Unbiased, Genome-Wide In Vivo Mapping of Transcriptional Regulatory Elements Reveals Sex Differences in Chromatin Structure Associated with Sex-

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th>Citation</th>
<th>Ling, Guoyu et al. “Unbiased, Genome-Wide In Vivo Mapping of Transcriptional Regulatory Elements Reveals Sex Differences in Chromatin Structure Associated with Sex-Specific Liver Gene Expression.” Molecular and Cellular Biology 30.23 (2010): 5531-5544.</th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1128/MCB.00601-10">http://dx.doi.org/10.1128/MCB.00601-10</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>American Society for Microbiology</td>
</tr>
<tr>
<td>Version</td>
<td>Author's final manuscript</td>
</tr>
<tr>
<td>Accessed</td>
<td>Thu Jan 03 06:37:33 EST 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/60028">http://hdl.handle.net/1721.1/60028</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td></td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Unbiased, Genome-wide in vivo Mapping of Transcriptional Regulatory Elements Reveals Sex Differences in Chromatin Structure Associated with Sex-specific Liver Gene Expression

Guoyu Ling*, Aarathi Sugathan*, Tali Mazor, Ernest Fraenkel, and David J. Waxman#

Division of Cell and Molecular Biology, Department of Biology,
Boston University, Boston, MA 02215 (GL, AS, and DW).
Department of Biological Engineering, Massachusetts Institute of Technology,
Cambridge, MA 02139 (TM and EF)

Running title – Global DNase hypersensitivity map of mouse liver

Keywords –DNase I hypersensitivity, growth hormone, liver, sex-specific gene expression, DNase-seq

*Co-first authors

Disclosure: The authors have no financial interests to declare.

# To whom correspondence should be addressed at:
Department of Biology, Boston University
5 Cummington Street, Boston, MA 02215
Tel: 617-353-7401; Fax: 617-353-7404 ; Email: djw@bu.edu

Word counts:
Materials and Methods: 2,327
Introduction, Results, and Discussion: 4,708 words.
Abstract – We have used a simple and efficient method to identify condition-specific transcriptional regulatory sites \textit{in vivo} to help elucidate the molecular basis of sex-differences in transcription, which are widespread in mammalian tissues and affect normal physiology, drug response, inflammation and disease. To systematically uncover transcriptional regulators responsible for these differences, we used DNase hypersensitivity analysis coupled with high-throughput sequencing to produce condition-specific maps of regulatory sites in male and female mouse liver, and for livers of male mice feminized by continuous infusion of growth hormone (GH). We identified 71,264 hypersensitive sites, with 1,284 showing robust sex-differences. Continuous GH infusion suppressed the vast majority of male-specific sites and induced a subset of female-specific sites in male liver. We also identified broad genomic regions (up to ~100kb) showing sex-dependent hypersensitivity and similar patterns of GH response. We found a strong association of sex-specific sites with sex-specific transcription; however, a majority of sex-specific sites were >100kb from sex-specific genes. By analyzing sequence motifs within regulatory regions, we identified two known regulators of liver sexual dimorphism, and several new candidates for further investigation. This approach can readily be applied to mapping condition-specific regulatory sites in mammalian tissues under a wide variety of physiological conditions.
Sexual dimorphism in gene expression is common in mammalian somatic tissues (23), and has broad implications for human health. Sex differences in gene expression may contribute to differences between men and women in the prevalence, extent, and progression of disease, including autoimmune diseases (54), kidney disease (37), cardiovascular disease (45) and liver diseases, including hepatocellular carcinoma (8, 58). In addition, sex differences in pharmacokinetics and pharmacodynamics are common and may affect drug response (52). Sex differences in gene expression have been widely studied in liver, where they affect >1,000 transcripts (5, 51, 57) and impact physiological and pathophysiological functions ranging from lipid and fatty acid metabolism to xenobiotic metabolism and disease susceptibility (52). In liver, sex differences in gene expression are primarily determined by growth hormone (GH) signaling (3, 21), which shows important sex-differences that reflect the sex-differences in plasma GH profiles seen in many species, including rats, mice and humans (53).

The underlying mechanisms of sexual dimorphism in mammalian tissues have been only partly elucidated at the molecular level. In male rat liver, intermittent plasma GH pulses repeatedly activate the latent cytoplasmic transcription factor STAT5b, whose activity is essential for liver sex differences (5). The more continuous, female-like pattern of pituitary GH secretion can be mimicked by continuous GH infusion of male mice, which abolishes the normal male, pulsatile, plasma GH profile and feminizes liver transcript patterns by suppressing many male-specific genes and inducing many female-specific genes (19). In spite of these findings, the molecular mechanisms whereby STAT5b and other transcription factors regulate liver sex specificity have remained elusive (26, 52).
DNase I hypersensitivity (DHS) analysis is a powerful tool to identify functional DNA elements involved in gene regulation. The temporal and spatial association of DHS sites with tissue-specific and developmentally regulated gene expression has long been established (14), and several instances have been reported where sex differences in DNase hypersensitivity characterize genes that show sex-dependent transcription. Early studies identified a male-predominant DHS site in mouse liver upstream of C4a (Sex-limited protein), where an open chromatin structure correlated with a male-predominant pattern of gene expression (16), and examples of sex-regulated DHS sites have been reported for two sex-specific cytochrome P450 (Cyp) genes in rat liver (10, 47).

In order to identify sex differences in mouse liver chromatin structure on a global scale, we combined DHS analysis with ultra high throughput sequencing (DNase-seq) to probe open chromatin structure at single bp resolution (1, 6, 40, 46). We show that DNase-seq, whose application until now has been limited to cultured cell lines, can readily be used to map DHS sites in mammalian liver, despite the added complexity of multiple cell types. We obtained high resolution, genome-wide DHS maps for both male and female mouse liver under physiological conditions, and we demonstrate that these maps can be utilized to identify transcriptional regulators of sex-biased liver gene expression. We characterize more than 70,000 DHS sites and show that they encompass a large fraction (~70-90%) of binding sites for six liver transcription factors identified earlier by ChIP-seq. We identify 1,284 DHS sites that show robust sex differences and may contribute to sex-dependent gene expression, ~20% of which mapped within 100 kb of a sex-specific liver gene. In addition, a subset of the sex-dependent DHS sites is shown to respond to continuous GH treatment in male mice, likely representing functional DNA elements that mediate hormone-dependent, sex-dependent gene expression. Finally, analysis of the sex-dependent DHS sequences for enriched
motifs identified binding sites for two transcription factors, STAT5b and HNF4α, known to be essential for sex-specific liver gene expression (5, 18, 20), as well as binding sites for several novel factors, not previously implicated in liver sexual dimorphism. These findings highlight the utility of DNase-seq for elucidating condition-specific transcriptional regulatory sites associated with complex biological processes in mammalian tissue in vivo on a genome-wide scale.

MATERIALS AND METHODS

Mouse studies – Adult male and female ICR mice (CD-1 mice) were purchased from Taconic Farms, Inc. (Germantown, NY) or Charles River Laboratories (Wilmington, MA), and were housed in the Boston University Laboratory of Animal Care Facility in accordance with approved protocols. Livers were collected from 8-wk old mice euthanized by CO₂ asphyxiation followed by cervical dislocation. Continuous GH treatment of 7-week male mice was achieved using Alzet model 1007D micro-osmotic pumps (Durect Corporation, Cupertino, CA) implanted s.c. under ketamine and xylazine anesthesia delivering recombinant rat GH (obtained from Dr. Arieh Gertler, Protein Laboratories Rehovot, Ltd, Rehovot, Israel) at 20 ng rat GH/hr/gram body weight for 7 days (19). RNA was extracted from individual livers using Trizol reagent (Invitrogen, Carlsbad, CA) followed by reverse transcription using 1 µg total RNA and a High Capacity Reverse Transcription Kit (Applied Biosystems, Carlsbad, CA). To verify feminization of liver gene expression by continuous GH infusion, real-time PCR analysis of a panel of established, continuous GH-responsive genes (19) was performed using Power SYBR Green PCR Master Mix and an ABS 7900HT sequence detection system (both from Applied Biosystems).
For \textit{in vivo} transfection assays, genomic regions corresponding to six individual DHS sites were PCR-amplified from ICR mouse genomic DNA and cloned into a modified Promega pGL4.10 vector (Promega, Madison, WI), designated pAlbpmo, that includes a minimal mouse \textit{Alb} promoter driving expression of a Firefly luciferase gene. The DHS regions were cloned into pAlbpmo, upstream of the \textit{Alb} promoter. A Renilla luciferase reporter vector was constructed in a similar way by inserting an \textit{Alb} enhancer into a modified pGL4.70 reporter vector containing the same mouse minimal \textit{Alb} promoter. Twelve \(\mu\)g of Firefly reporter vector and 3 \(\mu\)g of Renilla reporter plasmids were delivered to adult male and female mouse liver by hydrodynamic injection using the TransIT gene delivery system (Mirus, Madison, WI). Livers were harvested 7-days after injection, homogenized in 1X Passive Lysis Buffer, and Firefly luciferase activity normalized to Renilla luciferase activity was assayed using a dual luciferase assay kit (Promega).

\textbf{Nuclei isolation} – Nuclei were isolated using a high-sucrose based protocol to minimize perturbation of chromatin structure during nuclei isolation (27). Livers were pooled from 2-5 mice, minced and then homogenized in nuclear homogenization buffer (10 mM HEPES pH 7.9, 25 mM KCl, 0.15 mM spermine, 0.5 mM spermidine, 1 mM EDTA, 2 M sucrose, 10% glycerol, 10 mM NaF, 1 mM orthovanadate, 1 mM PMSF, 0.5 mM DTT) supplemented with 1X protease inhibitor cocktail (Sigma P8340) using a motor-driven Potter-Elvehjem homogenizer (5 ml homogenization buffer per g liver tissue). 25 ml of homogenate was layered over a 10 ml cushion comprised of the same buffer, and centrifuged at 25,000 rpm for 45 min at 4°C in an SW28 rotor. The pelleted nuclei were suspended in nuclei storage buffer (20 mM Tris-Cl, pH 8.0, 75 mM NaCl, 0.5 mM EDTA, 50% (v/v) glycerol, 0.85 mM DTT and 0.125 mM PMSF) using a Dounce homogenizer, counted using a hemacytometer after ~100-fold dilution in phosphate-buffered saline, and snap frozen in aliquots of ~5-10 x 10^7 nuclei/ml at –80°C.
DNase I digestion – DNase I digestion was performed as described (40) with some modifications. Frozen nuclei were thawed on ice and washed twice with ice-cold Buffer A (15 mM Tris-Cl, pH 8.0, 15 mM NaCl, 60 mM KCl, 1 mM EDTA, 0.5 mM EGTA, 0.5 mM spermidine). DNase digestion was initiated by incubating ~5 x 10^6 nuclei in 1 ml of Buffer D (9 vol of buffer A + 1 vol of 60 mM CaCl_2, 750 mM NaCl) for 2 min at 37°C with an optimized amount of RQ1 DNase I (Promega) (see below). Six tubes of nuclei were incubated in parallel. DNase digestion was halted by the addition of 1 ml stop buffer (50 mM Tris-Cl, pH 8.0, 100 mM NaCl, 0.1 % SDS, 100 mM EDTA) to each tube, followed by proteinase K digestion (100 μg/ml final concentration) overnight at 55°C. The next day, phenol chloroform extraction was performed, after which the aqueous phase (~11-12 ml from the 6 parallel DNase digestion reactions) was removed and adjusted to 0.8 M NaCl by addition of 5 M NaCl. Control samples were prepared in the same way by incubating ~5 μg purified genomic DNA in 0.1 ml digestion buffer with 0.0625, 0.125, 0.25, or 0.5 Unit of DNase I. The control DNase I digestion sample that yielded a smear of DNA fragments ranging from 100 bp to ~1.5 kb was selected and further purified by sucrose gradient ultracentrifugation.

Small DNA fragments (< 1.5 kb) released during DNase I digestion were isolated by sucrose step gradient ultracentrifugation of the DNase-digested nuclei, and also from the DNase-digested genomic DNA and the sonicated genomic DNA control samples. The small DNA fragments released from the digested nuclei correspond to DNase sensitive regions containing multiple cut sites in close proximity, whereas larger DNA fragments present in the samples primarily result from random, non-specific DNA degradation. The released fragments were size-selected on a sucrose step gradient prepared by sequentially layering 3 ml of each sucrose concentration (20 mM Tris-Cl, pH 8.0, 5 mM EDTA, 1 M NaCl containing either 40%, 35%, 30%, 25%, 20%, 17.5%, 15%, 12.5%, or 10%
sucrose) in a SW28 tube. Half of each sample (~5.7 ml) was loaded on the top of each gradient, and the gradient was centrifuged for 24 h at 25,000 rpm at 25°C in a SW28 rotor. Fractions of 1.9 ml were collected from the top and assayed by agarose gel electrophoresis. Gels were stained with 1X SYBR Green I nucleic acid gel staining solution (Invitrogen) in 1X TAE buffer (40 mM Tris-acetate, 2 mM EDTA, pH 8.5) at room temperature for 30 min, and visualized on a Typhoon Imager (GE Healthcare, Piscataway, NJ). Fractions that primarily contain DNA fragments <1.5 kb (typically fraction 7) were purified on Qiagen columns (Cat. No. 28704) and assayed by qPCR using primers designed to amplify known hypersensitive genomic regions (Table S1). The typical yield was 1-2 ng released DNA fragments per 10^6 nuclei, as determined using Quanti-IT kit (Invitrogen). Illumina sequencing (see below) was carried out for each of two independent pools of biological replicates (n=4-5 livers for each sex in each replicate). Each sequenced sample was comprised of ~ 30 ng DNA pooled from at least 3 independent batches of DNase digested nuclei, to minimize the impact of inter-sample variability in DNase digestion.

Optimization of DNase digestion conditions – qPCR primers used to optimize DNase digestion conditions and to assess the quality of DNase-released fragments are listed in Table S1: Intergenic-1 primers were used to amplify a genomic region distant (>100 kb) from any known genes, where no DHS regions were expected; and Rassf6 primers were selected to flank a strong DHS site near the Rassf6 promoter. For each batch of DNase I enzyme, initial experiments were carried out over a range of DNase I concentrations (5 to 120 Unit/ml) to identify a DNase concentration (typically ~ 40 Unit DNase I/ml) that resulted in < 20% copy-number loss in Intergenic-1 region, but >50% copy-number loss in the Rassf6 promoter region. Purified, released DNase fragments were routinely tested for enrichment of Alb promoter sequences over intergenic region sequences to verify the quality of the DHS samples prior to Illumina sequencing analysis using Alb primers, located within a
strong DHS site within the Alb promoter, and Intergenic-2 primers, which are close to Intergenic-1 primers. Typically, the DNase-released Alb promoter fragments showed >16-fold enrichment compared to control genomic DNA. Sex-dependent release of genomic DNA fragments was also routinely verified for a male-specific DHS region near Ttc39c and a female-specific DHS region near Cyp2b9 (Table S1).

Illumina sequencing – Sequences of DNase I-released DNA fragments were sequenced using an Illumina Genome Analyzer II instrument (Illumina, Inc., San Diego, CA). Briefly, about 30 ng of DNase-released DNA fragments were subjected to end repair, adaptor ligation, and PCR enrichment using an Illumina sample preparation kit following the manufacturer’s recommendation. A DNA library comprised of ~100-300 bp fragments was size-purified by gel electrophoresis. The concentration of properly ligated samples was estimated by qPCR, followed by cluster generation and sequencing. The sequencing reads were aligned to mouse genome mm9 using Illumina's Eland extended software with a maximum of 2 mismatch allowed in the first 25 bp. A total of 36 million, 32 million, and 28 million reads were sequenced from the male, female, and GH-treated male samples, respectively, with ~82% mapped to unique genomic positions (Table S2). All high-throughput sequencing data are available on the GEO database (accession GSE-21777).

Identification of DHS sites – PeakSeq (39) was modified as outlined below and used to identify DHS sites in male, female, and GH-treated male mouse liver nuclei in comparison to control samples; these consisted of sonicated mouse genomic DNA, as well as DNase I-digested mouse genomic DNA, and were processed in parallel to the DNase I-digested nuclei. Sex-specific DHS sites, and sites induced or suppressed in male liver by continuous GH treatment were also identified. The PeakSeq algorithm was modified as follows: (a) sequence read numbers from the two samples being
compared were square-root normalized to improve the linearity of the data for calculation of the scaling factor by linear regression; (b) when comparing sequence reads from male vs. female liver, or from male vs. continuous GH-treated male liver, only putative peak regions were included in linear regression due to the presence of varying amounts of background between samples (Table S3); (c) a minimum threshold of 7 sequence reads for autosomes and 5 sequence reads for sex chromosomes was applied to all putative DHS peaks to eliminate unusually long (>10 kb) peaks with few sequence reads; and (d) putative peaks that were <100 bp in length were extended to 100 bp and then evaluated for statistical significance. Further details are provided in Supplementary Methods and Results.

Association of DHS sites with genes, H3K4 methylation, and FoxA2 binding sites – DHS sites were classified as intergenic, coding, or associated with promoter regions based on mapping to known genes, mRNAs, and spliced and unspliced ESTs in the UCSC genome browser. To compare DHS sites with liver gene expression, the locations of sex-independent and sex-specific DHS sites were mapped to sex-specific and sex-independent genes expressed in mouse liver (50). ChIP-seq data for H3K4-me1, H3K4-me3, and FOXA2 binding in female mouse liver (38) were compared to locations of sex-independent and sex-specific DHS sites. The association of DHS sites with the above two histone modifications was calculated by determining the numbers of DHS sites that have a histone modification site within 150 bp of either side of the DHS site. To generate distribution plots, distances from the midpoints of DHS sites to sequence tags previously determined by ChIP-seq for H3K4-me1, H3K4-me3 and FOXA2 were computed.

Gene expression data – Liver gene expression data was obtained from an earlier study (50), where a total of 1,380 transcripts showed significant differences in expression between male and female liver. After removing microarray probes that do not map to any known gene, probes that map to the
same transcript as another probe, and probes mapping to chromosomes for which no DNase
hypersensitivity data is presented (chrY, chrUn, and chrN_random), a total of 1,209 genes showing
>2-fold sex-differences in expression at \( p \leq 0.005 \) remained (sex-specific genes), as well as 21,153
sex-independent genes. Of these, 343 sex-specific genes and 7,341 sex-independent genes met the
additional criterion of microarray signal intensity \( \geq 500 \) in liver.

Identification of broad DHS regions – SICER (59), an algorithm that uses a clustering approach to
identify extended enriched domains from histone modification ChIP-seq data, was applied to detect
broad regions of the genome that are enriched for DNase-seq reads in male or female liver samples
compared to control. Genomic regions that show sex-dependent DNase hypersensitivity were
defined as those that were significantly enriched in male compared to female, or in female compared
to male, DNase-digested liver nuclei. Similarly, genomic regions that respond to continuous GH
treatment were identified by comparing the untreated male samples to the GH-treated male samples.
A window size of 200 bp and a gap size 1200 bp were used, and significant regions were chosen that
met FDR < 10^{-3} and fold-difference \( \geq 2 \) for the pair of liver samples being compared.

Enriched transcription factor binding sites in sex-specific DHS sites – THEME, a hypothesis-based
algorithm that tests whether a given motif separates a foreground set of sequences from a
background set (29), was used to identify enriched motifs in the following 18 sets of sites compared
to sex-independent sites: (A) male-specific sites either (i) within 10 kb or (ii) up to 50 kb from the
transcription start site (TSS) of a sex-specific gene; (B) male-specific sites within 10 kb or 50 kb of a
sex-independent gene; and (C) male-specific sites distant from any gene; each group of DHS sites
was divided into subgroups that respond and subgroups that do not respond to continuous GH
treatment in males at \( p < 0.01 \) and fold-difference \( > 2 \), and similarly for female-specific DHS sites.
Transcription factor binding profiles for 97 families of transcription factors were generated by clustering the vertebrate transcription factor position-specific scoring matrices (PSSMs) from the TRANSFAC and JASPAR databases (2, 33). The corresponding 97 motifs were considered enriched if they met the following conditions: cross-validation error <0.4, \( p \)-value <0.001, normalized log-likelihood ratio score > 0.4, and enrichment compared to sex-independent sites > 2.

Motifs with total information content < 8 bits were eliminated from further consideration. To identify the transcription factor(s) associated with each motif, the refined family binding profiles were matched back to the TRANSFAC and JASPAR databases using STAMP (32), and the top factor(s) that matched with an E-value < \( 10^{-8} \) were identified. One exception was motif 44, for which the best matches, to Fox factors, had an E-value < \( 10^{-6} \). Discovered motifs were clustered by hierarchical clustering by Pearson correlation of fold enrichment over sex-independent sites in each of the 18 sets of sex-specific sites using the hierarchical clustering module of the GenePattern suite of tools (36).

RESULTS

Generation of liver DHS maps – DNase-seq was used to generate genome-wide DHS maps for liver tissue obtained from male and female mice, and from male mice given a continuous infusion of GH for 7 days, which feminizes the pattern of liver gene expression (19). Mouse liver nuclei prepared from two independent pools of biological replicates were digested with DNase I under optimized conditions (see Materials and Methods), and fragments released from hypersensitive regions were separated from randomly cut DNA fragments, which tend to be much larger in size (40). DNase-released fragments ranging from ~100-300 bp were sequenced using Illumina sequencing technology. The final combined data set is comprised of 29 million sequence reads mapped to
unique locations in the mouse genome for male liver and 26 million reads for female liver; 23
million additional sequence reads were obtained for continuous (7 day) GH-treated male liver (~82%
uniquely mappable reads; Table S2). The resultant DHS maps are of high quality, as seen in Fig. 1A
for the Alb gene region. Eight DHS regions were identified within ~47 kb of the Alb gene TSS, with
very low background between peaks of hypersensitivity. In addition to the DHS sites at -0.1 kb, -3.5
kb, -10.8 kb and -13.7 kb relative to the Alb TSS, previously identified using classical Southern
blotting methods (28), we identified DHS sites at four upstream locations, from -22 kb to -47 kb
(Fig. 1A). Closer examination of the -13.7 kb DHS site revealed a typical structure for a DHS peak,
with a roughly symmetric distribution of positive and negative strand digestion sites that clearly
define the DHS peak boundary (Fig. S1B). Using DNase I digested genomic DNA as a control, we
identified 71,264 DHS sites in male and female liver, covering 1.8% of the mouse genome. 48,762
of the DHS sites were high stringency sites, and the total number of DHS sites increased to 110,785
when using the combined data sets (Table 1 and Table S4). There is a high degree of overlap
between DHS sites and transcription factor binding sites identified by ChIP-chip or ChIP-seq
analysis of mouse liver (Table 2), ranging from 67% for CEBPA (42) to 93% for FXR/NR1H4 (49).
Thus, the DHS sites identified here likely include a large fraction of the active regulatory elements in
liver tissue.

DHS sites showing sex specificity and responsiveness to GH – We hypothesized that liver chromatin
is characterized by sex differences in accessibility to DNase, and that these differences relate to the
observed sex differences in liver gene expression. We further anticipated that continuous GH
treatment of male mice, which feminizes the overall pattern of liver gene expression (19), will alter
the sex-dependent patterns of chromatin accessibility. By comparing the DHS profiles of male and
female mouse liver we identified genomic regions showing significantly greater DNase I fragment
release from male compared to female mouse liver nuclei, i.e., male-specific DHS sites; correspondingly, female-specific DHS sites showed significantly greater DNase I cleavage in female liver. 850 male-specific peaks and 434 female-specific peaks were identified as high stringency sex-specific DHS sites based on their confirmation in each of two independent biological replicates (Table 1); examples are shown in Fig. 1B and Fig. 1C and in Fig. S4. A total of 4,182 sex-specific DHS sites were identified at lower stringencies (Table 1 and Tables S5B-S5E). Continuous GH treatment of male mice suppressed 82% of the high stringency male-specific DHS sites and induced 26% of the high stringency female-specific DHS sites, whereas <3% of sex-independent DHS sites were GH-responsive at $p<0.01$ and fold-difference $>2$ (Table 3). When weaker GH responses are included, 98% of male-specific DHS sites were suppressed and 44% of female-specific DHS sites were induced (Table S5H).

Examination of the distribution of the 1,284 high stringency sex-specific sites across chromosomes revealed greater enrichment of female-specific DHS sites on chromosomes 5 and X, and enrichment of male-specific DHS sites on chromosomes 3 and 18, compared to the overall list of DHS sites (Table S6). Overall, 65% of sex-specific DHS sites are in the coding region or within 5 kb of the TSS of a known transcript, as compared to 78% of sex-independent DHS sites (Fig. S5). The median lengths of sex-specific and sex-independent DHS were similar, 466-575 bp and 437-483 bp, respectively (Table S7), corresponding to the depletion of ~2 nucleosomes.

Association of liver DNase hypersensitivity with liver gene expression – In CD4$^+$ T cells, the probability that a given gene harbors a 5’-proximal DHS site increases with the level of gene expression (1). We observed the same trend in mouse liver, where the proportion of sex-independent genes that have a DHS site within 200 bp of the TSS increased with increasing intensity
of gene expression, leveling off at ~90% (Fig. 2A). In contrast, the proportion of genes that show sex-specific expression (50) and have a 5'-proximal DHS site increased more gradually with increasing expression (Fig. 2B; \( p = 0.0006 \)). The overall lower percentage of sex-specific genes with a 5'-proximal DHS site might indicate that these genes are more commonly regulated by distal elements. Alternatively, the sex-independent genes might simply be close to more non-functional DHS sites than are sex-specific genes. Next, we tested the hypothesis that genes that show sex-specific expression in mouse liver are more likely to be associated with sex-specific DHS sites, compared with sex-independent genes. Supporting this hypothesis, we observed that sex-specific genes are 8.1-fold more likely than sex-independent genes to have a sex-specific DHS site in the coding region, and 3.1-fold more likely within 100 kb; furthermore, the distance to the nearest sex-specific DHS site rises more steeply for sex-specific genes than for sex-independent genes (Fig. 3A, left). Conversely, sex-specific DHS sites are more likely than sex-independent DHS sites to be located near a sex-specific gene (Fig. 3B, left). Finally, the proportion of male-specific genes whose nearest DHS site is also male-specific (20%) is ~10-fold greater than those whose nearest DHS site is either female-specific or sex-independent (2% in both cases), and similarly for female-specific genes and female-specific DHS sites (Fig. 3C). Thus, there is a strong association between sex-specific DHS sites and sex-specific gene expression. However, this association is seen for only a subset of sex-specific genes, insofar as only 20% of sex-specific genes have a high stringency sex-specific DHS site in the coding region, and only 43% have a sex-specific DHS site within 100 kb. This compares to 90% of liver-expressed sex-independent genes with a sex-independent DHS site in the coding region, and 99% with at least one sex-independent DHS site within 100 kb (Fig. 3A, right). Moreover, only 23% of sex-specific DHS sites are within 100 kb of a sex-specific gene, while 76% of sex-independent DHS sites are within 100 kb of a sex-independent gene (Fig. 3B, right). Sex-specific DHS sites may therefore act as distant regulators. Alternatively, this finding
may reflect more complex regulatory mechanisms of sex-specific genes, involving interactions between multiple regulatory sites and multiple genes (30), regulatory changes that have no effect on chromatin structure, or post-transcriptional regulation. The subsets of sex-specific genes that do and do not have a sex-specific DHS site within 100 kb include equal proportions of male-specific and female-specific genes; however, the female-specific genes that are within 100 kb of a sex-specific DHS site are 2.7-fold enriched \( (p<10^{-4}) \) for the subset of female-specific genes that are suppressed in female liver upon ablation of pituitary of GH stimulation (class I female-specific genes; (50)). The extensive loss of male-specific DHS sites and the induction of a substantial fraction of female-specific DHS sites in livers of continuous GH-treated male mice (Table 3), where the gene expression profile is feminized (19), supports the conclusion that these sex-specific DHS sites play a functional role in the sex-specific expression of the genes associated with these sites. Indeed, the subset of female-specific DHS sites that respond to continuous GH is even more frequently associated with female-specific genes compared to the full set of female-specific DHS sites (Fig. 3C).

**Enhancer-like activity of sex-specific DHS sites** – Our finding, above, that sex-specific DHS sites are more likely to be associated with genes of the same sex specificity suggests that sex-specific DHS sites serve as enhancers of sex-specific gene expression. This possibility is supported by a comparison of our DHS map with maps of histone H3 lysine-4 mono- and tri-methylation (H3K4-me1 and H3K4-me3, respectively) reported for female mouse liver (38): 80% of high stringency female-specific DHS sites are within 150 bp of nucleosomes marked by H3K4-me1 but not H3K4-me3, whereas only 15% are associated with a H3K4-me3 mark (Fig. 4A). This pattern – presence of H3K4-me1 in the absence of H3K4-me3 – is indicative of an enhancer (15). A smaller proportion of sex-independent DHS sites exhibit an enhancer-like H3K4-methylation profile, with 32% of these...
DHS sites containing the H3K4-me3 mark and only 61% showing the H3K4-me1-only pattern (Fig. 4A). The frequency of the H3K4-me1 + H3K4-me3 double mark decreased dramatically with increasing distance from the promoter, as was seen for both female-specific and sex-independent DHS sites (Fig. 4A). Both histone marks exhibited a trough at the midpoint of female-specific and sex-independent DHS sites, indicating nucleosome depletion (Fig. 4B and Fig. 4C).

To assay for enhancer activity, we selected 6 sex-specific DHS sites, 5 of which were responsive to continuous GH treatment (Table S8), and cloned them into a reporter vector containing a modified Alb promoter linked to a luciferase reporter (56). The 6 sex-dependent DHS sites were assayed for their intrinsic ability to enhance the Alb promoter following in vivo liver transfection by hydrodynamic injection. Five of the six sites exhibited enhancer activity when assayed 7 days after liver transfection (Fig. 5). This time point was selected to allow for decay of the transiently high activity of the Alb promoter using this transfection method (56). The most active DHS fragment, from intron 2 of the highly female-specific Cux2 gene (Table S8) (24), was >200-fold more active than the Alb promoter alone, but showed similar activity in male and female mouse liver. Two male-specific DHS sites showed 8 to 17-fold higher activity than the Alb promoter, with 3-fold higher activity seen in male compared to female liver for a Cyp2d9 DHS site (Fig. 5). A female-specific DHS site adjacent to Acot4 showed female-enriched enhancer activity, albeit at a modest level. One of the six DHS sites (Cyp2c39) was inactive.

Broad genomic regions of DNase hypersensitivity – While 87% of the above-identified DHS sites are <1 kb in length, we observed genomic regions with considerably longer sex-dependent hypersensitivity, some extending up to ~100 kb. To identify such regions of extended DNase hypersensitivity, we used SICER (59), a clustering based algorithm designed to identify diffuse
domains of ChIP-enriched regions. We found 3,971 DHS regions >10 kb in length, 58 of which showed significant sex-differences (Table 4). Continuous GH treatment suppressed 50% of the extended male-specific regions and induced 47% of the female-specific regions, as compared to <0.2% of the sex-independent regions; the proportion of these >10 kb female-specific regions that are induced by GH is even higher than that for the short female-specific DHS peaks. Some of the extended DHS regions are comprised of clusters of the short DHS peaks identified above (Fig. 6A and 6B; also see Supplemental Text and Fig. S6), while other extended DHS regions contain few sites identified as DHS peaks by PeakSeq, which is optimized for identification of short, well-defined discrete peaks (Fig. 6C and Fig. 6D; track marked All DHS sites). Additional examples, including GH responses, are shown in Fig. S6. The full list of SICER-identified regions is provided in Table S9.

Transcription factor binding sites enriched in sex-specific DHS sites – THEME, a hypothesis-based algorithm that tests for enrichment of pre-defined motifs (29), was used to examine sex-specific DHS sites for enrichment of transcription factor binding site motifs using sex-independent DHS sites as a background. Given the expected heterogeneity of sex-specific DHS sites, we carried out these analyses using subsets comprised of male- and female-specific DHS sites that are: (1) within 10 kb or within 50 kb of the TSS of a sex-specific gene; (2) within 10 kb or within 50 kb of a sex-independent gene; and (3) distant (>50 kb) from any gene. Each set of DHS sites was further divided into sites that respond and sites that do not respond to continuous GH treatment in males at $p<0.01$ and fold-difference $>2$ (Table S5A). Starting with motif families derived from the TRANSFAC and JASPAR databases, we identified 16 enriched motifs (Table S10). The sets of sex-specific sites were then scanned for each of the 16 motifs, which were then clustered according to fold-enrichment in each of the sets of DHS sites (Fig. 7).
The discovered motifs include binding sites for two factors known to be required for sex-specific liver gene expression. Thus, a motif matching the binding site for STAT5b (motif 28) is enriched in male-specific GH-responsive sites, as is a motif matching HNF4α (motif 70), consistent with the earlier findings that these two transcription factors are essential for GH-regulated sex-specific gene expression in male mouse liver (5, 18, 20). While the HNF4α-like motif is most highly enriched in sites within 10 kb of a sex-specific gene, the STAT5 motif shows highest enrichment in more distal sites, including sites proximal to sex-independent genes, consistent with other studies on STAT5 binding (9, 25, 34). The STAT5 motif clusters together with motifs that match 9 other transcription factors (or transcription factor families), all of which exhibit a common pattern of enrichment in male-specific, GH-responsive DHS sites (Fig. 7; motif cluster A). Eight of these 10 motifs are under-represented in female-specific GH-responsive sites (cluster A1). These 10 motifs include binding sites for: CUX2, a highly female-specific, GH-regulated transcription factor (24); GFI1, a STAT-inducible transcriptional repressor (22, 60); OCT1 (POU2F1), which interacts with STAT5 in binding to the cyclin D1 promoter (31); PBX1, which interacts with OCT1 (35, 48) and may help penetrate repressive chromatin (41); and EVI1 (gene Mecom), a positive regulator of PBX1 (44) that interacts with the histone methyltransferase SUV39H1 (13). Two motifs, binding sites for MYC and MAX, were enriched in male-specific DHS sites not responsive to GH. Finally, motifs corresponding to binding sites for VDR, TCFAP2A, and TAL1 were most highly enriched at female-specific DHS sites. VDR activates the GH-responsive and female-predominant gene CYP3A4 in human hepatocytes (4, 7, 55), and a female-specific DHS site containing the VDR motif is associated with a female-specific mouse homolog, Cyp3a41a (19).
DISCUSSION

We present a set of detailed, high quality, genome-wide hypersensitivity maps comprised of more than 70,000 DHS sites, which encompass the transcriptional regulatory elements in mouse liver *in vivo*. DHS maps were generated for both male and female mouse liver, from which we were able to identify 1,284 high stringency sex-specific DHS sites, a subset of which was responsive to changes in plasma GH status, the major determinant of sex differences in liver gene expression. We demonstrate the utility of these maps for identifying binding sites for transcription factors previously shown to be essential for GH-regulated sex-specific gene expression (STAT5b and HNF4α; (5, 18, 20)), as well as binding sites for several novel factors, not previously implicated in this process.

These DHS sites encompass 1.8% of the mappable mouse genome, which substantially narrows down the sequence space in searches for gene regulatory sequences, including binding site motifs important for liver gene expression. The fine structure of DHS sites with a high density of sequence reads (Fig. S1B) suggests that it might be possible to visualize transcription factor binding directly in the form of digital footprints within DHS sites (17). Further analysis of hypersensitivity data collected at higher sequencing depth will be required to establish the feasibility of this approach in mammalian tissues. DHS sites are expected to encompass key regulatory elements, including promoters, enhancers, silencers and insulators associated with the expression of thousands of genes in their native chromatin structure under physiological conditions. The DHS maps presented here for mouse liver tissue, in combination with corresponding sets of genome-wide histone modification and transcription factor binding maps (11, 38, 42, 43, 49), can be expected to serve as a valuable resource for elucidation of transcriptional networks controlling a wide range of physiological and pathophysiological processes.
Most sex-dependent DHS sites were short and highly localized (median length ~ 500 nts), but in several cases sex-specific hypersensitivity extended over broad regions, up to ~100 kb in length (Fig. 6 and Fig. S6). The accessibility of many of these sex-dependent DHS sites and regions was altered by continuous GH infusion of male mice, which both feminizes the overall pattern of liver gene expression (19) and rendered the vast majority of the male-specific DHS sites less accessible to DNase while increasing the hypersensitivity of a substantial subset of the female-specific DHS sites and extended regions. These findings support the proposal that these GH-responsive DHS sites play a functional role in liver sexual dimorphism, and suggest that a common upstream pathway responsive to GH, such as the activation of STAT5b (26), regulates their differential chromatin accessibility in male and female liver.

We also observed a strong association between sex-specific DHS sites and sex-specific gene expression, with sex-specific genes more likely than sex-independent genes to have a nearby sex-specific DHS site. Moreover, sex-specific DHS sites were more likely than sex-independent DHS sites to have a nearby sex-specific gene. In some cases, multiple DHS sites, or extended hypersensitivity regions, discussed above, were associated with sex-specific genes. These may act in concert to increase the magnitude of differences in gene expression between male and female liver. However, in other cases, we observed groups of sex-specific DHS sites not located near sex-specific genes – one striking example is a cluster of female-specific DHS sites on the X chromosome (Fig. S6B), and another is a cluster of male-specific sites on chr13, whose nearest sex-specific genes (Cd180 and Sgtb) are weakly female-specific and located 800 kb and 500 kb, respectively, from the cluster (Fig. S6A). Indeed, for a majority of sex-specific DHS sites the closest gene was not a sex specific gene, and only a subset of sex-specific genes have a sex-specific DHS site within 100 kb of the gene. These findings suggest the importance of long-range DNA interactions for sex-specific
gene expression, as well as more complex interactions between multiple regulatory sites and multiple genes (30). While our results are consistent with regulation by distal sex-specific DHS sites, it is also possible that some sex-specific genes are regulated via sex-independent DHS sites whose cognate transcription factors are expressed or regulated in a sex-dependent manner. Other sex-specific genes may be regulated post-transcriptionally, i.e., by a mechanism that does not involve sex differences in chromatin accessibility.

Histone methylation marks, such as H3K4-me1 and H3K4-me3, are associated with active regions of chromatin, including enhancers and promoters, which are anticipated to coincide with DHS regions. Indeed, based on H3K4 methylation ChIP-seq maps for female mouse liver (38), we found a much higher fraction of H3K4-me1 marks than H3K4-me3 marks associated with DHS sites, particularly for female-specific DHS sites. As H3K4-me1 in the absence of H3K4-me3 is a characteristic of enhancers (15), we surmise that many liver DHS sites function as enhancers, some of which may exhibit sex-specific activities. This is supported by our in vivo reporter gene assays, where 5 out of 6 DHS sites investigated demonstrated intrinsic enhancer activity when delivered to mouse liver by hydrodynamic injection. Moreover, 3 of the 6 enhancer sequences showed sex differences in activity that match the sex specificity of the DHS site and their associated genes. The sex differences in in vivo enhancer activity seen here, were, however, considerably smaller than the sex differences in expression of the genes themselves, suggesting that multiple DHS sites may be required to confer a high degree of sex-specificity to gene expression. Indeed, multiple sex-dependent DHS sites are associated with the three genes whose enhancers showed sex differences in activity (Acot4, Cyp7b1, Cyp2d9). In the case of the Cux2 intron 2 DHS site tested, no sex-difference in enhancer activity was observed, indicating that the cloned fragment does not recapitulate the strong sex difference in DNA accessibility seen in intact liver chromatin.
(female/male DHS ratio = 7.3). Nevertheless, given the very strong enhancer activity of this genomic region (Fig. 5), coupled with 7-fold lower accessibility in male liver, this DHS site could make a substantial contribution to the strong (~100-fold) female specificity that characterizes Cux2 gene expression (24). These findings suggest that some sex-dependent DHS sites exhibit intrinsic sex differences in enhancer activity, e.g., due to the binding of transcription factors that are expressed or activated in a sex- and plasma GH pattern-dependent manner (e.g., STAT5b), while other sex-dependent DHS sites (e.g., the Cux2 intron 2 enhancer) impart strong sex differences to gene transcription by virtue of the large sex differences in their accessibility in intact liver chromatin per se, even though they might not directly bind sex-specific transcription factors. Further studies will be required to identify the factors and establish the underlying mechanisms that initiate and maintain these sex differences in chromatin structure.

Motif analysis identified 13 transcription factor binding motifs that are enriched in one or more subsets of male-specific DHS sites compared to sex-independent DHS sites (Fig. 7 and Table S10). Three other motifs were enriched in female-specific DHS sites, and one of these, the motif for TCFAP2A, was depleted in subsets of male-specific DHS sites. The male DHS site-enriched motifs include motifs that bind liver-expressed transcription factor families from the FOX and nuclear receptor families (e.g., HNF4α), as well as the binding site for STAT5b, which exhibits important sex differences in responsiveness to GH stimulation in vivo (52). Another male DHS-enriched motif, CDP, matches the binding site for CUX2 (12), a transcriptional repressor expressed at a 100-fold higher level in female compared to male liver (24), suggesting that CUX2 may enforce male specificity by binding to male-specific DHS sites in female liver, thereby suppressing the residual activity of enhancers that are partially accessible in females. A cluster of 8 male DHS site-enriched motifs (sub-cluster A1; Fig. 7), which includes motifs for STAT5b and CUX2, was depleted in a...
subset of female-specific DHS sites responsive to GH. Given the high frequency of these 8 male DHS site-enriched motifs, it is not surprising that many male DHS sites contain matches for 6 or more of the 8 motifs (Fig. S7). Further work will be needed to determine whether or not particular combinations of these 8 motifs have distinct functions, and to identify the specific factors that actually bind to their cognate sequences at DHS sites in male and female liver.

Our finding that male-specific, GH-responsive DHS sites are enriched for both STAT5b-like and HNF4α-like (nuclear receptor) motifs is consistent with our earlier observation that these factors are both essential for sex-specific gene expression in mouse liver. STAT5b is one of the major direct effectors of GH signaling in liver, and its deletion down-regulates ~90% of male-specific genes in male mouse liver (5, 18). Similarly, knockout of the liver-enriched nuclear receptor HNF4α abolishes liver sex differences (20). By carrying out the motif analysis separately for DHS sites that are near sex-specific genes, near sex-independent genes, and distant from genes, we showed that the HNF4α-like motif is most highly enriched at DHS sites within 10 kb of sex-specific genes, while the STAT5-like motif is highly enriched at distal sites, as well as at sites proximal to sex-independent genes. This latter finding is consistent with the occurrence of functional STAT5 binding sites at large distances from target genes (9).

In conclusion, the present investigation of sex differences in chromatin accessibility has identified condition-specific transcriptional regulatory sites in mouse liver on a genome-wide scale. These differences are manifested as sex-specific DHS sites, which in some cases encompass extended chromatin regions. A subset of these sex-specific genomic sites and regions is associated with genes expressed in a sex-dependent manner, strongly suggesting they play a functional role in liver sexual dimorphism; however, a majority of sex-specific DHS sites were distal to sex-specific genes, making
it more difficult to establish their significance. Transcription factor binding motifs identified as
enriched in these sites serve as candidates for further study of the molecular mechanisms that govern
sex-specific liver gene transcription. Further study will be required to determine how sex
differences in chromatin accessibility are established and maintained in response to sex-differences
in plasma GH patterns, which are programmed by early androgen exposure and first emerge at
puberty.

Acknowledgements - This work was supported in part by NIH grant DK33765 (to D.J.W.). A.S.
received Training Core support from the Superfund Research Program at Boston University (NIH grant
5 P42 ES07381). The authors thank Minita Holloway of this laboratory for contributions to initial
optimization of the DHS protocol and Dr. X. Shirley Liu (Dana-Farber Cancer Institute) for initial
discussions about experimental design.

Abbreviations – DHS, DNase hypersensitivity; GH, growth hormone; TSS, transcription start site;
ChIP, chromatin immunoprecipitation; seq, high throughput sequence analysis.
REFERENCES


**FIGURE LEGENDS**

Figure 1: DHS sites associated with the *Alb* gene (A), *Cyp2d9* (B), a male-specific gene, and *Cux2* (C), a female-specific gene. A, DHS sites appear as sharp, narrow peaks and are marked by distance upstream of the *Alb* gene. The four sites marked in red are the same as those discovered earlier using conventional methods (28). Green arrows at bottom mark regions of high species conservation that coincide with DHS sites. B, Male-specific DHS sites in the region of *Cyp2d9* are marked as horizontal bars in the merged DHS sites track, with dark blue color used for high stringency male DHS sites and lighter shades of blue for standard and low stringency male DHS sites, numbered as in Table S5A. Sex-independent DHS sites are shown in gray. Continuous GH treatment of male mice suppresses each of the four high stringency male DHS sites down to female levels (bottom track). C, Intron 2 of *Cux2*, a highly female-specific gene (female/male expression ratio ~100; (24)), showing female-specific DHS sites (marked as pink horizontal bars), four of which are markedly induced in continuous GH-treated male liver, as indicated by red asterisks. A–C, Individual sequence reads (35 nt) are represented as a single nt wide bar graphed at the chromosomal location of the DNase cut site; this presentation is clearest in panel B, where individual sequence reads can be seen at the resolution presented. Green and red indicate sequence reads from DNase-digested female liver, and blue and yellow indicate sequence reads from DNase-digested male liver, or continuous GH-treated male liver, as marked. Green and blue, (+) strand reads; red and yellow, (−) strand reads.

Figure 2: Relationship between level of gene expression and presence of a 5' DHS site. Shown is the fraction of sex-independent genes (A) or the fraction of sex-specific genes (B) that have at least one 5' DHS site within 200 bp of the TSS, graphed as a function of log2 liver expression level (microarray signal intensity). Sex-specific genes are those that show >2-fold sex differences at...
p<0.005 as determined by microarray analysis (50), using expression intensities for the sex showing higher expression. The total number of genes at each range of expression intensity is shown above each bar. The proportions of sex-independent and sex-specific genes that have a 5'-DHS site at increasing expression levels are significantly different (Wilcoxon signed rank test test, p = 0.0006).

Figure 3: Proximity of sex-specific and sex-independent DHS sites to genes expressed in mouse liver. (A) Cumulative distances from the gene body of sex-specific and sex-independent genes to the nearest sex-specific and sex-independent DHS sites. (B) Cumulative distances from sex-specific and sex-independent DHS sites to the nearest sex-specific or sex-independent gene. X-axes are on different scales in the left and right panels of A and B. For A and B, gene expression in liver is defined as a microarray signal intensity ≥ 500 in either male or female liver, and sex-specific genes are those that met the combined criteria of >2-fold sex difference at p<0.005 in a microarray comparison of male and female mouse liver (50). (C) Sex-specificity of the genes closest to the indicated categories of sex-specific and sex-independent DHS sites. Shown are the percentages of DHS sites whose closest gene shows male-specific, female-specific, or sex-independent expression, as indicated, using the above criteria for sex specificity, but independent of microarray signal intensity.

Figure 4: Histone modifications at sex-independent and sex-specific DHS sites in female mouse liver. (A) Percentages of sex-independent and sex-specific DHS sites within 150 bp of a H3K4-me1 and/or H3K4-me3 mark in female mouse liver. These data are shown for all DHS sites, and for the subsets comprised of sites that are proximal (within 200 bp, or between 0.2 and 5 kb, of an active sex-specific or sex-independent TSS) to liver-expressed genes (microarray signal intensity ≥ 500, as in Fig. 3), and those that are distal to genes (>5 kb from the TSS), regardless of liver expression.
intensity level. Proportions of DHS sites associated with the different patterns of H3K4 methylation marks are statistically significant (between all sex-specific and all sex-independent DHS sites: $p<10^{-14}$; between sites ≤ 200 bp, 0.2-5 kb, and > 5 kb from the TSS: $p<10^{-21}$ for sex-independent sites, and $p<10^{-5}$ for sex-specific sites; $\chi^2$ test). (B-C) Distributions of H3K4me1, H3K4me3, and FoxA2 ChIP-seq reads relative to sex-independent (B) and female-specific (C) DHS sites.

Figure 5: Enhancer activity of six sex-specific DHS sites assayed in mouse liver. Four high-stringency female DHS sites (Cux2, Acot4, Cyp3a16 and Cyp2c39) and two male DHS sites (Cyp7b1 and Cyp2d9), named after the closest gene to each site, were assayed for reporter gene activity after in vivo transfection of mouse liver by hydrodynamic injection. All 4 female-specific DHS sites are associated with H3K4me1 and not H3K4me3 marks (Table S8). Data shown represent normalized luciferase activity for at least 3 individual mice (mean $\pm$ SE), each of which is based on the average reading from 3 separate pieces of liver. Relative luciferase activity of the parental vector AlbPmo in male and female liver (mean value) was set to 1. *, significant difference in activity between male and female liver transfection at $p<0.05$.

Figure 6: Broad regions of DNase hypersensitivity. Tracks in each panel show (in order, from top) RefSeq genes, DHS peaks identified using PeakSeq, and DHS regions identified using SICER, with the length of the SICER-detected region marked in red. Grey, blue and pink indicate sex-independent, male-specific, and female-specific SICER regions, respectively, with lighter blue or pink indicating < 2-fold sex-specificity. Panel A depicts a 102 kb sex-independent region overlapping with six sex-independent genes. Panels B and C show 56 to 87 kb male-specific DHS regions in the vicinity of male-specific genes Ttc39c and Aox3, respectively; and panel D shows several female-specific DHS regions at the female-specific genes Cyp2b9 and Cyp2b13. While the
broad DHS regions in panels A and B contain multiple individual DHS peaks, as identified by
PeakSeq, the broad DHS regions shown in panels C and D were not identified by PeakSeq, which is
optimized for identification of short, well defined discrete peaks.

Figure 7: Heat map of transcription factor binding motifs enriched in sex-specific DHS sites
compared to sex-independent DHS sites. Motifs are clustered by fold-enrichment in male-specific
(M) and female-specific (F) DHS sites that respond (GH) or do not respond to GH (noGH) at $p<0.01$
and fold-difference $>2$, and are either 0-10 kb or 10-50 kb from the TSS of either sex-specific ('Sex')
or sex-independent ('Ind') genes, or $>50$ kb from any gene (distal), as marked. Combined data for 0
to 50 kb is presented for the M-noGH and F-GH sets, which both contained <15 DHS sites in the 10-
50 kb subset. Red indicates over-representation and green indicates under-representation $>1.45$,
with darker color indicating higher enrichment. The full data set is shown in Table S10A. Where
multiple factors are listed, they are in order of highest to lowest match to the motif matrix. Major
clusters of motifs (A-D) are identified on the left. Cluster A (10 motifs) is enriched in male-specific
DHS sites; 8 of these motifs (sub-cluster $A_1$) are enriched in several male-specific GH-responsive
(M-GH) sets and depleted in the F-GH, sex-independent (0-10 kb) set. Motifs in cluster B are most
highly enriched in F-GH sets, and are most highly depleted in various male DHS sites and in a F-
noGH set. Motif 94 (TAL1) is most highly enriched in the same F-GH set as TFAP2A. Motif 70 is
enriched in several M-GH sets, but in contrast to the motifs in cluster A, it is most highly enriched in
the DHS set that is within 10 kb of sex-specific genes. Two very similar motifs for MYC/MAX are
enriched in male-specific non-GH responsive sites.
Table 1: DHS sites in male and female mouse liver. Male vs. control and female vs. control DHS sites determined by PeakSeq analysis were merged to generate a single list of ‘all DHS sites’ in liver. Male-specific and female-specific DHS sites were determined by comparing male vs. female, and female vs. male DHS sites, respectively. Two biological replicates each from male and female mice were combined for each comparison. Stringency differences are as indicated and as described in Materials and Methods, and full lists of sites are presented in Table S5.

<table>
<thead>
<tr>
<th></th>
<th>High stringency*</th>
<th>Standard stringency**</th>
<th>Low stringency***</th>
</tr>
</thead>
<tbody>
<tr>
<td>All DHS sites</td>
<td>48,762</td>
<td>71,264</td>
<td>110,785</td>
</tr>
<tr>
<td>Male-specific DHS sites</td>
<td>850</td>
<td>2,315</td>
<td>2,800</td>
</tr>
<tr>
<td>Female-specific DHS sites</td>
<td>434</td>
<td>1,064</td>
<td>1,382</td>
</tr>
</tbody>
</table>

*Sites that meet $p<0.01$, fold-difference >2, and confirmed separately in both sets of biological replicates.

**Sites that meet $p<0.01$, fold-difference >2, and confirmed in at least 2 of the 3 combined samples (Male, Female, and GH-treated male) for sex-independent DHS sites, or at least one individual set of biological replicates for sex-specific DHS sites.

*** All sites that meet $p<0.01$ and fold-difference >2 in the either the combined male replicate or the combined female replicate sequencing data sets.
Table 2: Overlap between ChIP-chip/ChIP-seq transcription factor binding sites and DHS sites in liver. 67-93% of experimentally determined transcription factor binding sites in mouse liver overlap standard stringency liver DHS sites. The extent to which these sites overlap with DHS sites was very similar when only the subset of sites that contain the respective binding motif for the factor used for ChIP was considered (last column). Data for FOXA2 (38), ERα/ESR1 (11), FXR/NR1H4 (49), SREBP-1 (43), CEBP4A and HNF4A (42) are based on ChIP-chip or ChIP-seq in various mouse strains. Binding motifs were obtained from TRANSFAC and also identified de novo in each set of binding sites using the motif discovery module of CisGenome.

<table>
<thead>
<tr>
<th>Transcription factor</th>
<th>All binding regions</th>
<th>Binding regions containing specific motif</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># sites</td>
<td>% overlap with DHS</td>
</tr>
<tr>
<td>FOXA2</td>
<td>10,958</td>
<td>68%</td>
</tr>
<tr>
<td>ERα/ESR1</td>
<td>6,272</td>
<td>86%</td>
</tr>
<tr>
<td>FXR/NR1H4</td>
<td>7,794</td>
<td>93%</td>
</tr>
<tr>
<td>SREBP-1</td>
<td>426</td>
<td>91%</td>
</tr>
<tr>
<td>CEBP4A</td>
<td>29,188</td>
<td>67%</td>
</tr>
<tr>
<td>HNF4A</td>
<td>20,355</td>
<td>90%</td>
</tr>
</tbody>
</table>
Table 3: Effect of continuous GH treatment on DHS sites in male liver. Numbers of high-stringency sex-specific and standard stringency sex-independent DHS sites that are induced or suppressed in continuous GH-treated male liver ($p<0.01$ and $>2$-fold-difference between GH-treated males and untreated males). Values in parentheses are percentages of numbers of DHS sites shown in column 1. The GH-responses for lower-stringency sex-specific DHS sites, and GH-responses at lower stringency ($p<0.05$), are summarized in Table S5H.

<table>
<thead>
<tr>
<th>Sex-specificity of DHS site</th>
<th>Effect of continuous GH treatment in male</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># DHS sites Induced</td>
<td>Suppressed</td>
</tr>
<tr>
<td>Male-specific sites (n = 850)</td>
<td>0</td>
<td>693 (82%)</td>
</tr>
<tr>
<td>Female-specific sites (n = 434)</td>
<td>114 (26%)</td>
<td>1 (0.2%)</td>
</tr>
<tr>
<td>Sex-independent sites (n = 68,683)*</td>
<td>102 (0.1%)</td>
<td>1,848 (3%)</td>
</tr>
</tbody>
</table>

*This represents the 71,264 standard stringency DHS sites after removal of sites that show sex-specific at low stringency.
Table 4: Broad DHS regions and their response to continuous GH treatment in male liver.

Numbers of sex-specific and sex-independent hypersensitivity regions that are >10 kb in length, and numbers of each that are induced or suppressed in continuous GH-treated male mouse liver, as determined by SICER analysis. All DHS regions are significant with FDR < $10^{-3}$, and are ≥ 2-fold enriched for sequence reads in the relevant comparison (Male vs control or Female vs control for sex-independent regions; Male vs control and Male vs Female for male-specific regions; Female vs control and Female vs Male for female-specific regions; and Male vs Male+GH\textsubscript{cont} or Male+GH\textsubscript{cont} vs Male for GH response). Values in parentheses are percentages of numbers of regions shown in column 1. Complete lists of the SICER-determined DHS regions are in Table S9.

<table>
<thead>
<tr>
<th>Sex-specificity</th>
<th># &gt;10 kb regions</th>
<th>Effect of continuous GH treatment in male</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td># Regions Induced</td>
</tr>
<tr>
<td>Male-specific</td>
<td>34</td>
<td>0</td>
</tr>
<tr>
<td>Female-specific</td>
<td>24</td>
<td>12 (50%)</td>
</tr>
<tr>
<td>Sex-independent*</td>
<td>3,913</td>
<td>0</td>
</tr>
</tbody>
</table>

*DHS regions significant in Male vs control or Female vs control merged into a single list, excluding those regions that are sex-specific.