STAT5 regulation of sex-dependent hepatic CpG methylation at distal regulatory elements mapping to sex-biased genes

Pengying Hao^a and David J. Waxman^{a, #}

^a Department of Biology and Bioinformatics Program, Boston University, Boston, MA 02215

Running title: Liver STAT5-dependent CpG methylation

Keywords: RRBS, enhancers, CpG methylation, sex-bias, DNase hypersensitive site

Address correspondence to:

David J. Waxman

Dept. of Biology, Boston University

5 Cummington Mall

Boston, MA 02215

Email: djw@bu.edu

Word counts:

Abstract (**199**)

Other text (**7,828**), excluding Materials and Methods, Tables and References

34 **Abstract**

35 Growth hormone-activated STAT5b is an essential regulator of sex-differential gene expression in
 36 mouse liver, however, its impact on hepatic gene expression and epigenetic responses is poorly
 37 understood. Here, we found a substantial, albeit incomplete loss of liver sex bias in hepatocyte-specific
 38 STAT5a/STAT5b (collectively, STAT5)-deficient mouse liver. In male liver, many male-biased genes
 39 were down regulated in direct association with the loss of STAT5 binding; many female-biased genes,
 40 which show low STAT5 binding, were de-repressed, indicating an indirect mechanism for repression by
 41 STAT5. Extensive changes in CpG-methylation were seen in STAT5-deficient liver, where sex
 42 differences in DNA methylation were abolished at 88% of ~1,500 differentially-methylated regions,
 43 largely due to an increase in methylation at the hypomethylated sites. STAT5-dependent CpG-
 44 hypomethylation was rarely found at proximal promoters of STAT5-dependent genes. Rather, STAT5
 45 primarily regulated the methylation of distal enhancers, where STAT5 deficiency induced widespread
 46 hypermethylation at genomic regions enriched for accessible chromatin, enhancer histone marks
 47 (H3K4me1, H3K27ac), STAT5 binding, and DNA motifs for STAT5 and other transcription factors
 48 implicated in liver sex differences. In conclusion, the sex-dependent binding of STAT5 to liver chromatin
 49 is closely linked to sex-dependent demethylation of distal regulatory elements mapping to STAT5-
 50 dependent genes important for liver sex bias.

51

52 **Introduction**

53

54 The mammalian STAT gene family encodes seven latent cytoplasmic transcription factors activated by
 55 cell surface receptor-induced tyrosine phosphorylation, which stimulates STAT protein dimerization and
 56 nuclear translocation, followed by DNA binding and transcriptional activation of STAT target genes (1).
 57 One STAT gene, STAT5a, is highly expressed in mammary tissues, where it is activated by prolactin (2,
 58 3), while the closely related (96% similar) STAT5b is particularly abundant in hepatocytes, where it is
 59 activated by growth hormone (GH) and regulates lipid metabolism (4-6) as well as body growth, in part
 60 through production of IGF-1 (7). STAT5b is also a key mediator of the sex differential transcriptional

61 networks that GH regulates in the liver (8, 9), as seen in a STAT5b whole body knockout mouse model
 62 (10). GH activates STAT5b signaling by binding to its liver cell surface receptor, which in turn activates
 63 the receptor-associated tyrosine kinase JAK2, stimulating phosphorylation of GH receptor on multiples
 64 intracellular tyrosine residues, thereby creating docking sites for downstream signaling proteins,
 65 including STAT5b. JAK2-catalyzed phosphorylation of STAT5b on Tyr-699 enables STAT5b to dimerize
 66 and then translocate to the nucleus (11), where it binds to STAT5 motifs enriched at open chromatin
 67 regions and stimulates gene transcription (12).

68
 69 Liver STAT5 activity is highly responsive to GH stimulation (13). In many mammalian species, including
 70 humans (14), rats and mice (8, 15), the temporal pattern of pituitary GH secretion differs between males
 71 and females. In male mice and rats, pituitary GH secretion is episodic, with strong pulses every 3-4 h
 72 followed by a GH-free interval, whereas in females, pituitary GH secretion is more frequent, resulting in
 73 a persistent (near continuous) plasma GH profile. Liver STAT5 activity mirrors the sex differences in
 74 circulating GH patterns, and consequently, it oscillates between active and inactive states in male liver,
 75 but is more persistently active in female liver (12, 16, 17). These sex-dependent temporal differences in
 76 liver STAT5 activity are the dominant determinant of the sex-biased expression of hundreds of genes in
 77 mouse and rat liver, including many cytochromes P450 and other enzymes of steroid and drug
 78 metabolism, as was demonstrated in mice with a global deficiency in STAT5b (10), and are an important
 79 factor in sex-differences in drug and steroid metabolism and disease susceptibility (8, 18). STAT5a
 80 deficiency can also impact sex-biased gene expression in the liver, albeit to a much lesser extent than
 81 STAT5b, consistent with the much lower levels of STAT5a present in liver (19). Perturbation of liver
 82 STAT5 activity by altering plasma GH patterns by hypophysectomy (16, 20) or by continuous GH
 83 infusion (21, 22) largely abolishes sex differences in gene expression, including the expression of sex-
 84 dependent regulatory micro-RNAs (23) and lncRNAs (24, 25). Importantly, while exogenous GH
 85 replacement can restore sex-specific expression of *Cyp* and other genes in livers of hypophysectomized
 86 mice, this restoration is not achieved in livers of hypophysectomized STAT5b knockout mice (22, 26).
 87 While these studies demonstrate the essential nature of STAT5b for sex differentiated liver gene

88 expression, they do not distinguish effects due to the loss of STAT5b in liver *per se* versus effects due to
 89 loss of STAT5 expression in other tissues, including the hypothalamus, which can impact plasma GH
 90 profiles and liver responses to GH (27). These issues are addressed here, where we use RNA-seq to
 91 evaluate the impact of hepatocyte-specific deletion of the *Stat5a/Stat5b* locus (28) on the liver
 92 transcriptome in both male and female mice, and its relationship to sex-specific binding sites for STAT5
 93 (12, 29) associated with regulation of sex-biased gene expression in the liver.

94
 95 DNA methyltransferase-catalyzed cytosine methylation at the C-5 position is a heritable epigenetic mark
 96 in higher vertebrates and is key for many biological processes, including X chromosome inactivation and
 97 imprinting (30, 31). A large majority of DNA methylation occurs in the context of CpG dinucleotides in
 98 mammalian somatic cells, with most CpG sites being highly methylated, except for CpGs within gene
 99 promoters. These patterns of DNA methylation are stably propagated in terminally differentiated cells
 100 throughout life; however, changes in CpG methylation can occur in response to hormonal and other
 101 environmental factors (32, 33) and diseased states (30). DNA methylation generally leads to gene
 102 silencing and chromatin condensation, although DNA methylation can also stimulate gene expression
 103 (34). Sex-differences in DNA methylation have been reported for mouse liver (35-37) and may, in part,
 104 be determined by CpG demethylation associated with testosterone exposure at puberty (38-40). It is
 105 unclear, however, what role STAT5 might play in regulating these sex differences in CpG methylation, or
 106 how they may relate to the STAT5-dependent regulation of sex-specific gene expression. Here, we use
 107 reduced representation bisulfite sequencing (41) to characterize the effects of STAT5 on CpG
 108 methylation, including sex differences in liver CpG methylation. Our findings establish that liver STAT5
 109 deficiency abolishes the sex bias in DNA methylation at more than 1,300 genomic sites, and that
 110 STAT5-dependent sex-biased genes are frequently regulated by STAT5-regulated demethylation
 111 associated with the activation of distal enhancers.

112

113 Results

114

115 Impact of hepatocyte-specific STAT5 loss on sex-biased liver gene expression. Global inactivation
116 of the mouse *Stat5b* gene leads to widespread loss of sex-biased gene expression in the liver (10). As
117 the observed changes in liver gene expression could, in part, be due to extrahepatic actions of STAT5b
118 on pituitary GH secretion (26), which is a major regulator of sex-specific gene expression in the liver (21),
119 we used RNA-seq to determine the impact of hepatocyte-specific deletion of the *Stat5a/Stat5b* locus
120 (STAT5-KO) (28, 42) on sex-biased gene expression. Differential expression analysis was carried out
121 using liver RNA purified from wild-type male (WTM) and female (WTF) floxed control mice, and from
122 STAT5-KO male (KOM) and female (KOF) mice. A total of 274 liver-expressed genes showed
123 differential expression between male and female controls (WTM/WTF comparison, FDR < 0.05): 132
124 male-biased and 142 female-biased genes. Genes that were responsive to STAT5-KO in male liver (365
125 genes) and in female liver (278 genes) were identified in two additional comparisons, KOM/WTM and
126 KOF/WTF (Table S1). The gene overlap across the three comparisons is presented in **Fig. 1A**. 155 of all
127 274 sex-biased genes (57%) responded to STAT5 deficiency in either male or female liver vs. only 2.2%
128 of stringently sex-independent genes (**Table 1**). Male-biased genes whose expression was altered were
129 primarily down regulated in STAT5-KO male liver or were up regulated in STAT5-KO female liver.
130 Conversely, female-biased genes whose expression was altered were primarily up regulated in STAT5-
131 KO male liver or were down regulated in STAT5-KO female liver (**Fig. 1B** and **Table 1**). This relationship
132 between sex specificity and response to STAT5 deficiency is visualized in quantitative scatter plots
133 showing highly significant differences in the response of male-biased and female-biased genes to
134 STAT5-KO loss in male liver, and to a lesser degree in female liver (**Fig. 1C**). The extent of sex bias and
135 magnitude of response to STAT5 deficiency were significantly correlated in male liver ($r^2=0.4$, $p = 1.2E-$
136 13), but not in female liver. Thus, STAT5 acts at the level of the hepatocyte to regulate sex-biased gene
137 expression *via* both activating and repressive mechanisms, strongly in male liver and weakly in female
138 liver.
139
140 Sex-biased expression was lost in STAT5-KO liver for 168 (61%) of the 274 sex-biased genes, as
141 judged by comparison of male and female STAT5-KO livers (KOM/KOF comparison; Table S1A).

Moreover, sex bias was reduced in the STAT5-KO strain for 80% of the 106 genes that did not meet our cutoff for loss of sex-biased expression. Further supporting our conclusion that STAT5 deficiency impacts the vast majority of sex-biased genes, we found that 81 of the 106 genes that were apparently unresponsive to STAT5-KO actually showed the same overall trend in their response to STAT5 deficiency in male liver as was seen for the 168 significantly responsive sex-biased genes, namely, 43 of 53 male-biased genes decreased in expression, and 38 of 53 female-biased genes increased in expression. Of note, strong female-specific expression was retained in STAT5-KO liver for the GH-regulated transcriptional repressor *Cux2* (43, 44). The absence of an effect on *Cux2* expression may in part explain the incomplete loss of sex-biased gene expression in STAT5-KO livers, given the widespread role of CUX2 in mediating GH-dependent regulation of sex-biased genes in mouse liver (43). Other highly female-biased genes (WTF/WTM >4) that appeared to be truly unresponsive to STAT5 loss (KOM/WTM |fold-change| < 1.2) include *Cyp3a16*, *Hsd3b1*, *Ugt2b37* and *Nipal1* (Table S1B).

STAT5 binding contributes to sex-biased gene expression and STAT5 response. STAT5 binds to liver chromatin in a sex-differential manner nearby a subset of sex-biased genes, as determined by ChIP-seq analysis (12). Here, we examined the set of 98 of 274 sex-biased genes (36%) with a nearby (within 10 kb of gene body) sex-biased STAT5 ChIP-seq peak (Table S1A). Female-biased genes were significantly associated with female-biased STAT5 ChIP-seq peaks, and male-biased genes with male-biased STAT5 ChIP-seq peaks ($p < 0.0001$, Fisher's Exact test) (**Fig. 1D, Fig. 1E, left panels**), as expected (12). These positive associations indicate that proximal sex-biased STAT5 binding imparts sex-biased expression, primarily by activation of gene expression. Furthermore, genes down regulated in STAT5-KO male liver were significantly enriched for having proximal male-biased STAT5 ChIP-seq peaks in wild-type liver, consistent with gene down regulation being due to a direct loss of the transcriptional stimulatory effects of STAT5 (**Fig. 1D, Fig. 1E, middle panels**). In contrast, genes up regulated in STAT5-KO male liver were significantly enriched for proximal female-biased STAT5 ChIP-seq peaks, i.e., STAT5 binding nearby those genes is low in male liver as compared to female liver. This indicates that the STAT5-dependent repression of female-biased genes in wild-type male liver is indirect,

169 e.g., is mediated by a STAT5-dependent repressor whose loss in STAT5-KO liver leads to the observed
170 gene de-repression. The association between gene response to STAT5 deficiency and proximal sex-
171 biased STAT5 ChIP-seq peaks was weaker in female liver (**Fig. 1D, Fig. 1E, right panels**). STAT5-
172 dependent sex-biased genes that did not show any proximal STAT5 binding (Table S1A) may be
173 regulated by distal STAT5 binding (29).

174

175 **Sex-independent genes that respond to STAT5 deficiency.** Analysis of sex-independent genes
176 responsive to STAT5-KO gave additional insights into the biological changes occurring in this mouse
177 model. We identified 338 sex-independent genes whose expression was significantly altered in STAT5-
178 KO liver (**Fig. 1A**, Venn regions D, F, G). Many of these genes responded to STAT5 deficiency in the
179 same manner in both sexes, as seen in a heat map displaying 140 of the 338 genes, whose expression
180 was stringently sex-independent in wild-type liver (**Fig. 2A**). Examples include the classic STAT5 target
181 genes *Igf1*, *Socs2*, and *Onecut1* (HNF6), whose expression was down regulated in livers of both sexes.
182 Pathway analysis identified immunity, cellular response to interferon- β , and peptide antigen
183 binding/MHC as the top enriched pathways for the subset of 71 genes that were induced in common in
184 male and female STAT5-KO livers (**Fig. 2B, top**). This contrasts to the strong enrichment of cytochrome
185 P450 and drug and steroid metabolism genes in the set of 155 STAT5-dependent sex-specific genes
186 (**Fig. 2B, bottom**). The transcription factor STAT1 was among the genes up regulated in both males and
187 females (**Fig. 2C**), consistent with (28), and may contribute to the observed increases in immune
188 signaling and interferon response gene expression (45). Some residual expression of STAT5b was seen
189 in STAT5-KO liver (**Fig. 2C**), reflecting its continued expression in non-parenchymal cells (42), where
190 the *Alb*-Cre transgene used to excise the STAT5a-TAT5b locus is not active (46).

191

192 **Impact of STAT5 knockout on DNA methylation in male and female mouse liver.** The methylation
193 of state of CpG dinucleotides is a key epigenetic mark regulating chromatin structure and gene
194 expression. Given the major impact of liver STAT5 deficiency on sex-biased gene expression, shown
195 above, we carried out reduced representation bisulfite sequencing (RRBS) to capture and analyze the

196 methylation status of genomic regions with high CpG content in wild-type and STAT5-KO mouse liver.
197 We used the same set of livers as in the RNA-seq analysis, but analyzed a larger number of individual
198 livers (biological replicates; N=5) to increase statistical power. We identified genomic regions that show
199 significantly differential CpG methylation (DMRs, differentially methylated regions), either between male
200 and female wild-type mouse liver (sex-biased DMRs), or between STAT5-KO and wild-type mouse liver
201 (STAT5-dependent DMRs) (Table S2).

202
203 Each DMR covers a 100 bp genomic region ('tile'), which may be either hypermethylated or
204 hypomethylated in the treatment sample compared to its control (e.g., male compared to female, or
205 STAT5-KO compared to wild-type). The extent of differential methylation was quantified as a differential
206 methylation (diff.meth) value, corresponding to the difference in percent methylation between the
207 samples being compared (**Fig. 3A**). Hypermethylation of the treatment relative to control is indicated by
208 a positive diff.meth value; and hypomethylation, where treatment is demethylated relative to control, is
209 indicated by a negative diff.meth value. DMRs with diff.meth between |15%| and |25%| at FDR < 0.05
210 were considered significant at standard stringency; diff.meth > |25%| at FDR < 0.05 identified robust
211 DMRs, as shown in **Fig. 3A** for each of four pair-wise comparisons. DMRs sensitive to the loss of
212 STAT5 were more frequently associated with hypermethylation than hypomethylation when compared to
213 wild-type liver, most notably in males (Hyper/Hypo methylation ratio = 2.14 for KOM/WTM vs. 0.9-1.01
214 for WTM/WTF and KOM/KOF). The trend of increased hypermethylation with STAT5 deletion was also
215 apparent by boxplot analysis (**Fig. 3B**). Furthermore, unsupervised (hierarchical) clustering of raw
216 methylation values for all robust DMRs identified in the KOM/WTM comparison correctly separated all of
217 the male from female wild-type individuals, and all but one of the STAT5-KO from wild-type individuals
218 (**Fig. 3C**, heat map).

219
220 Of the 1493 sex-differential DMRs seen in wild-type liver, 738 (49%) showed significant differential
221 methylation between STAT5-KO and wild-type livers, i.e., are STAT5-responsive DMRs, in either males
222 or females (**Fig. 3D**; sections A+B+C; **Fig. 3F**). Moreover, STAT5 deficiency caused a loss of significant

223 sex-differences in CpG methylation at 1310 (88%) of the 1493 sex-differential DMRs (WTM/WTF vs
224 KOM/KOF comparison; **Fig. 3E**). In male STAT5-KO liver, DMRs with a significantly lower methylation
225 level in wild-type male compared to wild-type female liver (i.e., male-hypomethylated DMRs) primarily
226 became hypermethylated, while many male-hypermethylated DMRs became hypomethylated (**Fig. 3F**
227 and **Table 2**). In contrast, in female STAT5-KO liver, many female-hypermethylated DMRs (i.e., male-
228 hypomethylated DMRs) became hypomethylated, and female-hypomethylated DMRs became
229 hypermethylated. **Fig. 3F** shows that STAT5-KO induces a loss of sex-differential methylation at male-
230 hypomethylated sites by increasing methylation in male liver or by decreasing methylation in female liver.
231 In the case of female-hypomethylated sites, STAT5-KO induces a loss of sex-differential methylation by
232 increasing methylation in female liver or by decreasing methylation in male liver. Quantitative scatter
233 plots illustrate this significant reversal of sex-biased differential DNA methylation in the absence of liver
234 STAT5 (**Fig. 3G**). Further, the extent of methylation differences between sexes correlated significantly
235 with the level of DNA methylation changes in the STAT5-KO livers (**Fig. 3G**). Thus, liver STAT5 is
236 required for sex-specific methylation and demethylation at several hundred genomic regions.

237
238 **Few DMRs are in promoter regions of sex-biased, STAT5-dependent genes.** CpG islands, found at
239 many gene promoters, generally exhibit low methylation levels and are correlated with active
240 transcription (47). Further, hypermethylation of CpG islands is associated with stable silencing of genes
241 (48). To determine if promoter CpG methylation contributes to the sex-biased and STAT5-dependent
242 gene expression changes we observed, we examined DMRs proximal (5 kb upstream or downstream) to
243 the TSS of sex-biased, STAT5-responsive genes. Very few such sex-biased genes have a proximal sex-
244 biased DMR (**Fig. 4A**). Further, only two STAT5-dependent, male-biased genes, *Cyp7b1* and *Cspg5*,
245 showed significant promoter DNA methylation changes in STAT5-KO male liver that matched the
246 general trends described above (**Fig. 4B, Fig. 4C**, Table S3A). For these two genes, DNA methylation in
247 the promoter region was significantly lower in wild-type male than female liver, and was increased in
248 STAT5-KO male liver in association with the decrease in gene expression. Thus, the large majority of
249 sex-biased, STAT5-dependent genes do not display sex differences in promoter region DNA methylation.

250 We also identified examples of sex-independent STAT5-KO responsive genes that showed STAT5-
 251 dependent changes in promoter region DNA methylation in a direction consistent with the change in
 252 their expression in STAT5-deficient male and/or female liver (Table S3B).
 253
 254 **Chromatin states and chromatin accessibility at DMRs.** To determine the chromatin context of
 255 STAT5-dependent DNA methylation, we mapped each set of DMRs to 14 ChromHMM-derived
 256 chromatin states (**Fig. 5A**) previously learned for male and female mouse liver using genome-wide
 257 datasets for open chromatin regions (DHS, DNase-I hypersensitive sites) and histone marks (29). A
 258 substantial fraction of the >1.3 million CpGs captured by our RRBS samples were within a promoter
 259 state, reflecting the preferential capture of promoter CpG island sequences by the RRBS protocol.
 260 Promoter states show low CpG methylation overall, while enhancer states show a large range of CpG
 261 methylation levels (Fig. S2). CpGs in sex-biased and STAT5-responsive DMRs were significantly
 262 depleted of promoter states (blue) and showed various levels of enrichment for enhancer states (green)
 263 (**Fig. 5B, Fig. 5C**). Frequency profiles of the 14 chromatin states in a 4-kb window surrounding each
 264 DMR revealed specific enrichment of enhancer state 6 for DMRs that were hypermethylated, but not
 265 hypomethylated, in the absence of STAT5 (KOM/WTM and KOF/WTF comparisons) (**Fig. 5D**). A
 266 modest enrichment of enhancer state 6 was also seen in DMRs hypomethylated in wild-type male
 267 relative to wild-type female liver. Enhancer state 6 represents active enhancers characterized by high
 268 probabilities of H3K4me1 and H3K27ac histone marks and presence of a DHS (**Fig. 5A**). Thus, loss of
 269 STAT5 leads to a preferential hypermethylation of genomic regions in an active enhancer state.
 270
 271 **TAD-based analysis of association of sex-biased and STAT5-dependent genes with distal DMRs.**
 272 Next, we investigated whether the enhancer state-associated DMRs can be linked to STAT5-dependent
 273 transcriptional regulation. As it is difficult to directly link a given enhancer to its specific gene target(s),
 274 we evaluated such interactions at the level of topologically associating domains (TADs), which are
 275 megabase-scale genomic regions with high propensity for intra-domain contact in three-dimensional
 276 chromatin organization (49). Using TAD boundaries defined for mouse liver (50, 51), we investigated

whether genes that are sex-biased or STAT5-responsive are more likely to be in the same TAD as a sex-biased or STAT5-responsive DMR, when compared to sets of random CpG regions. DMRs, grouped by chromatin state (**Fig. 5B**), were mapped to TADs, and their overlap with the set of 437 TADs containing at least one sex-biased or STAT5-responsive gene was then determined (Table S4). The statistical significance of TAD overlap was determined for each chromatin state by comparison to TAD overlap for a random sampling, performed 50 times, of comparably sized sets of 100 bp RRBS-captured CpG regions that did not show significant differential methylation. DMRs in chromatin states 6 and 9, but not DMRs in other chromatin states, showed significantly higher co-occupancy with TADs containing sex-biased or STAT5-responsive genes than did random CpG regions (z-score: 2.54 and 2.98; p-value: 0.011 and 0.004, for chromatin states 6 and 9, respectively) (**Fig. 5E**, Table S4C). Chromatin states 6 and 9 are enhancer states with high levels of histone H3K4me1 and histone H3K27ac marks, either with (state 6) or without (state 9) a DNase hypersensitive site (DHS) (**Fig. 5A**), with the DHS-free enhancer state 9 showing much higher overall CpG methylation than enhancer state 6 (Fig. S2). These enrichments support the proposal that the regulation of CpG methylation at these gene-distal enhancers plays a regulatory role for the sex-biased and STAT5-responsive genes within these TADs. Consistent with this proposal, the sex-biased, STAT5-responsive DMRs in chromatin states 6 and 9 were closer in median distance to sex-biased, STAT5-responsive genes than DMRs in other chromatin states (**Fig. 5F**).

Association of DMRs with nearby open chromatin regions – Many DMRs were located nearby at least one liver DHS (Fig. S3). These DMR-proximal DHS include 308 male-biased DHS and 66 female-biased DHS identified previously (52) (Table S5). Gene set enrichment analysis showed that these 374 sex-biased DHS were non-randomly distributed: the 308 male-biased DHS were associated with DMRs that were hypomethylated in wild-type male liver (normalized enrichment score, NES: -3.18, FDR=0), and the 66 female-biased DHS were associated with DMRs hypermethylated in wild-type male liver (and consequently, hypomethylated in wild-type female liver) (NES=1.57, FDR=0.007) (**Fig. 6**, *left panels*). Thus, sex-biased CpG demethylation is associated with sex-biased chromatin opening in wild-type livers of both sexes. Moreover, in male liver, loss of STAT5 resulted in significant hypermethylation in genomic

304 regions with a male-biased DHS, but not in region with female-biased DHS (*middle* panels), whereas the
 305 absence of STAT5 in female liver resulted in hypermethylation at sites nearby both male and female
 306 DHS (*right* panels). Thus, sex-biased chromatin opening is associated with sex-biased CpG
 307 demethylation, and loss of liver STAT5 enables hypermethylation of nearby sex-biased DHS, consistent
 308 with the enrichment of the DHS-containing enhancer state 6 seen in **Fig. 5D**.

309
 310 **STAT5 binding nearby DMRs** – Given our finding that the loss of STAT5 in male liver increases
 311 methylation at male-biased DHS regions associated with hypomethylated DMRs, we scanned each set
 312 of DMRs (WTM/WTF; KOM/WTM; KOF/WTF) for STAT5 motifs. We found strong enrichment of the
 313 STAT5 motif at hypermethylated DMRs in STAT5-KO male liver, and to a lesser extent in STAT5-KO
 314 female liver, but not at the corresponding hypomethylated DMRs, consistent with hypermethylation
 315 occurring at or nearby STAT5 binding sites (**Fig. 7A**). Next, we examined whether STAT5 ChIP-seq
 316 binding sites identified in wild-type male and female liver (12) are enriched in the set of DMRs
 317 hypermethylated in STAT5-KO liver. Overall, 59% of all liver STAT5 ChIP-seq peaks were nearby one or
 318 more of the CpG sites covered by our RRBS dataset (**Fig. 7B**). Cumulative frequency plots of distance
 319 to the nearest STAT5 binding site (**Fig. 7C**) showed that a much higher fraction of DMRs that were
 320 hypermethylated in STAT5-KO liver were directly at (y-intercept values; solid lines) or nearby a STAT5
 321 binding site as compared to DMRs that were hypomethylated (dashed lines), consistent with the STAT5
 322 motif analysis. When considering STAT5 binding sites within a few kb of a DMR, male-biased STAT5
 323 binding sites showed greater enrichment for DMRs hypermethylated by STAT5-KO in male liver than
 324 those hypermethylated in STAT5-KO female liver (**Fig. 7C, middle**), while female-biased STAT5 binding
 325 was most enriched in hypermethylated DMRs found in STAT5-KO female liver (**Fig. 7C, bottom**). Further,
 326 male-biased STAT5 binding sites (and to a lesser extent sex-independent STAT5 sites) were
 327 preferentially enriched at DMRs hypomethylated in male liver (*middle*, purple dashed line). We conclude
 328 that STAT5 binding sites are preferentially hypomethylated in wild-type male, but not female liver, and
 329 that the loss of STAT5 increases CpG methylation proximal to STAT5 binding in both male and female
 330 liver.

331

332 **Motifs enriched at hypermethylated DMRs** – Next, we used *de novo* motif discovery, implemented
 333 using Homer (53), to determine whether motifs for other transcription factors are enriched in the sets of
 334 sex- or STAT5-dependent hypermethylated and hypomethylated DMRs described above. The largest
 335 number of significantly enriched motifs ($n = 47$, q -value < 0.05) was discovered in the set of
 336 hypermethylated DMRs from the KOM/WTM comparison, followed by 13 motifs for hypermethylated
 337 DMRs from the KOF/WTF comparison (**Fig. 8A**). Analysis of the overlaps of enriched motifs between
 338 datasets revealed that a majority of the hypermethylated KOF/WTF motifs overlapped the set of
 339 hypermethylated KOM/WTM motifs (**Fig. 8B**). The top motifs discovered are shown in **Fig. 8C**. Notably,
 340 motifs for STAT5, HNF6 and CUX2, all previously implicated in regulation of liver sex differences (29, 43,
 341 44), were unique to the sets of hypermethylated DMRs common to the KOM/WTM and KOF/WTF
 342 comparisons. Thus, loss of STAT5 resulted in hypermethylation of regions enriched in DNA motifs for
 343 STAT5 and its co-regulating transcription factors. These regions are enriched for enhancer histone
 344 marks and chromatin states (**Fig. 5**), DHS (**Fig. 6**), and STAT5 binding (**Fig. 7**) in wild-type liver, and
 345 their presumed loss of activity upon hypermethylation (54, 55) is expected to contribute to the loss of
 346 sex-biased expression of their associated genes.

347

348 **Discussion**

349 We investigated the relationship between sex-biased DNA methylation and sex-biased liver gene
 350 expression in mice with a hepatocyte-specific deletion of STAT5, which plays a central role in GH
 351 regulation of adult liver sexual differentiation. Loss of STAT5 in hepatocytes led to a substantial, albeit
 352 incomplete loss of sex-biased liver gene expression, with STAT5 inducing male-biased genes in male
 353 liver via a positive, STAT5-binding-dependent regulatory mechanism, and with STAT5 repressing
 354 female-biased genes in male liver by an indirect regulatory mechanism. Liver STAT5 deficiency also
 355 resulted in widespread changes in DNA methylation, abolishing the sex bias in DNA methylation at 88%
 356 of 1,493 sex-biased DMRs. Hypermethylated regions identified in male and female STAT5-KO liver were
 357 enriched for open chromatin regions (DHS), enhancer chromatin states, and STAT5 ChIP-seq peaks in

358 wild-type liver, and encompassed transcription factor motifs for STAT5 and other transcription factors
 359 previously implicated as regulators of liver sex bias. Thus, many DMRs that become hypermethylated in
 360 STAT5-KO liver represent STAT5-bound enhancers that require STAT5 to maintain a low DNA
 361 methylation state. As DNA methylation is often associated with condensed chromatin that is inaccessible
 362 to transcription factor binding (56), our findings suggest that the loss of STAT5 leads to silencing of
 363 these regulatory elements via CpG methylation. Proximal promoter regions of STAT5-responsive genes
 364 showed very few significant changes in DNA methylation in STAT5-KO liver, and correspondingly,
 365 promoter states were significantly under-represented, especially in the sets of DMRs that become
 366 hypomethylated in the absence of STAT5. The later DMRs may represent latent regulatory elements
 367 that become activated in the absence of STAT5, e.g., due to the up regulation of STAT1 (**Fig. 2**), or by
 368 other compensatory mechanisms that have yet to be identified.

369
 370 The present findings in hepatocyte-specific STAT5a/STAT5b-KO liver differ somewhat from prior
 371 findings in whole body STAT5b-KO liver based on microarray analysis (10), where liver sex bias was
 372 abolished for a larger fraction of sex-biased genes (78% vs. 61%). However, a full 80% of genes that
 373 apparently retained sex-biased expression in the present hepatocyte-specific STAT5-KO model actually
 374 showed a decrease in the magnitude of their sex bias in the absence of STAT5. Furthermore, and
 375 consistent with our prior work (10), the impact of hepatocyte-specific STAT5 loss was less pronounced
 376 in female liver, where fewer genes were dysregulated, and where there was a weaker correlation
 377 between sex specificity and STAT5 response and a weaker association with sex-biased proximal STAT5
 378 binding. Nevertheless, hepatocyte-specific STAT5 deficiency did impact sex-biased gene expression in
 379 female liver, consistent with the female-biased binding of STAT5 to liver chromatin seen by ChIP-seq
 380 (12) and with the finding that STAT5a is specifically required for expression of 15% of female-specific
 381 genes in female liver (57).

382
 383 One factor contributing to the more extensive loss of sex-biased gene expression in the whole body
 384 STAT5b-KO mouse model (10) may be the loss of STAT5 throughout prenatal and postnatal

development, as well as its absence in all liver cell types, including non-parenchymal cells. In contrast, in the present, hepatocyte-specific STAT5 knockout mouse model, STAT5 deletion was achieved by activation of a Cre transgene under the control of the *Alb* gene promoter (28), which does not occur in liver non-parenchymal cells, and consequently, there is significant residual expression of STAT5 (42). Another factor contributing to the less extensive loss of sex-biased gene expression in the hepatocyte-specific STAT5-KO mouse model may be CUX2, a female-specific transcriptional repressor that is rapidly induced by continuous GH infusion in male liver (21) and is de-repressed in whole body STAT5b-KO male liver (58), but not in hepatocyte-specific STAT5-KO mouse liver, as shown here. The induction of CUX2 seen in whole body STAT5b-KO male liver is likely due to its direct, STAT5-independent induction by the persistent elevation of circulating GH found in that mouse model (42), which would enable CUX2 to contribute to a more complete feminization of male mouse liver through its widespread repression of male-biased genes (43). Other strongly female-biased genes that were not induced in hepatocyte STAT5-deficient male liver include the drug metabolizing genes *Cyp3a16*, *Hsd3b1* and *Ugt2b37*. Given that these genes, as well as *Cux2*, are strongly induced in livers of male mice treated with GH as a continuous infusion (female plasma profile) (21), it seems that their sex-biased expression is regulated by a GH-dependent but STAT5-independent signaling pathway.

CpG methylation at a specific promoter-proximal CpG dinucleotide has been implicated in the regulation of two sex-biased liver genes, *Cyp2d9* and *C4a/Slp* (38, 59, 60). Our studies identified differential methylation proximal to *C4a/Slp* (Table S2F); however, our RRBS analysis did not identify CpGs nearby *Cyp2d9*, which could be due to insufficient CpG density in that region. While many sex-biased, STAT5-responsive gene promoters contain nearby CpGs captured by our RRBS analysis, very few differentially methylated CpG tiles were found at such promoters. Rather, our findings indicate that STAT5-dependent sex-biased genes are frequently regulated by STAT5-regulated demethylation associated with the activation of distal enhancers. Thus, we observed a high correlation between the sex-bias and STAT5 responsiveness of CpG methylation changes at sex-biased DMRs in both male and female liver. Loss of STAT5 preferentially increased CpG hypermethylation, with the sites that are hypermethylated in the

412 absence of STAT5 being more strongly enriched for enhancer chromatin states, presence of open
413 chromatin (DHS), and STAT5 binding as compared to hypomethylated sites. This indicates that de-
414 methylation at regulatory elements is dependent on STAT5, and that loss of STAT5 leads to retention of
415 DNA methylation at genomic loci that would otherwise be active, demethylated enhancers that bind
416 STAT5 and undergo trans-activation. This regulatory mechanism encompasses many distal enhancers,
417 as indicated by the significant enrichment of DMRs in enhancer chromatin states for co-localization with
418 sex-biased and/or STAT5-responsive genes within their TAD boundaries. Whether DNA demethylation
419 plays a direct, instructive role in promoting STAT5 binding, or is a passive consequence of binding of
420 STAT5 or associated transcription factors is unknown.

421

422 The widespread effects of STAT5a/STAT5b-KO on liver sex differences in gene expression and DNA
423 methylation seen here can primarily be attributed to the loss of STAT5b, which is GH-responsive and is
424 the predominant liver STAT5 form. STAT5a is also GH responsive, but is expressed in mouse liver at a
425 much lower level than STAT5b, as confirmed here by RNA-seq; furthermore, its loss has much less
426 impact on hepatic expression of sex-biased genes (57). STAT1, although classically activated by
427 interferon- γ stimulation (45), can also be activated by GH in liver (61). STAT1 is up-regulated in STAT5-
428 KO liver (Fig. 2), perhaps as an adaptive response in the absence of competition by STAT5 for binding
429 to their shared intracellular docking sites on GH receptor (28). STAT1 has a DNA-binding motif distinct
430 from that of STAT5 (61) and consequently has unique target genes, and thus is unable to compensate
431 for the loss of STAT5-regulated sex-biased gene expression (as was seen here) or for other phenotypic
432 consequences of liver STAT5 deficiency, including impaired cell proliferation and fatty liver development
433 (28). Although STAT5-KO dysregulated only ~2% of stringently sex-independent genes in liver (vs. 57%
434 of sex-biased genes), at least some of the sex-independent genes and DMRs responsive to STAT5-KO
435 may reflect adaptive mechanisms in the liver. Indeed, the sex-independent genes induced by STAT5-KO
436 in both male and female liver are enriched for immune-specific genes (Fig. 2), at least some of which
437 may be STAT1 targets.

438

439 In the *Alb*-Cre-induced hepatocyte-specific STAT5-KO model used here, STAT5 deletion occurs during
 440 late gestation, when the Cre transgene is activated by the *Alb* gene promoter (46). Therefore, the DNA
 441 methylation changes we observed may reflect late gestational or postnatal effects of STAT5 deficiency.
 442 Mammalian DNA methylation patterns are established at many sites during embryogenesis and then
 443 maintained throughout life (62, 63); however, DNA methylation continues to evolve postnatally at a
 444 subset of genomic regions in terminally differentiated cells (39, 40, 64). In mouse hepatocytes,
 445 demethylation at enhancer elements occurs after birth, when it is required to establish postnatal
 446 chromatin accessibility and gene transcription patterns (40). Our finding of enhancer hypermethylation in
 447 STAT5-KO liver indicates that STAT5 regulates demethylation of these genomic regions, which may
 448 contribute to chromatin opening and activation of enhancers critical for postnatal liver function. Our
 449 finding that these regions are also enriched for DNA motifs for other transcription factors important for
 450 sex-biased liver function (Fig. 8) suggests STAT5 may work with these other factors to recruit the TET
 451 demethylases (65) that actively demethylate these enhancer regions. Such demethylation events likely
 452 occur around puberty (39), when pituitary GH secretion becomes pulsatile in males, leading to
 453 intermittent activation of liver STAT5 (16, 17, 66), one of the key driving forces for the striking sex-
 454 differences in liver gene expression that become widespread at that time (67). Further study is needed
 455 to determine whether CpG demethylation of these enhancers by STAT5 is a permanent, irreversible
 456 event (39), or whether the methylation status of these enhancers is dynamically responsive to the
 457 repeated activation of STAT5 in male liver by plasma GH pulses, as was recently found for other sex-
 458 biased epigenetic events (16).

459
 460 In contrast to STAT5-dependent enhancer DNA demethylation, we found that alteration of promoter
 461 DNA methylation is not a common mechanism for STAT5 regulation of sex-biased gene expression.
 462 Although CpG islands are frequently found at or nearby gene promoters, where they are mostly
 463 hypomethylated, even when the gene is transcriptionally silent (68), few promoters become activated in
 464 cells deficient in DNA methyltransferase activity, despite a 95% reduction in global DNA methylation and
 465 the creation of thousands of active enhancers (69-71). This suggests that DNA methylation does not

drive promoter activation or repression, but rather, may be a more common regulatory mechanism in enhancers. The significance of DNA methylation at enhancer sites is apparent in cancers (69, 72, 73), where enhancer CpG methylation is more closely associated with changes in gene expression than is promoter methylation. Our finding that STAT5 primarily alters DNA methylation at enhancer elements is consistent with these findings, and is reminiscent of our earlier observation that sex-biased histone marks are more pronounced at distal sex-biased DHS than at sex-biased gene promoters (29).

STAT5 can respond to many extracellular cytokines and growth factor signals, including GH, and is the direct downstream effector of the hypothalamo-pituitary-liver GH axis. Liver STAT5 signaling is susceptible to disruption by xenobiotics, which can dramatically alter hepatic transcriptomic profiles and impact the ability of liver to metabolize foreign chemicals, increasing susceptibility to liver diseases, including metabolic disorders, fatty liver, hepatocellular carcinoma and obesity (4, 9). Our studies elucidating the interplay between STAT5-dependent DNA methylation and sex-biased gene expression have strong implications for understanding mechanisms underlying diseases associated with dysregulation of liver STAT5 function, which is a common occurrence (74, 75). Our findings suggest that STAT5 plays a critical role in shaping both sex-independent and sex-dependent enhancer elements via DNA methylation, which ultimately contributes to the control of proper RNA output for normal liver function. Given that DNA methylation is considered one of the more stable chromatin modifications, that once established, contribute to durable epigenetic memory through multiple cell divisions (76), dysregulation of STAT5 activity could lead to persistent effects on DNA methylation profiles at regulatory elements in the liver.

487

488

489 **Materials and Methods**

Liver tissue – Livers used for the analysis reported here were the same tissue samples used in our earlier work (42) and were originally obtained from Dr. Lothar Hennighausen (NIDDK, NIH). The livers were from 8 to 12-wk old hepatocyte STAT5-deficient males and females (KOM and KOF, respectively), and floxed controls (wild-type with respect to STAT5, i.e., WTM and WTF, respectively). The

494 hepatocyte-specific STAT5-deficient mouse line (STAT5-KO) was developed by mating albumin
 495 promoter-driven Cre transgenic mice (FVB/N) with C57BL/6 x 129J mice having a floxed *Stat5a-Stat5b*
 496 *locus* (28). All mouse work was performed according to the Animal Research Advisory Committee
 497 Guidelines of NIH (<https://oacu.oir.nih.gov/animal-research-advisory-committee-guidelines>) and was
 498 approved by the NIDDK Animal Care and Use Committee.

499

500 RNA-seq sample preparation and data analysis - A portion of each frozen liver was used for RNA
 501 extraction with TRIzol reagent (Invitrogen Life Technologies Inc., Carlsbad, CA) following the
 502 manufacturer's instructions. Total liver RNA was isolated from n=5-6 individual livers per sex and
 503 genotype (WTM, WTF, KOM, KOF). Two RNA-seq libraries (biological replicates) were prepared for
 504 each group, with each sequencing library comprised of a pool of RNAs obtained from n=2-3 mouse
 505 livers. Libraries were prepared using the Illumina TruSeq RNA library preparation kit (Illumina, cat# RS-
 506 122-2001) starting with total liver RNA depleted of ribosomal RNA using NEBNext rRNA Depletion Kit.
 507 Paired-end sequence reads (50 nt reads) were generated on an Illumina HiSeq instrument at the New
 508 York Genome Center (New York, NY). RNA-seq data was analyzed using a custom pipeline developed
 509 in our lab (16). Briefly, FASTQ files with raw sequence reads were aligned to mouse genome build mm9
 510 (NCBI 37) using Tophat (version 2.0.13) (77) using default parameters. FeatureCounts (78) was used to
 511 count sequence reads mapping to gene bodies within gene bodies of RefSeq genes, and EdgeR (79)
 512 was used to identify differentially expressed genes. 9,175 genes (**Table S1**) were identified as liver-
 513 expressed genes based on the criteria of >1 FPKM average expression in either male or female wild
 514 type mouse liver. An FDR < 0.05 was used to identify differentially expressed genes between male and
 515 female liver (sex-specific genes; 132 male-biased and 142 female-biased genes, based on WTM/WTF
 516 comparison) and for genes responsive to STAT5 loss (KOM/WTM, n=365 genes; and KOF/WTF, n=278
 517 genes). Genes with FDR > 0.05 were deemed non-responsive to STAT5 loss (Table S1A). Stringent
 518 sex-independent genes (n=6,306) were identified by |fold-difference| < 1.2 and FDR > 0.1 for the
 519 WTM/WTF comparison.

520

521 STAT5 ChIP-seq peaks nearby genes - Genomic coordinates of male-biased, female-biased and sex-
522 independent STAT5 ChIP-seq peaks were downloaded from the Supplementary Materials of (12).
523 STAT5 ChIP-seq peaks within 10 kb upstream or downstream to genes of interest were identified using
524 the Bedtools (80) command: `bedtools intersect -c -w 10000` and are shown in Table S1A.
525
526 Calculation of correlation - Gene expression fold change values and differentially methylated values
527 contain both positive and negative values. When calculating correlations between two conditions, fold
528 change and differentially methylated values were first converted to absolute values to avoid artificial
529 overestimation of correlation. Linear regression was performed in R, and R^2 and p-values are reported.
530
531 Reduced representation bisulfite sequencing (RRBS) - RRBS libraries were prepared from frozen liver
532 tissues as previously described, with modifications (41). Briefly, ~50 mg of frozen liver tissue was
533 homogenized by vigorously pipetting up and down the sample in 370 μ l TE buffer (10 mM Tris, 1mM
534 EDTA, pH 8) containing 10 μ l of proteinase K solution (10 mg/ml), and 20 μ l of 10% (v/v) SDS was
535 quickly added to achieve 0.5% SDS final concentration. The homogenate was incubated at 50°C for 2 hr.
536 Genomic DNA (gDNA) was purified by extraction with 400 μ l phenol chloroform with the addition of 32 μ l
537 of 5 M NaCl, using a large-bore pipet tip with gentle inversions and without vortexing. Intact gDNA was
538 spooled from the aqueous phase following the addition of an equal volume pf 100% ethanol. The
539 spooled gDNA was treated with RNase in PBS buffer (2 μ l of 10 mg/ml RNase A in 298 μ l PBS) for 1 hr
540 at room temperature. The RNA-free gDNA was washed in ethanol and thoroughly resuspended in TE
541 buffer for Qubit quantification of DNA, taking care to first sample a large volume (10 μ l) and to then
542 shear the DNA by vigorous pipeting to ensure that the subsequent 2 μ l input for Qubit analysis is as
543 homogenous as possible. 500 ng of intact gDNA was digested overnight with 2.5 μ l of MspI (20,000
544 units/ml; NEB cat. R0106S) in the manufacturer's supplied buffer (40 μ l total volume) at 37°C, with the
545 lid heated to 40°C to insure more complete digestion. The reaction was subjected to end-repair in the
546 same reaction by adding 1 μ l of Klenow fragment (New England Biolabs cat. M0210S, 5000 U/ml) and 1
547 μ l of dNTP mix (10 mM dA, 1mM dG, 1 mM dC). Samples were then incubated at 30°C for 20 min,

548 followed by 37°C for 20 min, and then purified by phenol chloroform extraction. 450 ng of MpsI-digested,
 549 end-repaired material (quantified by Nanodrop measurement) was ligated to Illumina adaptors supplied
 550 in the NEB DNA-seq kit (NEB cat. E7370S, using 1/20 diluted adaptor) and NEB Quick DNA ligase kit
 551 (NEB cat. E6056S) by incubation at 16°C overnight without a heated lid. After USER enzyme digestion
 552 and cleanup with Zymo's DNA Clean and Concentrator-5 (Zymo cat. D4003T), the adaptor-ligated MspI
 553 fragments were bisulfite-treated for two consecutive rounds using Zymo's EZ DNA Methylation-Gold Kit
 554 (Zymo cat. D5005) per the manufacturer's instructions. A portion (5 µl) of each 12 µl bisulfite-treated
 555 DNA sample was amplified using the KAPA HiFi HS Uracil+ polymerase system (Kapa
 556 Biosystems/Roche cat. KM2800) for ~12-15 PCR cycles, with the optimal cycle number determined
 557 empirically to find the lowest number of cycles that yielded sufficient material without over-amplification.
 558 Finally, size selection of fragments 160-340 bp was accomplished by performing a double SPRI
 559 purification of 0.65x and 0.55x, as follows. SPRI beads (32.5 µl) were added to each 50 µl reaction
 560 (0.65x). After mixing well, the tubes were placed on magnetic rack until the beads were separated from
 561 the supernatant. The supernatant was transferred to fresh tubes, to which fresh SPRI beads (27.5 µl;
 562 0.55x) were then added to the supernatant. The beads were washed twice in 200 µl of 80% ethanol
 563 solution, air dried for 5 min and eluted in 0.1X TE buffer. The final RRBS sequencing libraries were
 564 analyzed on a 1% agarose gel stained with SYBR green, quantified by Qubit analysis and sequenced on
 565 an Illumina HiSeq instrument (paired-end, 50 bp reads) with no more than ~30% of a sequencing lane
 566 being RRBS material. RRBS libraries were prepared and sequenced for the following samples with the
 567 indicated numbers of biological replicate livers: WT male (n=6), WT female (n=5), STAT5-KO male (n=6),
 568 and STAT5-KO female (n=6).

569

570 RRBS data processing - About 15 million paired-end sequence reads were obtained for each RRBS
 571 sample. Raw FASTQ files were first processed by Trim Galore (version 0.4.3) using options: --rrbs --
 572 paired. The trimmed read1 and read2 files were mapped by Bismark (81) using default parameters.
 573 PileOMeth was used to extract and quantify all CpG dinucleotides in the bam files, outputting the
 574 location, total coverage, number of thymines and number of cytosines of each CpG in MethylKit-

compatible format (parameter: --methylKit), while avoiding any double-counting of reads where paired-end reads for a given fragment overlap. The PileOMeth results were directly input into MethylKit, two sample groups at a time, for pair-wise differential analysis. MethylKit (82) differential analysis was performed at the level of both 100 bp tiles and individual CpGs using default parameters combined with the following: context="CpG", filterByCoverage (count=10), destrand=T. A total of 4 comparisons were performed in a Treatment vs. Control configuration, yielding differential methylation values (diff.meth), where the absolute value indicates the magnitude of change, and positive or negative signs indicate hypermethylation or hypomethylation of the treatment group relative to the control group, respectively. Tiles and individual CpGs with a significant change in DNA methylation were defined using a threshold fold-difference of 15% and q-value < 0.05, and were combined such that individual CpGs not captured at the tile level were included. Significant tiles and significant individual CpGs exhibit significant overlap with each other, and are summarized in **Table S2** (also see below). Here, differentially methylated regions (DMRs) refers to 100 bp genomic regions that fit one of the following criteria: 1) 100 bp tiles that were significantly differentially methylated at the level of overall tile statistics and also contain one or more CpGs that showed significant differences in methylation at the individual CpG level (this describes the majority of DMRs); 2) 100 bp tiles that were significant in their overall tile statistics but did not show statistically significant differences in individual CpGs; and 3) 100 bp regions that were not significantly differentially methylated at the level of overall tile statistics, but that contain at least one significant individual CpG, whose diff.meth values were averaged based on those CpGs that showed significant differences in methylation. The criteria by which the DMRs were identified is flagged in **Table S2**. In the majority of analyses, all DMRs were utilized.

A total 1,355,894 CpGs were captured by all 23 liver RRBS samples (Fig. S1, Supplemental file: all_cpG_covered_by_STAT5KO_23samples.txt). To generate this set of individual CpGs, raw PileOmeth files were de-stranded using a custom Python script, script27_destranding_pileometh.py; see Supplementary materials), which collapses CpG coordinates from opposite strands into a single coordinate, to facilitate downstream analyses, as follows. Each CpG is symmetrical in the genome,

which means that each CpG will have two coordinates, e.g. chr1:100 (+) and chr1:101 (-). The de-stranding script aggregates the genomic position from the (+) strand and from the (-) strand into a single genomic position. Next, the frequency of occurrence of each CpG in each of the 23 RRBS-seq samples was determined, and only those CpGs that were covered by all 23 samples were considered. By segmenting the mouse genome into 100 bp non-overlapping windows, these 1,355,894 CpGs were aggregated into 232,882 windows (Supplemental file: complete_23STAT5_unique_tiles_unique.txt). The number of 100 bp windows (tiles) is lower than the number of individual CpGs because many windows contain multiple CpGs.

610

Rationale for merging of tile and individual CpG statistics – As described above, DMRs were determined by MethyKit pair-wise comparison using a tile-based method and also an individual CpG method. The tile-based analysis gives higher statistical power due to the aggregation of signals from multiple CpGs within a 100 bp genomic region. For example, if a tile contains 3 CpGs, all of which are only slightly differential, the tile statistics may give a modest differential methylation value but with a low FDR, whereas individual CpG analysis would yield a high FDR for each CpG. In a different scenario, if a single CpG within a tile region is highly differentially methylated but is situated nearby other CpGs that are not differentially methylated, the signal from the highly differential CpG will be ‘diluted’ by the averaging used in the tile analysis, causing the loss of a significantly differential CpG. Because most tiles contain only one CpG, results of the tile and individual CpG analyses are similar for the most part. The final list of differential CpGs includes 2,375 significant tiles, as well as 1,514 100-bp genomic regions with individual CpGs that were captured by the individual CpG analysis but not by the tile analysis, for a combined total of 3,889 differential 100 bp genomic regions, is presented and further explained in Table S2.

625

An example of the R code tile and individual CpG differential analysis as follows:

```
#my.list refers to file locations for 11 PileOmeth output files
myobj_1 <-
```

```

629 methRead(my.list, sample.id=list('M1_WT', 'M2_WT', 'M3_WT', 'M4_WT', 'M5_WT', 'M6_W
630 T', 'F1_WT', 'F2_WT', 'F3_WT', 'F4_WT', 'F5_WT'), assembly='mm9', context='CpG', trea
631 tment=c(1,1,1,1,1,1,0,0,0,0,0))
632
633 filtered.myobj_1<-filterByCoverage(myobj_1, lo.count=10, lo.perc=NULL,
634 hi.count=NULL, hi.perc=99.9)
635
636 meth_1<-unite(myobj_1, destrand=T)
637
638 tiles_1<- tileMethylCounts(meth_1, win.size=100, step.size=100)
639
640 #calculate differential 100 bp tiles
641 myDiff_tiles_1<-calculateDiffMeth(tiles_1)
642
643 #calculate differential individual CpG
644 myDiff_cpg<-calculateDiffMeth(meth_1)
645
646 Generation of raw methylation heat maps – An intermediate file containing raw methylation values for all
647 100 bp tiles was generated by inputting Raw PileOMeth files into MethylKit and using the following
648 command: tileMethylCounts(pileometh_input, win.size=100, step.size=100). Raw
649 methylation values were retrieved for only the differential tiles, which were visualized as a heat map
650 using R. These raw methylation values in WTF, KOM and KOF samples were tested against WTM and
651 evaluated using two-tailed Student t-test. For the heat maps shown in Fig. 3C, only DMRs determined
652 by tile statistics with stringent threshold (diff.meth > 25%, qval < 0.05) were included. The heat map in
653 Fig. 3C presents raw methylation values of DMRs that were determined to be significantly differential by
654 tile analysis between STAT5-KO male liver and wild-type male liver when determined by tile analysis,
655 i.e., excluding DMRs identified by individual CpGs only.

```

656

657 DMRs in gene promoters - Gene promoters were defined as 5 kb upstream of the gene transcription
 658 start site. Thus, for genes on the + strand, the promoter coordinates were set as (start position – 5000,
 659 start position); for genes on the - strand, coordinates were set as (end position, end position + 5000).
 660 DMRs positioned in a promoter region (1 nt overlap) were identified using the Bedtools windows
 661 command (Table S3).

662

663 Distribution of CpGs across chromatin states and annotated genomic features – To determine the
 664 distribution of chromatin states across All Tiles (see Fig. 5B), the 232,882 CpG-containing 100 bp tiles
 665 were mapped to 14 chromatin states previously defined in male mouse liver (29) using ChromHMM (83)
 666 and the Intersect command of Bedtools (80). Custom Bash commands and R scripts were used to
 667 calculate the frequencies of the 14 chromatin states in each bin and the results were plotted by ggplot.
 668 Chromatin state distributions were similarly determined for the sets of DMRs identified in the WTM/WTF,
 669 KOM/WTM and KOF/WTF comparisons, using the full set of DMRs (i.e., DMRs identified by tiles and
 670 DMRs identified by individual CpGs) described in Table S2 for each comparison. The relative frequency
 671 of occurrence of each chromatin state in each 100 bp bin across a 4 kb window surrounding each
 672 hypomethylated and hypermethylated DMR. To calculate enrichment scores, the 14 chromatin states
 673 were grouped into five super states, designated as inactive (states 1, 2, 3), enhancer (states 5, 6, 8, 9,
 674 10), bivalent (state 12), transcribed (states 13, 14) and promoter (states 7, 8). Enrichment of DMRs for
 675 these five super states relative to ‘all CpGs’ (i.e., the set of 232,882 tiles covered by our RRBS samples)
 676 was calculated as follows: (number of differential DMRs in chromatin state-X / total number of differential
 677 DMRs) / (number of all DMRs in chromatin state-X / 232,882). Statistical significance was determined in
 678 R by two proportion Z-test using the *prop.test* function.

679

680 TAD analysis - DMRs were grouped by their chromatin states defined in wild type mouse liver (29).
 681 DMRs and genes were then mapped to topologically associating domains (TADs) defined for mouse
 682 liver (51), and the overlap of DMR-containing TADs and gene-containing TADs was analyzed using

683 datasets provided in Table S4. As a control, for each chromatin state, comparably sized sets of RRBS
684 tiles that did not show significant sex differences or response to STAT5 loss were randomly sampled 50
685 times and mapped to TADs. The significance of DMRs occurring in TADs containing sex-biased,
686 STAT5-responsive genes compared to the extent of overlap observed for the random controls was
687 assessed by t-test. In addition, the distance from the start of the DMR to the TSS of sex-biased, STAT5
688 responsive genes was determined and analyzed using box plots, where DMRs that are $> 10^6$ bp away
689 from genes are deemed outliers and were excluded from analysis.

690
691 Gene set enrichment analysis (GSEA) - DMRs determined by tile statistics with stringent threshold (qval
692 < 0.05) were used. DMRs were ranked from high (hypermethylated) to low (hypomethylated). This list of
693 ranked DMRs as well as the list of DMRs with nearby (within 10 kb) male-biased and female-biased
694 DHS (52) (Table S5), were input into the GSEA desktop application for pre-ranked test analysis (84),
695 with the number of permutations set to 1000.

696
697 MEME DNA motif analysis - The Analysis of Motif Enrichment (AME) from the MEME suite (85) was
698 used to assess the enrichment of the STAT5 motif in the tile sequences. The following command was
699 used: `ame --control random_sequences.txt --evaluate-report-threshold 1 --o comp1_N`
700 `comp1_N_sequence.txt STAT5_meme_motif.txt`, where `random_sequences.txt` is a FASTA files of 100
701 bp sequences from randomly selected genomic regions that are not covered by our RRBS
702 experiments; `comp1_N_sequences.txt` is the fasta file of the DMR tile sequences; and
703 `STAT5_meme_motif.txt` is the MEME file of the STAT5 motif (see Supplemental materials).

704
705 Cumulative frequency plots of STAT5 ChIP-seq peaks nearby DMRs - DMRs based on both tiles and
706 individual CpG statistics for three comparisons (WTM/WTF; KOM/WTM; KOF/WTF) were separated into
707 hypomethylated and hypermethylated DMRs. In addition, a file containing 1000 randomly selected non-
708 differential methylated genomic regions was used. Sex-independent, male-biased and female-biased
709 STAT5 ChIP-seq peaks were from (12). The analysis was based on 14,910 ChIP-seq peaks for STAT5

(see stat5b_merged_ss_sorted.txt, in Supplemental materials) from (12) after excluding 184 peaks on ChrX and ChrM from the full set of 15,094 ChIP-seq peaks reported in that study. Male-biased STAT5 binding occurs at 1,661 of the 14,910 sites, and female-biased binding occurs at 1,782 sites. The Bedtools command `bedtools closest -d -t` was used to find the distance to the nearest STAT5 peak for each DMR. For each of the seven sets of DMRs, DMRs were first ranked by distance to the nearest STAT5 peak from low to high, then cumulative values (which range from 0 to 1) were assigned to the ranked DMRs as follows: (position in ranked list / total number in list). The cutoff for the distance to nearest STAT5 peak was set at 10 kb. The distance (x axis) and cumulative frequency values (y axis) were plotted for all sets of DMRs.

Homer DNA motif analysis – The Homer motif analysis tool (53) was used to find enriched motifs at DMRs relative to a background set comprised of a random selection of one thousand 100 bp regions that were detected by RRBS but not differentially methylated. Homer enrichment analysis of known motifs (Table S6) is based on 332 motifs in the HOMER motif database that were mostly curated from published ChIP-seq datasets (<http://homer.ucsd.edu/homer/motif/motifDatabase.html>). The command for Homer motif analysis was as follows: `findMotifsGenome.pl input.txt mm9 output_folder -size 100 -bg random_1000.txt` where input.txt contains the genomic coordinates of DMRs to be analyzed for enriched regulatory elements; mm9 specifies the genome; "-size 100" indicates that the genomic regions in input.txt are 100 bp long; and "-bg random_1000.txt" specifies the coordinates for the background regions.

Data availability - Raw and processed sequencing data are available at GEO (<https://www.ncbi.nlm.nih.gov/geo/>) under accession numbers GSE103885 (RNA-seq samples) and GSE103886 (RRBS samples).

Acknowledgments

736 This work was supported in part by NIH grant DK121998 (to DJW). The funding agency had no role in
737 study design, data collection and interpretation, or the decision to submit the work for publication. The
738 authors thank Dr. Lothar Hennighausen, NIDDK, for originally providing the mouse liver tissue samples
739 used in these analyses. The authors state they have no financial conflicts of interest to disclose.

740

741 **Author contributions:** PH and DJW jointly conceptualized the project and designed the study. PH
742 carried out all of the laboratory experiments, developed and implemented the methods for computational
743 analysis, carried out data analysis and figure preparation. DJW obtained grant funding and supervised
744 the study. PH and DJW jointly drafted the manuscript and DJW revised and edited the manuscript for
745 publication.

746

747 References

748

- 749 1. **Levy DE, Darnell JE, Jr.** 2002. Stats: transcriptional control and biological impact. *Nat Rev Mol Cell*
750 *Biol* **3**:651-662.
- 751 2. **Furth PA, Nakles RE, Millman S, Diaz-Cruz ES, Cabrera MC.** 2011. Signal transducer and activator of
752 transcription 5 as a key signaling pathway in normal mammary gland developmental biology and
753 breast cancer. *Breast Cancer Res* **13**:220.
- 754 3. **Hughes K, Watson CJ.** 2012. The spectrum of STAT functions in mammary gland development. *Jakstat*
755 **1**:151-158.
- 756 4. **Kaltenecker D, Themanns M, Mueller KM, Spirk K, Suske T, Merkel O, Kenner L, Luis A, Kozlov A,**
757 **Haybaeck J, Muller M, Han X, Moriggl R.** 2019. Hepatic growth hormone - JAK2 - STAT5 signalling:
758 Metabolic function, non-alcoholic fatty liver disease and hepatocellular carcinoma progression.
759 *Cytokine* **124**:154569.
- 760 5. **Waters MJ.** 2016. The growth hormone receptor. *Growth Horm IGF Res* **28**:6-10.
- 761 6. **Dodington DW, Desai HR, Woo M.** 2018. JAK/STAT - Emerging Players in Metabolism. *Trends*
762 *Endocrinol Metab* **29**:55-65.
- 763 7. **Chia DJ.** 2014. Minireview: mechanisms of growth hormone-mediated gene regulation. *Mol*
764 *Endocrinol* **28**:1012-1025.
- 765 8. **Waxman DJ, Holloway MG.** 2009. Sex differences in the expression of hepatic drug metabolizing
766 enzymes. *Mol Pharmacol* **76**:215-228.
- 767 9. **Lichanska AM, Waters MJ.** 2008. How growth hormone controls growth, obesity and sexual
768 dimorphism. *Trends Genet* **24**:41-47.
- 769 10. **Clodfelter KH, Holloway MG, Hodor P, Park SH, Ray WJ, Waxman DJ.** 2006. Sex-dependent liver
770 gene expression is extensive and largely dependent upon signal transducer and activator of
771 transcription 5b (STAT5b): STAT5b-dependent activation of male genes and repression of female
772 genes revealed by microarray analysis. *Mol Endocrinol* **20**:1333-1351.
- 773 11. **Herrington J, Smit LS, Schwartz J, Carter-Su C.** 2000. The role of STAT proteins in growth hormone
774 signaling. *Oncogene* **19**:2585-2597.

12. **Zhang Y, Laz EV, Waxman DJ.** 2012. Dynamic, sex-differential STAT5 and BCL6 binding to sex-biased, growth hormone-regulated genes in adult mouse liver. *Mol Cell Biol* **32**:880-896.
13. **Baik M, Yu JH, Hennighausen L.** 2011. Growth hormone-STAT5 regulation of growth, hepatocellular carcinoma, and liver metabolism. *Ann N Y Acad Sci* **1229**:29-37.
14. **Veldhuis JD, Bowers CY.** 2003. Human GH pulsatility: an ensemble property regulated by age and gender. *J Endocrinol Invest* **26**:799-813.
15. **Xu J, Bekaert AJ, Dupont J, Rouve S, Annesi-Maesano I, De Magalhaes Filho CD, Kappeler L, Holzenberger M.** 2011. Exploring endocrine GH pattern in mice using rank plot analysis and random blood samples. *J Endocrinol* **208**:119-129.
16. **Connerney J, Lau-Corona D, Rampersaud A, Waxman DJ.** 2017. Activation of Male Liver Chromatin Accessibility and STAT5-Dependent Gene Transcription by Plasma Growth Hormone Pulses. *Endocrinology* **158**:1386-1405.
17. **Tannenbaum GS, Choi HK, Gurd W, Waxman DJ.** 2001. Temporal relationship between the sexually dimorphic spontaneous GH secretory profiles and hepatic STAT5 activity. *Endocrinology* **142**:4599-4606.
18. **Seton-Rogers S.** 2014. Hepatocellular carcinoma: gender differences. *Nat Rev Cancer* **14**:578.
19. **Park SH, Liu X, Hennighausen L, Davey HW, Waxman DJ.** 1999. Distinctive roles of STAT5a and STAT5b in sexual dimorphism of hepatic P450 gene expression. Impact of STAT5a gene disruption. *J Biol Chem* **274**:7421-7430.
20. **Wauthier V, Sugathan A, Meyer RD, Dombkowski AA, Waxman DJ.** 2010. Intrinsic sex differences in the early growth hormone responsiveness of sex-specific genes in mouse liver. *Mol Endocrinol* **24**:667-678.
21. **Lau-Corona D, Suvorov A, Waxman DJ.** 2017. Feminization of male mouse liver by persistent growth hormone stimulation: Activation of sex-biased transcriptional networks and dynamic changes in chromatin states. *Mol Cell Biol* doi:10.1128/MCB.00301-17.
22. **Holloway MG, Laz EV, Waxman DJ.** 2006. Codependence of growth hormone-responsive, sexually dimorphic hepatic gene expression on signal transducer and activator of transcription 5b and hepatic nuclear factor 4alpha. *Mol Endocrinol* **20**:647-660.
23. **Hao P, Waxman DJ.** 2018. Functional Roles of Sex-Biased, Growth Hormone-Regulated MicroRNAs miR-1948 and miR-802 in Young Adult Mouse Liver. *Endocrinology* **159**:1377-1392.
24. **Melia T, Hao P, Yilmaz F, Waxman DJ.** 2015. Hepatic Long Intergenic Noncoding RNAs: High Promoter Conservation and Dynamic, Sex-Dependent Transcriptional Regulation by Growth Hormone. *Molecular and cellular biology* **36**:50-69.
25. **Melia T, Waxman DJ.** 2019. Sex-Biased lncRNAs Inversely Correlate With Sex-Opposite Gene Coexpression Networks in Diversity Outbred Mouse Liver. *Endocrinology* **160**:989-1007.
26. **Davey HW, Park SH, Grattan DR, McLachlan MJ, Waxman DJ.** 1999. STAT5b-deficient mice are growth hormone pulse-resistant. Role of STAT5b in sex-specific liver p450 expression. *J Biol Chem* **274**:35331-35336.
27. **Bennett E, McGuinness L, Gevers EF, Thomas GB, Robinson IC, Davey HW, Luckman SM.** 2005. Hypothalamic STAT proteins: regulation of somatostatin neurones by growth hormone via STAT5b. *J Neuroendocrinol* **17**:186-194.
28. **Cui Y, Hosui A, Sun R, Shen K, Gavrilo O, Chen W, Cam MC, Gao B, Robinson GW, Hennighausen L.** 2007. Loss of signal transducer and activator of transcription 5 leads to hepatosteatosis and impaired liver regeneration. *Hepatology* **46**:504-513.
29. **Sugathan A, Waxman DJ.** 2013. Genome-wide analysis of chromatin states reveals distinct mechanisms of sex-dependent gene regulation in male and female mouse liver. *Mol Cell Biol* **33**:3594-3610.
30. **Greenberg MVC, Bourc'his D.** 2019. The diverse roles of DNA methylation in mammalian development and disease. *Nat Rev Mol Cell Biol* **20**:590-607.
31. **da Rocha ST, Gendrel AV.** 2019. The influence of DNA methylation on monoallelic expression. *Essays Biochem* **63**:663-676.

32. **Kouzmenko A, Ohtake F, Fujiki R, Kato S.** 2010. Hormonal gene regulation through DNA methylation and demethylation. *Epigenomics* **2**:765-774.
33. **Singh S, Li SS.** 2012. Epigenetic effects of environmental chemicals bisphenol A and phthalates. *Int J Mol Sci* **13**:10143-10153.
34. **Angeloni A, Bogdanovic O.** 2019. Enhancer DNA methylation: implications for gene regulation. *Essays Biochem* **63**:707-715.
35. **Ito S, Hirabayashi K, Moriishi K, Matsui Y, Moriya K, Koike K, Matsuura Y, Shiota K, Yagi S.** 2015. Novel sex-dependent differentially methylated regions are demethylated in adult male mouse livers. *Biochem Biophys Res Commun* **462**:332-338.
36. **McCormick H, Young PE, Hur SSJ, Booher K, Chung H, Cropley JE, Giannoulatou E, Suter CM.** 2017. Isogenic mice exhibit sexually-dimorphic DNA methylation patterns across multiple tissues. *BMC Genomics* **18**:966.
37. **Takasugi M, Hayakawa K, Arai D, Shiota K.** 2013. Age- and sex-dependent DNA hypomethylation controlled by growth hormone in mouse liver. *Mech Ageing Dev* **134**:331-337.
38. **Krebs CJ, Schultz DC, Robins DM.** 2012. The KRAB zinc finger protein RSL1 regulates sex- and tissue-specific promoter methylation and dynamic hormone-responsive chromatin configuration. *Mol Cell Biol* **32**:3732-3742.
39. **Reizel Y, Spiro A, Sabag O, Skversky Y, Hecht M, Keshet I, Berman BP, Cedar H.** 2015. Gender-specific postnatal demethylation and establishment of epigenetic memory. *Genes Dev* **29**:923-933.
40. **Reizel Y, Sabag O, Skversky Y, Spiro A, Steinberg B, Bernstein D, Wang A, Kieckhaefer J, Li C, Pikarsky E, Levin-Klein R, Goren A, Rajewsky K, Kaestner KH, Cedar H.** 2018. Postnatal DNA demethylation and its role in tissue maturation. *Nat Commun* **9**:2040.
41. **Gu H, Smith ZD, Bock C, Boyle P, Gnirke A, Meissner A.** 2011. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat Protoc* **6**:468-481.
42. **Holloway MG, Cui Y, Laz EV, Hosui A, Hennighausen L, Waxman DJ.** 2007. Loss of sexually dimorphic liver gene expression upon hepatocyte-specific deletion of Stat5a-Stat5b locus. *Endocrinology* **148**:1977-1986.
43. **Conforto TL, Zhang Y, Sherman J, Waxman DJ.** 2012. Impact of CUX2 on the female mouse liver transcriptome: activation of female-biased genes and repression of male-biased genes. *Mol Cell Biol* **32**:4611-4627.
44. **Conforto TL, Steinhardt GF, Waxman DJ.** 2015. Cross Talk Between GH-Regulated Transcription Factors HNF6 and CUX2 in Adult Mouse Liver. *Mol Endocrinol* **29**:1286-1302.
45. **Au-Yeung N, Mandhana R, Horvath CM.** 2013. Transcriptional regulation by STAT1 and STAT2 in the interferon JAK-STAT pathway. *JAKSTAT* **2**:e23931.
46. **Weisend CM, Kundert JA, Suvorova ES, Prigge JR, Schmidt EE.** 2009. Cre activity in fetal albCre mouse hepatocytes: Utility for developmental studies. *Genesis* **47**:789-792.
47. **Jeziorska DM, Murray RJS, De Gobbi M, Gaentzsch R, Garrick D, Ayyub H, Chen T, Li E, Telenius J, Lynch M, Graham B, Smith AJH, Lund JN, Hughes JR, Higgs DR, Tufarelli C.** 2017. DNA methylation of intragenic CpG islands depends on their transcriptional activity during differentiation and disease. *Proc Natl Acad Sci U S A* **114**:E7526-e7535.
48. **Cedar H, Bergman Y.** 2012. Programming of DNA methylation patterns. *Annu Rev Biochem* **81**:97-117.
49. **Yu M, Ren B.** 2017. The Three-Dimensional Organization of Mammalian Genomes. *Annu Rev Cell Dev Biol* **33**:265-289.
50. **Vietri Rudan M, Barrington C, Henderson S, Ernst C, Odom DT, Tanay A, Hadjur S.** 2015. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep* **10**:1297-1309.
51. **Matthews BJ, Waxman DJ.** 2018. Computational prediction of CTCF/cohesin-based intra-TAD loops that insulate chromatin contacts and gene expression in mouse liver. *Elife* **7**.
52. **Ling G, Sugathan A, Mazor T, Fraenkel E, Waxman DJ.** 2010. Unbiased, genome-wide in vivo mapping of transcriptional regulatory elements reveals sex differences in chromatin structure associated with sex-specific liver gene expression. *Mol Cell Biol* **30**:5531-5544.

- 877 53. **Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK.** 2010.
878 Simple combinations of lineage-determining transcription factors prime cis-regulatory elements
879 required for macrophage and B cell identities. *Mol Cell* **38**:576-589.
- 880 54. **Wade PA.** 2001. Methyl CpG-binding proteins and transcriptional repression. *Bioessays* **23**:1131-1137.
- 881 55. **Cedar H, Bergman Y.** 2009. Linking DNA methylation and histone modification: patterns and
882 paradigms. *Nat Rev Genet* **10**:295-304.
- 883 56. **Newell-Price J, Clark AJ, King P.** 2000. DNA methylation and silencing of gene expression. *Trends*
884 *Endocrinol Metab* **11**:142-148.
- 885 57. **Clodfelter KH, Miles GD, Wauthier V, Holloway MG, Zhang X, Hodor P, Ray WJ, Waxman DJ.** 2007.
886 Role of STAT5a in regulation of sex-specific gene expression in female but not male mouse liver
887 revealed by microarray analysis. *Physiol Genomics* **31**:63-74.
- 888 58. **Laz EV, Holloway MG, Chen CS, Waxman DJ.** 2007. Characterization of three growth hormone-
889 responsive transcription factors preferentially expressed in adult female liver. *Endocrinology*
890 **148**:3327-3337.
- 891 59. **Yokomori N, Kobayashi R, Moore R, Sueyoshi T, Negishi M.** 1995. A DNA methylation site in the
892 male-specific P450 (Cyp 2d-9) promoter and binding of the heteromeric transcription factor GABP.
893 *Mol Cell Biol* **15**:5355-5362.
- 894 60. **Yokomori N, Moore R, Negishi M.** 1995. Sexually dimorphic DNA demethylation in the promoter of
895 the Slp (sex-limited protein) gene in mouse liver. *Proc Natl Acad Sci U S A* **92**:1302-1306.
- 896 61. **Ram PA, Park SH, Choi HK, Waxman DJ.** 1996. Growth hormone activation of Stat 1, Stat 3, and Stat
897 5 in rat liver. Differential kinetics of hormone desensitization and growth hormone stimulation of
898 both tyrosine phosphorylation and serine/threonine phosphorylation. *J Biol Chem* **271**:5929-5940.
- 899 62. **Smith ZD, Meissner A.** 2013. DNA methylation: roles in mammalian development. *Nat Rev Genet*
900 **14**:204-220.
- 901 63. **Weaver JR, Susiarjo M, Bartolomei MS.** 2009. Imprinting and epigenetic changes in the early embryo.
902 *Mamm Genome* **20**:532-543.
- 903 64. **Cannon MV, Pilarowski G, Liu X, Serre D.** 2016. Extensive Epigenetic Changes Accompany Terminal
904 Differentiation of Mouse Hepatocytes After Birth. *G3 (Bethesda)* **6**:3701-3709.
- 905 65. **Zhao H, Chen T.** 2013. Tet family of 5-methylcytosine dioxygenases in mammalian development. *J*
906 *Hum Genet* **58**:421-427.
- 907 66. **Choi HK, Waxman DJ.** 2000. Plasma growth hormone pulse activation of hepatic JAK-STAT5 signaling:
908 developmental regulation and role in male-specific liver gene expression. *Endocrinology* **141**:3245-
909 3255.
- 910 67. **Conforto TL, Waxman DJ.** 2012. Sex-specific mouse liver gene expression: genome-wide analysis of
911 developmental changes from pre-pubertal period to young adulthood. *Biol Sex Differ* **3**:9.
- 912 68. **Deaton AM, Bird A.** 2011. CpG islands and the regulation of transcription. *Genes Dev* **25**:1010-1022.
- 913 69. **Aran D, Sabato S, Hellman A.** 2013. DNA methylation of distal regulatory sites characterizes
914 dysregulation of cancer genes. *Genome Biol* **14**:R21.
- 915 70. **Blattler A, Yao L, Witt H, Guo Y, Nicolet CM, Berman BP, Farnham PJ.** 2014. Global loss of DNA
916 methylation uncovers intronic enhancers in genes showing expression changes. *Genome Biol* **15**:469.
- 917 71. **Charlet J, Duymich CE, Lay FD, Mundbjerg K, Dalsgaard Sorensen K, Liang G, Jones PA.** 2016.
918 Bivalent Regions of Cytosine Methylation and H3K27 Acetylation Suggest an Active Role for DNA
919 Methylation at Enhancers. *Mol Cell* **62**:422-431.
- 920 72. **Jin W, Li QZ, Zuo YC, Cao YN, Zhang LQ, Hou R, Su WX.** 2018. The Relationship Between DNA
921 Methylation in Key Region and the Differential Expressions of Genes in Human Breast Tumor Tissue.
922 *DNA Cell Biol* doi:10.1089/dna.2018.4276.
- 923 73. **Neiman D, Moss J, Hecht M, Magenheimer J, Piyanzin S, Shapiro AMJ, de Koning EJP, Razin A, Cedar H,**
924 **Shemer R, Dor Y.** 2017. Islet cells share promoter hypomethylation independently of expression, but
925 exhibit cell-type-specific methylation in enhancers. *Proc Natl Acad Sci U S A* **114**:13525-13530.

74. **Oshida K, Vasani N, Waxman DJ, Corton JC.** 2016. Disruption of STAT5b-Regulated Sexual Dimorphism of the Liver Transcriptome by Diverse Factors Is a Common Event. *PLoS One* **11**:e0148308.
75. **Cordoba-Chacon J, Sarmiento-Cabral A, Del Rio-Moreno M, Diaz-Ruiz A, Subbaiah PV, Kineman RD.** 2018. Adult-Onset Hepatocyte GH Resistance Promotes NASH in Male Mice, Without Severe Systemic Metabolic Dysfunction. *Endocrinology* **159**:3761-3774.
76. **Kim M, Costello J.** 2017. DNA methylation: an epigenetic mark of cellular memory. *Exp Mol Med* **49**:e322.
77. **Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L.** 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**:562-578.
78. **Liao Y, Smyth GK, Shi W.** 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**:923-930.
79. **Robinson MD, McCarthy DJ, Smyth GK.** 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**:139-140.
80. **Quinlan AR.** 2014. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics* **47**:11 12 11-34.
81. **Krueger F, Andrews SR.** 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**:1571-1572.
82. **Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE.** 2012. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol* **13**:R87.
83. **Ernst J, Kellis M.** 2017. Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc* **12**:2478-2492.
84. **Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP.** 2007. GSEA-P: a desktop application for Gene Set Enrichment Analysis. *Bioinformatics* **23**:3251-3253.
85. **McLeay RC, Bailey TL.** 2010. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics* **11**:165.

Table 1. Liver-expressed genes showing significant expression changes in hepatocyte specific STAT5-KO mouse liver. 414 liver-expressed genes (FPKM >1 in either wild-type male or female liver) were analyzed: 132 male-biased genes, 142 female-biased genes, and 140 stringently sex-independent genes whose expression changes significantly (FDR < 0.05) with STAT5 deficiency in either male or female mouse liver (Table S1A, Table S1B). Data are shown for the 155 of 274 sex-biased genes (57%; 87 of 132 male-biased genes and 68 of 142 female-biased genes) and for the 140 of 6,323 stringently sex-independent genes (2.2%) that are liver-expressed and whose expression changes significantly with STAT5 deficiency in either male or female mouse liver (genes 'responsive' to STAT5-KO). Percentage values indicate the distribution of each gene group between up and down regulation in STAT5-deficient mouse liver. Response patterns to STAT5 deficiency were significantly different by Fisher's Exact test: between male-biased and female-biased genes, and between female-biased and sex-independent genes, in male livers, and also in female livers; and between male-biased and sex-independent genes in male liver only. Response patterns were also significantly different for male-biased genes between male and female livers, and for female-biased genes between male and female livers.

970

Sex-bias in wild-type liver	Fraction of KO-responsive genes	Change in expression in STAT5-KO	Response to STAT5-KO in male liver		Response to STAT5-KO in female liver	
			Gene count	%	Gene count	%
Male-biased and KO-responsive	(87/132 = 66%)	Increase	6	9.7	33	73.3
		Decrease	56	90.3	12	26.7
Female-biased and KO-responsive	(68/142 = 48%)	Increase	51	94.4	13	38.2
		Decrease	3	5.6	21	61.8
Sex-independent and KO-responsive	(140/6323 = 2.2%)	Increase	65	61.3	58	66.7
		Decrease	41	38.7	29	33.3

971

972

973

974 **Table 2. Differential methylated regions in hepatocyte-specific STAT5-KO liver.** 1493 sex-biased
 975 DMRs (**Fig. 3D**) were classified into DMRs that were more hypomethylated in male liver (787) or more
 976 hypomethylated in female liver (706). These DMRs were further sorted by their direction of change in
 977 STAT5-deficient compared to wild-type liver, as indicated, in both male liver (KOM/WTM) and in female
 978 liver (KOF/WTF).

979

	Methylation change in KOM/WTM	# DMR	%	Methylation change in KOF/WTF	# DMR	%
Male- hypomethylated DMRs (n=787)	Decrease	31	3.9	Decrease	194	24.7
	Increase	282	35.8	Increase	26	3.3
	No change	474	60.2	No change	567	72
Female- hypomethylated DMRs (n=706)	Decrease	124	17.6	Decrease	4	0.6
	Increase	17	2.4	Increase	195	27.6
	No change	565	80	No change	507	71.8

980
 981
 982

983 Figure legends

984 **Fig. 1 – Gene expression changes in STAT5a/STAT5b-deficient mouse liver. (A)** Venn diagram
 985 showing 612 significantly changing genes (FDR<0.05) identified in wild-type (WT) and STAT5 knockout
 986 (KO) male (M) and female (F) mouse liver for three comparisons: WTM/WTF (wild-type male versus
 987 wild-type female; sex-biased genes); KOM/WTM (knockout male versus wild-type male; STAT5
 988 responsive genes in male liver) and KOF/WTF (knockout female versus wild-type female; STAT5
 989 response genes in female liver). The number of differentially expressed genes in each comparison
 990 (labeled A through G) is indicated. **(B)** Heat map representation of log2 fold-change of 155 genes that
 991 showed sex specific expression in wild-type liver (WTM/WTF) and their corresponding changes in
 992 STAT5-KO male (KOM/WTM) and STAT5-KO female (KOF/WTF) livers. These genes correspond to
 993 regions A, B and C in the Venn diagram, as indicated at the top. n, number of genes. **(C)** Scatter plots of
 994 log2 expression fold-change values of sex-biased genes (WTM/WTF) that responded significantly to
 995 STAT5 loss in male liver (116 genes; regions A and B of Venn diagram; top), or that responded
 996 significantly to STAT5 loss in female liver (79 genes; regions B, C of Venn diagram; bottom). The
 997 number of genes in each quadrant of the plot was analyzed by Fisher's exact test to assess whether the
 998 proportions of male and female biased genes that were up regulated and down regulated by STAT5 loss
 999 are significantly different: $p<2.2E-16$ (top) and $p=0.003$ (bottom). Linear regression was performed on
 1000 the absolute values of fold change values to determine the level of correlation between magnitude of sex
 1001 bias and strength of response to STAT5 deletion: $r^2=0.40$, $p=1.2E-13$ (top), and $r^2=0.07$, $p=0.01$ (bottom).
 1002 **(D)** Relationship between STAT5 transcription factor binding and liver gene expression. Genes with one
 1003 or more sex-biased STAT5 ChIP-seq peaks within 10 kb were counted (y-axis) and is indicated in blue
 1004 text as a fraction of all genes in each gene set (top). (Left) In wild-type liver, female-biased genes were
 1005 significantly enriched for having nearby female-biased STAT5 ChIP-seq peaks, and male-biased genes
 1006 for having nearby male-biased STAT5 ChIP-seq peaks. (Middle) In male liver, genes down regulated by
 1007 STAT5-KO were significantly enriched for nearby male-biased STAT5 ChIP-seq peaks, and genes up
 1008 regulated by STAT5-KO were enriched for nearby female-biased STAT5 ChIP-seq peaks. P-values for
 1009 the sex-difference in these response patterns are shown (Fisher's exact test). (Right) In female liver,

1010 STAT5 responsive genes followed the same pattern as those in male liver, albeit with less significant
1011 differential enrichment for sex-biased STAT5 ChIP-seq peaks (higher p-value). (E) Shown are the
1012 number of STAT5 ChIP-seq peaks (male-biased, *top* row of graphs, in blue; female-biased, *middle* row,
1013 in red; or sex-independent, *bottom* row, in green) within 10 kb of each individual gene that is differentially
1014 expressed at FDR<0.05 between: WTM/WTF (274 genes, *left*), KOM/WTM (365 genes, *middle*) and
1015 KOF/WTF (278 genes, *right*), ranked from low to high gene expression fold-change. This represents a
1016 per gene visual representation of the results calculated in (D). See Table S1 for detailed listing of
1017 expression values for these 612 genes.

1018 **Fig. 2 – Sex-independent genes responsive to STAT5 knockout (KO).** (A) Heat map showing impact
1019 of STAT5 deficiency on 140 of 6,323 stringent sex-independent genes whose expression was
1020 significantly increased (red) or decreased (blue) in either male or female STAT5-KO liver (Table S1A,
1021 column S). (B) DAVID pathway analysis of 71 genes up regulated in common in STAT5-KO male and
1022 female liver (Table S1A, column T) (*top*), and of 155 sex-biased genes responsive to STAT5-KO in male
1023 or female mouse liver (Table S1A, column U) (*bottom*). Shown are the top 5 enriched clusters, with bars
1024 indicating $-\log_{10}$ p-values (Benjamini-Hochberg corrected) for enrichment. Results are detailed in Table
1025 S1C and Table S1D. (C) Expression levels of seven STAT family members determined by RNA-seq,
1026 expressed as FPKM values. STAT1 was significantly up regulated in both male and female STAT5-KO
1027 liver. The incomplete loss of STAT5a and STAT5b is consistent with prior reports (42) and likely reflects
1028 STAT5 expressed in non-hepatocytes in mouse liver.

1029 **Fig. 3 – DNA methylation changes in STAT5-KO compared to wild-type mouse liver.** (A) Shown
1030 are the numbers of all DMRs (based on both tile statistic and individual CpG-based DMRs, i.e., the
1031 combined set in Table S2F) that were hypermethylated (orange) or hypomethylated (blue) at standard
1032 ($|15-25|\%$) or robust ($>25|\%$) thresholds for differential methylation in each of the four indicated
1033 comparisons. (B) Box plots of the raw methylation levels of each of the four indicated groups, for the set
1034 of 383 tile-based DMRs that exhibited $> 25\%$ diff.meth between STAT5-KO and wild-type male liver,
1035 showing that STAT5 loss largely results in hypermethylation. ****, $p < 0.0005$ for Student t-test

1036 comparison of STAT5-KO to the sex-matched wild-type group. Analysis is based on n=5-6 individual
 1037 livers per group, as shown in C. **(C)** Heat map showing raw methylation levels for the 383 DMRs
 1038 analyzed in B. Shown are datasets for 23 individual livers, subjected to hierarchical clustering. Color bar:
 1039 raw methylation levels range from 0% methylation (blue) to 100% methylation (yellow). **(D)** Venn
 1040 diagram showing the overlap of sex-biased DMRs in wild-type liver compared to STAT5-KO male and
 1041 female liver, representing a total of 3,615 DMRs (>15% diff. meth). The number of DMRs in each
 1042 comparison (labeled A through G) is indicated. **(E)** Venn diagram overlap of sex-biased DMRs in wild-
 1043 type compared to STAT5-KO mouse liver. Analyses in D and E are based on both tile statistic and
 1044 individual CpG-based DMRs. **(F)** Heat map showing hypermethylated DMRs (orange) and
 1045 hypomethylated DMRs (blue) from Venn diagram comparisons A, B and C in panel D. **(G)** Scatter plots
 1046 showing the relationship between sex-specificity and STAT5 responsiveness of DMRs. Left graph: 454
 1047 sex-biased DMRs, from panels D and E, that are responsive to loss of STAT5 in male liver. Of the sex-
 1048 biased DMRs that are hypomethylated in wild-type male compared to female liver, 282 show increased
 1049 methylation and 31 show increased demethylation in STAT5-KO male liver, and those that are
 1050 hypermethylated in wild-type male compared to female liver, 124 become demethylated and 17 show
 1051 increased methylation in STAT5-KO male liver. Right graph: 419 sex-biased DMRs, from panels D and
 1052 E, that are responsive to STAT5 loss in female liver. STAT5 deficiency results in methylation and
 1053 demethylation of these DMRs with response patterns opposite to those shown in left graph. Fisher's
 1054 exact test for DMR responses between the four quadrants: $p < 2.2E-16$ for both graphs. Linear regression
 1055 analysis of |diff. meth| values showed significant correlations between magnitude of sex-specificity and
 1056 response to STAT5-KO (r^2 values).

1057

1058 **Fig. 4. Few DMRs are at promoters of sex-biased or STAT5-dependent genes.** **(A)** Shown are the
 1059 number of significant differentially expressed genes identified in each comparison (as in Fig. 1A), for
 1060 genes with an RRBS-covered tile or with a significant DMR tile within 5 kb upstream of gene
 1061 transcription start site (promoter region). **(B)** Differential expression and differential methylation data for
 1062 promoter DMR/gene pairs that are consistently male-biased at both the gene expression level and the

1063 promoter DMR level. FC, fold-change. (C) Browser screen shot of genomic region encompassing
1064 *Cyp7b1*, with the methylation level of individual RRBS tiles shown for male and female wild-type and for
1065 male and female STAT5-KO liver. Only the 3'-most RRBS tiles (boxed) correspond to DMRs. The latter
1066 region is expanded at the bottom and encompasses two adjacent DMRs, each containing 4 CpGs,
1067 whose percent methylation is indicated along the y-axis for each of the four conditions. The extent of
1068 differential methylation for these CpGs is shown in B.

1069

1070 **Fig. 5 – Chromatin states at DMRs responsive to STAT5-KO in liver.** (A) ChromHMM emission
1071 probabilities for the 14 chromatin states discovered in male and female mouse liver, from (29). Each
1072 state is colored and given a descriptive name (*left*) based on the frequency of each of six chromatin
1073 mark and presence of DHS, as indicated at the bottom. (B) Cumulative distribution of chromatin states
1074 for the set of 232,882 tiles captured by the 23 RRBS samples used in this study, and for the subsets of
1075 tiles identified as DMRs in the WTM/WTF, KOM/WTM, and KOF/WTF comparisons. The first bar shows
1076 that the whole mouse genome is mostly covered by chromatin state 2, an inactive state. All Tiles
1077 indicates the chromatin state distribution of the full set of 232,882 tiles, which is most strongly enriched
1078 for promoter states (blue) and poised states (yellow) when compared to the whole genome. DMRs
1079 identified in each of the 3 pairwise comparisons are marked >0 (hypermethylation) or <0
1080 (hypomethylation), and show a relative depletion of promoter states and variable increases in the
1081 proportions in enhancer states (various shades of green). For example, hypermethylated DMRs in the
1082 KOM/WTM and KOF/WTF datasets showed an increased proportion of enhancer states. (C) Enrichment
1083 scores for inactive, enhancer, bivalent, transcribed and promoter chromatin states in each DMR group
1084 relative to the full set of tiles captured by RRBS. To calculate enrichment scores, the 14 ChromHMM
1085 chromatin states were grouped into five super states, designated as inactive (states 1, 2, 3), enhancer
1086 (states 5, 6, 8, 9, 10), bivalent (state 12), transcribed (states 13, 14) and promoter (states 7, 8).
1087 Significant enrichment is shown in green and significant depletion in purple (p-value cutoff: 1E-5, two
1088 proportion z-test). (D) Frequency of occurrence of each of the 14 chromatin states in male and female
1089 liver, colored as indicated in B, was calculated in 100 bp windows extending 2 kb upstream and 2 kb

1090 downstream of the hypermethylated and hypomethylated DMRs in each comparison. Chromatin state 6
1091 showed a prominent peak in hypermethylated DMRs for the KOM/WTM and KOF/WTF comparisons, in
1092 both male and female liver, and to a less extent, in hypomethylated DMRs of the WTM/WTF comparison
1093 in the male liver, as marked. (E) Genes that are sex-biased or STAT5-responsive show a significant
1094 enrichment for being in the same TAD as a sex-biased and/or STAT5-responsive DMR. **, Z-score >2.5.
1095 Inactive chromatin state E2 shows fewer overlapping TADs than a random sampling of CpG regions
1096 (also see Table S4C). (F) Box plots showing distribution of distances to the nearest sex-biased and/or
1097 STAT5-responsive gene TSS for sex-biased and/or STAT5-responsive DMRs grouped by their
1098 chromatin state, as shown along the x-axis. See Table S4 for datasets used in these analyses.
1099

1100 **Fig. 6 – GSEA analysis of STAT5-responsive DMRs nearby sex-biased DHS.** DMRs in the three
1101 comparisons indicated at the bottom were ranked from hypermethylation (red) to hypomethylation (blue).
1102 GSEA was used to determine whether sex-biased DHS (Table S5) are preferentially found within 5 kb of
1103 the hypermethylated or the hypomethylated end of the ranked DMR spectrum. NES: normalized
1104 enrichment score; FDR < 0.05 was considered significant. For example, DMRs that were
1105 hypomethylated in male liver were enriched for male-biased DHS (NES: -3.18, FDR=0), while DMRs
1106 hypomethylated in male liver (and thus hypomethylated in female liver) were enriched for female-biased
1107 DHS (NES=1.57; FDR=0.007).
1108

1109 **Fig. 7 – STAT5 motifs and ChIP-seq peaks at DMRs.** (A) Enrichment of STAT5 motifs at DMRs
1110 determined using the MEME Analysis of Motif Enrichment tool. (B) Categorization of 14,910 STAT5
1111 ChIP-seq sites in mouse liver (12) based on whether they are within 5 kb of one of the 232,882 RRBS
1112 tiles captured in this study, or whether they are within 5 kb of one of the RRBS that is also a DMR. Here,
1113 DMR refers to 3,615 regions that showed significant differential methylation in one or more of these
1114 three comparisons: WTM/WTF, KOM/WTM, and KOF/WTF. (C) Cumulative frequency plots of
1115 hypermethylated and hypomethylated sex-biased and STAT5-responsive DMRs and their distances to
1116 the nearest STAT5 binding site in each set of sites identified in B (Table S2F, last columns).

1117

1118 **Fig. 8 – Enriched DNA motifs at DMRs.** (A) Number of enriched DNA motifs determined by Homer that
 1119 passed the significance filter of q-value >0.05. (B) Heat map showing the clustering of DNA motifs
 1120 numbered in A, indicating the extent of overlap between the 6 indicated sets of motifs (top). (C) Motif
 1121 name, q-value and sequence logo for the top six *de novo*-discovered DNA motifs (lowest q- value) for
 1122 the indicated 6 sets of hypermethylated and hypomethylated DMRs. See Table S6 for a complete listing
 1123 of significant motifs.

A

B Sex-independent STAT5-dependent genes (Up, common)

Sex-specific STAT5-dependent genes

C

Fig. 4

A

Comparison	# Genes	# Genes with RRBS tiles in promoter	# Genes with DMR in promoter
WTM / WTF	274	152	6
KOM / WTM	373	247	13
KOF / WTF	291	185	7

B

				RNA expression: log2FC			DNA methylation: %		
Gene category	Gene Name	bp to DMR in promoter	DMR_ID	WTM/ WTF	KOM/ WTM	KOF/ WTF	WTM/W TF	KOM/ WTM	KOF/ WTF
Male-biased genes	Cyp7b1	3113	chr3_18146401	2.40	-2.85	-1.51	-24.4	49.4	37.2
	Cyp7b1	3213	chr3_18146501	2.40	-2.85	-1.51	-23.0	47.5	30.8
	Cspg5	4136	chr9_110142101	4.35	-2.84	0	-19.1	16.2	-37.0

DMR, expanded:

Fig. 6

Fig. 7