 (2023). https://doi.org/10.1093/nar/gkad772

107 Shukla, S. *et al.* CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**, 74-79 (2011). https://doi.org/10.1038/nature10442

108 Shayevitch, R., Askayo, D., Keydar, I. & Ast, G. The importance of DNA methylation of exons on alternative splicing. *RNA* **24**, 1351-1362 (2018). https://doi.org/10.1261/rna.064865.117

109 Hadar, S., Meller, A., Saida, N. & Shalgi, R. Stress-induced transcriptional readthrough into neighboring genes is linked to intron retention. *iScience* **25**, 105543 (2022). https://doi.org/10.1016/j.isci.2022.105543

110 Zyner, K. G. *et al.* G-quadruplex DNA structures in human stem cells and differentiation. *Nat Commun* **13**, 142 (2022). https://doi.org/10.1038/s41467-021-27719-1

111 Biffi, G., Tannahill, D., Miller, J., Howat, W. J. & Balasubramanian, S. Elevated levels of G-quadruplex formation in human stomach and liver cancer tissues. *PLoS One* **9**, e102711 (2014). https://doi.org/10.1371/journal.pone.0102711

112 Hansel-Hertsch, R. *et al.* G-quadruplex structures mark human regulatory chromatin. *Nat Genet* **48**, 1267-1272 (2016). https://doi.org/10.1038/ng.3662

113 Lago, S. *et al.* Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. *Nat Commun* **12**, 3885 (2021). https://doi.org/10.1038/s41467-021-24198-2

114 Figueiredo, J., Mergny, J. L. & Cruz, C. G-quadruplex ligands in cancer therapy: Progress, challenges, and clinical perspectives. *Life Sci* **340**, 122481 (2024). https://doi.org/10.1016/j.lfs.2024.122481

115 Grundy, S. M. *et al.* Diagnosis and management of the metabolic syndrome: an American Heart Association/National Heart, Lung, and Blood Institute scientific statement: Executive Summary. *Crit Pathw Cardiol* **4**, 198-203 (2005). https://doi.org/10.1097/00132577-200512000-00018

116 Team, R. C. R: A Language and Environment for Statistical Computing. (2023).

117 Team, R. S. RStudio: Integrated Development for R. (2020).
118 Vienna, U. o. RNAfold Web Server. (2024).
119 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009). https://doi.org/10.1093/bioinformatics/btp324
120 Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26 (2011). https://doi.org/10.1038/nbt.1754
121 Jalili, V. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Res* **48**, W395-W402 (2020). https://doi.org/10.1093/nar/gkaa434
122 Zheng, K. W., Chen, Z., Hao, Y. H. & Tan, Z. Molecular crowding creates an essential environment for the formation of stable G-quadruplexes in long double-stranded DNA. *Nucleic Acids Res* **38**, 327-338 (2010). https://doi.org/10.1093/nar/gkp898
123 Lam, E. Y., Beraldi, D., Tannahill, D. & Balasubramanian, S. G-quadruplex structures are stable and detectable in human genomic DNA. *Nat Commun* **4**, 1796 (2013). https://doi.org/10.1038/ncomms2792
124 Li, Y., Ge, D. & Lu, C. The SMART App: an interactive web application for comprehensive DNA methylation analysis and visualization. *Epigenetics Chromatin* **12**, 71 (2019). https://doi.org/10.1186/s13072-019-0316-3