Hydroxysteroid 17-beta Dehydrogenase 13 (HSD17B13) Variants And Uses Thereof

Reference To A Sequence Listing

This application includes a Sequence Listing submitted electronically as a text file named 18923800802SEQ, created on January 18, 2018, with a size of 147 kilobytes. The Sequence Listing is incorporated herein by reference.

Field

The disclosure relates generally to the field of genetics. More particularly, the disclosure relates to gene alterations and polypeptide variants in hydroxysteroid 17-beta dehydrogenase 13 (HSD17B13) that associate with, for example, liver disease.

Background

Various references, including patents, patent applications, accession numbers, technical articles, and scholarly articles are cited throughout the specification. Each reference is incorporated by reference herein, in its entirety and for all purposes.

Chronic liver disease and cirrhosis are leading causes of morbidity and mortality in the United States, accounting for 38,170 deaths (1.5% of total deaths) in 2014 (Kochanek et al., Natl. Vital Stat. Rep., 2016, 65, 1-122). The most common etiologies of cirrhosis in the U.S. are alcoholic liver disease, chronic hepatitis C, and nonalcoholic fatty liver disease (NAFLD), together accounting for about 80% of patients awaiting liver transplant between 2004 and 2013 (Wong et al., Gastroenterology, 2015, 148, 547-555). The estimated prevalence of NAFLD in the U.S. is between 19 and 46 percent (Browning et al., Hepatology, 2004, 40, 1387-1395; Lazo et al., Am. J. Epidemiol., 2013, 178, 38-45; and Williams et al., Gastroenterology, 2011, 140, 124-131) and is rising over time (Younossi et al., Clin. Gastroenterol. Hepatol., 2011, 9, 524-530 e1; quiz e60, 2011), likely in conjunction with increased rates of obesity, its primary risk factor (Cohen et al., Science, 2011, 332, 1519-1523). While significant advances have been made in the treatment of hepatitis C (Morgan et al., Ann. Intern. Med., 2013, 158, 329-337; and van der Meer et al., J. Amer. Med. Assoc., 2012, 308, 2584-2593), there are currently no evidence-based treatments for alcoholic or nonalcoholic liver disease and cirrhosis.

Previous genome wide association studies (GWAS) have identified a limited number of genes and variants associated with chronic liver disease. The most robustly validated genetic association to date is to a common missense variant in the patatin-like phospholipase domain

containing 3 gene (*PNPLA3* p.Ile148Met, rs738409), initially found to be associated with increased risk of nonalcoholic fatty liver disease (NAFLD) (Romeo et al., Nat. Genet., 2008, 40, 1461-1465; and Speliotes et al., PLoS Genet., 2011, 7:e1001324), and subsequently found to be associated with disease severity (Rotman et al., Hepatology, 2010, 52, 894-903; and Sookoian et al., J. Lipid Res., 2009, 50, 2111-2116) and progression (Trepo et al., J. Hepatol., 2016, doi:10.1016/j.jhep.2016.03.011). Variation in the transmembrane 6 superfamily member 2 (*TM6SF2*) gene has also been shown to confer increased risk for NAFLD (Kozlitina et al., Nat. Genet., 2014, 46, 352-356; Liu et al., Nat. Commun., 2014, 5, 4309; and Sookoian et al., Hepatology, 2015, 61, 515-525). The normal functions of these two proteins are not well understood, though both have been proposed to be involved in hepatocyte lipid metabolism. How variants in *PNPLA3* and *TM6SF2* contribute to increased risk of liver disease has yet to be elucidated. GWAS have also identified several genetic factors to be associated with serum alanine aminotransferase (ALT) and aspartate aminotransferase (AST) (Chambers et al., Nat. Genet., 2011, 43,1131-1138; and Yuan et al., Am. J. Hum. Genet., 2008, 83, 520-528), quantitative markers of hepatocyte injury and liver fat accumulation that are frequently measured clinically. To date, there are no described protective genetic variants for chronic liver disease. The discovery of protective genetic variants in other settings, such as loss-of-function variants in PCSK9 that reduce the risk of cardiovascular disease, has been the catalyst for development of new classes of therapeutics.

Knowledge of genetic factors underlying the development and progression of chronic liver disease could improve risk stratification and provide the foundation for novel therapeutic strategies. A better understanding of underlying genetic factors is needed to improve risk stratification and generate novel therapies for liver disease.

The present disclosure provides novel HSD17B13 variants that will aid in understanding the biology of HSD17B13, and will facilitate the diagnosis and treatment of subjects with liver disease.

Summary

The present disclosure provides nucleic acid molecules, polypeptides, probes, primers, compositions, and methods related to the variant *HSD17B13* rs72613567 gene, variant *HSD17B13* transcripts, and variant HSD17B13 protein isoforms.

The present disclosure also provides nucleic acid molecules encoding variant HSD17B13 protein isoforms. In some embodiments, the nucleic acid molecule encodes the variant HSD17B13 protein Isoform C, Isoform D, Isoform F, Isoform G, or Isoform H. In some

embodiments, the nucleic acid molecule encodes the variant HSD17B13 protein Isoform D.

The present disclosure also provides nucleic acid molecules comprising or consisting of at least 15 contiguous nucleotides of an HSD17B13 gene, wherein the contiguous nucleotides are at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to a corresponding sequence in SEQ ID NO:2, and having a thymine at a position corresponding to position 12666 of SEQ ID NO:2.

Disclosed herein is a nucleic acid molecule comprising or consisting of at least 15 contiguous nucleotides of an HSD17B13 gene, wherein the contiguous nucleotides are at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to a corresponding sequence in SEQ ID NO:2, and having a thymine at a position corresponding to position 12666 of SEQ ID NO:2 when used for detecting a variant HSD17B13 gene in a human subject, wherein the nucleic acid molecule is linked to a heterologous nucleic acid or comprises a heterologous label.

The present disclosure also provides nucleic acid molecules comprising or consisting of a nucleotide sequence that encodes a polypeptide having an amino acid sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to the amino acid sequence of HSD17B13 Isoform D (SEQ ID NO:42). In some embodiments, the nucleic acid molecules comprise or consist of a nucleotide sequence that encodes a polypeptide comprising the amino acid sequence of HSD17B13 Isoform D (SEQ ID NO:42). In some embodiments, the nucleic acid molecules comprise or consist of a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a nucleotide sequence of a HSD17B13 Transcript D (SEQ ID NO:6, 15, 24, or 33). In some embodiments, the nucleic acid molecule is RNA and comprises or consists of SEQ ID NO:6, or a cDNA thereof comprising or consisting of SEQ ID NO:24, or wherein the nucleic acid molecule is mRNA and comprises or consists of SEQ ID NO:15, or a cDNA thereof comprising or consisting of SEQ ID NO:33.

Disclosed herein is a nucleic acid molecule comprising or consisting of a nucleotide sequence that encodes a polypeptide having an amino acid sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to the amino acid sequence of HSD17B13 Isoform D (SEQ ID NO:42) when used for detecting a variant HSD17B13 gene in a human subject, wherein the nucleic acid molecule is linked to a heterologous nucleic acid or comprises a heterologous label.

The present disclosure also provides nucleic acid molecules, such as probes and

primers, such as alteration-specific probes or alteration-specific primers, that hybridize to or near the alteration in the variant *HSD17B13* rs72613567 gene or an *HSD17B13* transcript.

The present disclosure also provides nucleic acid molecules comprising or consisting of from about 5 nucleotides up to about 50 nucleotides that specifically hybridize to a variant HSD17B13 gene at a region that includes a position corresponding to position 12666 in SEQ ID NO:2, or the complement thereof, and wherein the nucleic acid molecules specifically hybridize to a variant HSD17B13 gene having a thymine at a position corresponding to position 12666 of SEQ ID NO:2, or to the complement thereof.

Disclosed herein is a nucleic acid molecule comprising or consisting of from about 5 nucleotides up to about 50 nucleotides that specifically hybridizes to an HSD17B13 gene at a region that includes a position corresponding to position 12666 in SEQ ID NO:2, or the complement thereof, and wherein the nucleic acid molecule specifically hybridizes to an HSD17B13 gene having a thymine at a position corresponding to position 12666 of SEQ ID NO:2, or to the complement thereof when used for detecting a variant HSD17B13 gene in a human subject, wherein the nucleic acid molecule is linked to a heterologous nucleic acid or comprises a heterologous label.

The present disclosure also provides nucleic acid molecules comprising or consisting of from about 5 nucleotides up to about 50 nucleotides that specifically hybridize to variant HSD17B13 Transcript D, wherein the nucleic acid molecules specifically hybridize to: i) a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a nucleotide sequence of a SEQ ID NO:6, 15, 24, or 33, or ii) to the complement of the nucleotide sequence of i).

Disclosed herein is a nucleic acid molecule when used for detecting a variant HSD17B13 gene in a human subject, wherein the nucleic acid molecule comprises or consists of from about 5 nucleotides up to about 50 nucleotides that: a) specifically hybridizes to variant HSD17B13 Transcript D, and wherein the nucleic acid molecule specifically hybridizes to: i) a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a nucleotide sequence of a SEQ ID NO:6, 15, 24, or 33, or ii) to the complement of the nucleotide sequence of i) and wherein the nucleic acid molecule is linked to a heterologous nucleic acid or comprises a heterologous label; or b) comprises or consists of: i) the nucleic acid molecule of a); ii) a nucleic acid molecule which specifically hybridizes to exon 2 of Transcript D; and/or iii) a nucleic acid molecule which specifically hybridizes to the region which bridges exons 3 and 4 of Transcript D, wherein the nucleic acid molecule is linked to a heterologous nucleic acid or comprises a heterologous label.

The present disclosure also provides nucleic acid molecules comprising or consisting of from about 5 nucleotides up to about 50 nucleotides comprising: i) a nucleic acid molecule which specifically hybridizes to a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a nucleotide sequence of a SEQ ID NO:6, 15, 24, or 33, or to the complement thereof; ii) a nucleic acid molecule which specifically hybridizes to exon 2 of Transcript D; and/or iii) a nucleic acid molecule which specifically hybridizes to the region which bridges exons 3 and 4 of Transcript D. In some embodiments, the nucleic acid molecule specifically hybridizes to an RNA molecule comprising or consisting of SEQ ID NO:6, or a cDNA thereof comprising or consisting of SEQ ID NO:24, or wherein the nucleic acid molecule specifically hybridizes to an mRNA comprising or consisting of SEQ ID NO:15, or a cDNA thereof comprising or consisting of SEQ ID NO:33, or the complement thereof. In some embodiments, the nucleic acid molecule is linked to a heterologous nucleic acid or comprises a heterologous label.

The present disclosure also provides vectors comprising any of these nucleic acid molecules.

The present disclosure also provides cells comprising any of these nucleic acid molecules.

The present disclosure also provides cells comprising any of these vectors.

The present disclosure also provides compositions comprising any of these nucleic acid molecules.

The present disclosure also provides compositions comprising any of these vectors.

The present disclosure also provides compositions comprising any of these cells.

The present disclosure also provides uses of any of these nucleic acid molecules for detecting a variant HSD17B13 gene or transcript, for determining a human subject's susceptibility or risk of developing a liver disease, or for determining a human subject's risk for progression to more clinically advanced stages of fatty liver disease.

The present disclosure also provides polypeptides corresponding with different variant HSD17B13 protein isoforms.

The present disclosure also provides polypeptides comprising or consisting of an amino acid sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to the amino acid sequence of HSD17B13 Isoform D (SEQ ID NO:42). In some embodiments, the polypeptide comprises or

consists of the amino acid sequence of SEQ ID NO:42. In some embodiments, the polypeptide is linked to a heterologous molecule.

The present disclosure also provides compositions comprising any of the polypeptides.

The present disclosure also provides methods of detecting a variant *HSD17B13* rs72613567 gene, variant *HSD17B13* transcripts (such as Transcript D), and variant HSD17B13 Isoforms (such as Isoform D).

The present disclosure also provides methods of detecting a variant HSD17B13 gene in a human subject, comprising performing an assay on a biological sample obtained from the human subject, wherein the assay determines whether a thymine is inserted between positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene, or whether a thymine is present at a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene, wherein the presence of the thymine is indicative for a variant HSD17B13 gene. In some embodiments, the assay comprises sequencing a portion of the HSD17B13 gene including positions corresponding to positions 12665 and 12666 of SEQ ID NO:1, or including a position corresponding to position 12666 of SEQ ID NO:2. In some embodiments, the assay comprises or consists of: i) contacting the biological sample with a primer hybridizing to a region of the HSD17B13 gene that is within 50 nucleotides of a position corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene, or within 50 nucleotides of a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene; ii) extending the primer at least through the position corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene, or corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene; and iii) determining whether a thymine is inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene, or whether a thymine is present at a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene, in an extension product of the primer. In some embodiments, the method further comprises determining whether the human subject is homozygous for the variant HSD17B13 gene.

Disclosed herein are methods of detecting a variant HSD17B13 gene in a human subject, comprising or consisting of performing an assay on a biological sample obtained from the human subject, wherein the assay determines whether a thymine is inserted between positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene, or whether a thymine is present at a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene, wherein the presence of said thymine is

indicative for a variant HSD17B13 gene, wherein the presence of said thymine is indicative for a variant HSD17B13 gene encoding HSD17B13 Isoforms D-F, F', G, or H.

The present disclosure also provides methods of detecting the presence of an HSD17B13 Transcript D in a human subject, comprising performing an assay on a biological sample obtained from the subject, wherein the assay determines the presence of an HSD17B13 Transcript D in the biological sample. In some embodiments, the assay comprises or consists of contacting the biological sample with one or more primers or probes that specifically hybridize to a nucleic acid sequence of HSD17B13 Transcript D, or to the complement thereof, and determining whether hybridization has occurred. In some embodiments, the method further comprises or consists of specifically detecting Transcript D by using a nucleic acid molecule comprising or consisting of from about 5 nucleotides up to about 50 nucleotides comprising or consisting of: i) a nucleic acid molecule which specifically hybridizes to a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a nucleotide sequence of a SEQ ID NO:6, 15, 24, or 33, or to the complement thereof; ii) a nucleic acid molecule which specifically hybridizes to exon 2 of Transcript D; and/or iii) a nucleic acid molecule which specifically hybridizes to the region which bridges exons 3 and 4 of Transcript D. In some embodiments, the HSD17B13 Transcript D comprises or consists of a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to SEQ ID NO:6, 15, 24, or 33. In some embodiments, the one or more primers or probes specifically hybridize to SEQ ID NO:6, SEQ ID NO:15, SEQ ID NO:24, and/or SEQ ID NO:33. In some embodiments, the assay further comprises using a primer or probe which specifically hybridizes to one or more of Transcripts A, B, C, E, F, F', G and/or H but not to Transcript D, and determining that hybridization has not occurred. In some embodiments, the assay comprises reverse transcription polymerase chain reaction (RT-PCR). In some embodiments, the assay comprises sequencing.

The present disclosure also provides methods of detecting the presence of HSD17B13 Isoform D in a human subject, comprising performing an assay on a biological sample obtained from the human subject, wherein the assay determines the presence of HSD17B13 Isoform D in the biological sample. In some embodiments, the HSD17B13 Isoform D comprises or consists of an amino acid sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to SEQ ID NO:42. In some embodiments, the assay comprises sequencing.

The present disclosure also provides methods of determining a subject's susceptibility

to developing a liver disease and/or methods of diagnosing a subject with a liver disease by determining whether the variant *HSD17B13* rs72613567 gene, variant *HSD17B13* transcripts (such as Transcript D), and variant HSD17B13 Isoforms (such as Isoform D) is present in a biological sample from the subject.

Disclosed herein is are methods of determining a human subject's susceptibility or risk of developing a liver disease, comprising or consisting of: a) performing an assay on a biological sample obtained from the human subject, wherein the assay determines whether a thymine is inserted between positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1, or whether a thymine is present at a position of the variant HSD17B13 gene corresponding to position 12666 of SEQ ID NO:2; and b) classifying the human subject as being at decreased risk for developing the liver disease if a thymine is inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene or if a thymine is present at a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene, or classifying the human subject as being at increased risk for developing the liver disease if a thymine is not inserted between the positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1 or if a thymine is not present at a position of the variant HSD17B13 gene corresponding to position 12666 of SEQ ID NO:2. In some embodiments, the liver disease is a chronic liver disease. In some embodiments, the liver disease is selected from the group consisting of fatty liver disease, nonalcoholic fatty liver disease (NAFLD), alcoholic liver fatty liver disease, cirrhosis, viral hepatitis, hepatocellular carcinoma, simple steatosis, steatohepatitis, fibrosis, and non-alcoholic steatohepatitis (NASH). In some embodiments, the assay comprises or consists of: i) contacting the biological sample with a primer hybridizing to a region of the HSD17B13 gene that is within 50 nucleotides of positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1, or corresponding to position 12666 of SEQ ID NO:2; ii) extending the primer at least through the positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1, or corresponding to position 12666 of SEQ ID NO:2; and iii) determining whether a thymine is inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene, or whether a thymine is present at a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene, in an extension product of the primer. In some embodiments, the assay comprises or consists of contacting the biological sample with a primer or probe that specifically hybridizes to the variant HSD17B13 gene having a thymine at a position corresponding to position 12666 of SEQ ID NO:2, and does not hybridize to the corresponding wild type HSD17B13 gene under stringent conditions, and determining

whether hybridization has occurred. In some embodiments, the variant HSD17B13 gene is detected by sequencing. In some embodiments, the method further comprises determining whether the human subject is homozygous for the variant HSD17B13 gene.

The present disclosure also provides methods of determining a human subject's susceptibility or risk for developing a liver disease, comprising or consisting of: a) performing an assay on a biological sample obtained from the human subject, wherein the assay determines the presence of an HSD17B13 Transcript D in the biological sample; and b) classifying the human subject as being at decreased risk for developing the liver disease if an HSD17B13 Transcript D is present in the biological sample, or classifying the human subject as being at increased risk for developing the liver disease if an HSD17B13 Transcript D is not present in the biological sample. In some embodiments, the HSD17B13 Transcript D comprises or consists of a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to SEQ ID NO:6, 15, 24, or 33. In some embodiments, the HSD17B13 Transcript D is RNA and comprises or consists of SEQ ID NO:6, or a cDNA thereof comprising or consisting of SEQ ID NO:24, or wherein the HSD17B13 Transcript D is mRNA and comprises or consists of SEQ ID NO:15, or a cDNA thereof comprising or consisting of SEQ ID NO:33. In some embodiments, the assay determines the expression level of HSD17B13 Transcript D in the biological sample, wherein an increased expression level of HSD17B13 Transcript D compared to a control sample from a control human subject homozygous for a wild type HSD17B13 allele indicates a decreased risk for developing the liver disease, and wherein the same or a decreased expression level of HSD17B13 Transcript D compared to the control sample indicates an increased risk for developing the liver disease. In some embodiments, the liver disease is a chronic liver disease. In some embodiments, the liver disease is selected from the group consisting of fatty liver disease, nonalcoholic fatty liver disease (NAFLD), alcoholic liver fatty liver disease, cirrhosis, viral hepatitis, hepatocellular carcinoma, simple steatosis, steatohepatitis, fibrosis, and non-alcoholic steatohepatitis (NASH). In some embodiments, the assay comprises or consists of contacting the biological sample with one or more primers or probes that specifically hybridize to a nucleic acid sequence of an HSD17B13 Transcript D, or to the complement thereof, and determining whether hybridization has occurred. In some embodiments, the method further comprises specifically detecting Transcript D by using a nucleic acid molecule comprising from about 5 nucleotides up to about 50 nucleotides comprising: i) a nucleic acid molecule which specifically hybridizes to a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a nucleotide sequence of

a SEQ ID NO:6, 15, 24, or 33, or to the complement thereof; ii) a nucleic acid molecule which specifically hybridizes to exon 2 of Transcript D; and/or iii) a nucleic acid molecule which specifically hybridizes to the region which bridges exons 3 and 4 of Transcript D. In some embodiments, the HSD17B13 Transcript D comprises or consists of a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to SEQ ID NO:6, 15, 24, or 33. In some embodiments, the one or more primers or probes specifically hybridize to SEQ ID NO:6, SEQ ID NO:15, SEQ ID NO:24, and/or SEQ ID NO:33. In some embodiments, the assay further comprises using a primer or probe which specifically hybridizes to one or more of Transcripts A, B, C, E, F, F', G and/or H but not to Transcript D, and determining that hybridization has not occurred. In some embodiments, the assay comprises reverse transcription polymerase chain reaction (RT-PCR) or quantitative RT-PCR (qRT-PCR). In some embodiments, the assay comprises sequencing.

Disclosed herein are methods of determining a human subject's susceptibility or risk for developing a liver disease, comprising or consisting of: a) performing an assay on a biological sample obtained from the human subject, wherein the assay determines the presence of an HSD17B13 Transcript D in the biological sample; and b) classifying the human subject as being at decreased risk for developing the liver disease if an HSD17B13 Transcript D is present in the biological sample, or classifying the human subject as being at increased risk for developing the liver disease if an HSD17B13 Transcript D is not present in the biological sample: wherein the HSD17B13 Transcript D comprises or consists of a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to SEQ ID NO:6, 15, 24, or 33; or wherein the HSD17B13 Transcript D is RNA and comprises or consists of SEQ ID NO:6, or a cDNA thereof comprising or consisting of SEQ ID NO:24, or wherein the HSD17B13 Transcript D is mRNA and comprises or consists of SEQ ID NO:15, or a cDNA thereof comprising or consisting of SEQ ID NO:33.

Disclosed herein are methods of determining a human subject's susceptibility or risk for developing a liver disease, comprising or consisting of: a) detecting whether HSD17B13 Isoform D is present in a biological sample obtained from the human subject; and b) classifying the human subject as being at decreased risk for developing the liver disease if HSD17B13 Isoform D is detected in the biological sample, or classifying the human subject as being at decreased risk for developing the liver disease if HSD17B13 Isoform D is not detected in the biological sample. In some embodiments, the HSD17B13 Isoform D comprises or consists of an amino acid

sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to SEQ ID NO:42. In some embodiments, the liver disease is a chronic liver disease. In some embodiments, the liver disease is selected from the group consisting of fatty liver disease, nonalcoholic fatty liver disease (NAFLD), alcoholic liver fatty liver disease, cirrhosis, viral hepatitis, hepatocellular carcinoma, simple steatosis, steatohepatitis, fibrosis, and non-alcoholic steatohepatitis (NASH). In some embodiments, the detecting comprises sequencing.

Disclosed herein are methods of determining a human subject's risk for progression to more clinically advanced stages of fatty liver disease, comprising or consisting of: a) performing an assay on a biological sample obtained from the human subject, wherein the assay determines whether a thymine is inserted between positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene, or whether a thymine is present at a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene; and b) classifying the human subject as being at decreased risk for progression to more clinically advanced stages of fatty liver disease if a thymine is inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene or if a thymine is present at a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene, or classifying the human subject as being at increased risk for progression to more clinically advanced stages of fatty liver disease if a thymine is not inserted between the positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1, or if a thymine is not present at a position of the variant HSD17B13 gene corresponding to position 12666 of SEQ ID NO:2. In some embodiments, the assay comprises or consists of: i) contacting the biological sample with a primer hybridizing to a region of the HSD17B13 gene that is within 50 nucleotides of positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1, or corresponding to position 12666 of SEQ ID NO:2; ii) extending the primer at least through the positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1, or corresponding to position 12666 of SEQ ID NO:2; and iii) determining whether a thymine is inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene, or whether a thymine is present at a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene, in an extension product of the primer. In some embodiments, the assay comprises or consists of contacting the biological sample with a primer or probe that specifically hybridizes to the variant HSD17B13 gene having a thymine at a position corresponding to position 12666 of SEQ ID NO:2, and not to the corresponding wild type

HSD17B13 gene under stringent conditions, and determining whether hybridization has occurred. In some embodiments, the variant HSD17B13 gene is detected by sequencing. In some embodiments, the method further comprises determining whether the human subject is homozygous for the variant HSD17B13 gene.

Disclosed herein are methods of determining a human subject's risk for progression to more clinically advanced stages of fatty liver disease, comprising or consisting of: a) performing an assay on a biological sample obtained from the human subject, wherein the assay determines the presence of an HSD17B13 Transcript D in the biological sample; and b) classifying the human subject as being at decreased risk for progression to more clinically advanced stages of fatty liver disease if an HSD17B13 Transcript D is present in the biological sample, or classifying the human subject as being at increased risk for progression to more clinically advanced stages of fatty liver disease if an HSD17B13 Transcript D is not present in the biological sample. In some embodiments, the HSD17B13 Transcript D comprises or consists of a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to SEQ ID NO:6, 15, 24, or 33. In some embodiments, the HSD17B13 Transcript D is RNA and comprises or consists of SEQ ID NO:6, or a cDNA thereof comprising or consisting of SEQ ID NO:24, or wherein the HSD17B13 Transcript D is mRNA and comprises or consists of SEQ ID NO:15, or a cDNA thereof comprising or consisting of SEQ ID NO:33. In some embodiments, the assay determines the expression level of HSD17B13 Transcript D in the biological sample, wherein an increased expression level of HSD17B13 Transcript D compared to a control sample from a control human subject homozygous for a wild type HSD17B13 allele indicates a decreased risk for progression to more clinically advanced stages of fatty liver disease, and wherein the same or a decreased expression level of HSD17B13 Transcript D compared to the control sample indicates an increased risk for progression to more clinically advanced stages of fatty liver disease. In some embodiments, the assay comprises or consists of contacting the biological sample with one or more primers or probes that specifically hybridize to a nucleic acid sequence of HSD17B13 Transcript D, or to the complement thereof, and determining whether hybridization has occurred. In some embodiments, the method further comprises specifically detecting Transcript D by using a nucleic acid molecule comprising from about 5 nucleotides up to about 50 nucleotides comprising: i) a nucleic acid molecule which specifically hybridizes to a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a nucleotide sequence of a SEQ ID NO:6, 15, 24, or 33, or to the complement thereof; ii) a nucleic acid molecule which specifically

hybridizes to exon 2 of Transcript D; and/or iii) a nucleic acid molecule which specifically hybridizes to the region which bridges exons 3 and 4 of Transcript D. In some embodiments, the one or more primers or probes specifically hybridize to SEQ ID NO:6, SEQ ID NO:15, SEQ ID NO:24, and/or SEQ ID NO:33. In some embodiments, the assay further comprises using a primer or probe which specifically hybridizes to one or more of Transcripts A, B, C, E, F, F', G and/or H but not to Transcript D, and determining that hybridization has not occurred. In some embodiments, the assay comprises reverse transcription polymerase chain reaction (RT-PCR) or quantitative RT-PCR (qRT-PCR). In some embodiments, the assay comprises sequencing.

Disclosed herein are methods of determining a human subject's risk for progression to more clinically advanced stages of fatty liver disease, comprising: a) detecting whether HSD17B13 Isoform D is present in a biological sample obtained from the human subject; and b) classifying the human subject as being at decreased risk for progression to more clinically advanced stages of liver disease if HSD17B13 Isoform D is detected in the biological sample. In some embodiments, the HSD17B13 Isoform D comprises an amino acid sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to SEQ ID NO:42. In some embodiments, the detecting comprises sequencing.

Brief Description Of The Drawings

The accompanying figures, which are incorporated in and constitute a part of this specification, illustrate several aspects and together with the description serve to explain the principles of the present disclosure.

Figures 1A and 1B show Manhattan plots (left) and quantile-quantile plots (right) of single nucleotide variant associations with serum transaminase levels in the GHS discovery cohort; 1A: variants in genes significantly associated with alanine aminotransferase (ALT) levels at $P < 1.0 \times 10^{-7}$; 1B: variants in genes significantly associated with aspartate aminotransferase (AST) levels at $P < 1.0 \times 10^{-7}$; the association tests were well calibrated, as shown by exome-wide quantile-quantile plots and genomic control lambda values.

Figures 2A and 2B show that *HSD17B13* rs72613567:TA is associated with reduced risk of alcoholic and nonalcoholic liver disease phenotypes; 2A: *HSD17B13* rs72613567 was associated with lower odds of nonalcoholic and alcoholic liver disease, cirrhosis, and

hepatocellular carcinoma in an allele dosage-dependent manner; odds ratios were calculated using logistic regression, with adjustment for age, age^2, sex, BMI, and principal components of ancestry; genotypic odds ratios for heterozygous (Het OR) and homozygous (Hom OR) carriers are also shown; 2B: in the Dallas Liver Study, *HSD17B13* rs72613567 was associated with lower odds of any liver disease in an allele dosage-dependent manner; similar allele dosage-dependent effects were observed across liver disease subtypes; odds ratios were calculated using logistic regression, with adjustment for age, age^2, sex, BMI, and self-reported ethnicity.

Figures 3A and 3B show that *HSD17B13* rs72613567:TA is associated with reduced risk of progression from simple steatosis to steatohepatitis and fibrosis; 3A: prevalence of histopathologically-characterized liver disease according to HSD17B13 rs72613567 genotype in 2,391 individuals with liver biopsies from the GHS bariatric surgery cohort; the prevalence of normal liver did not appear to differ by genotype (P = 0.5 by Chi-squared test for trend in proportions), but the prevalence of NASH decreased ($P = 1.6x10^{-4}$) and that of simple steatosis increased ($P = 1.1x10^{-3}$) with each TA allele; 3B: in the GHS bariatric surgery cohort, *HSD17B13* rs72613567 was associated with 13% and 52% lower odds of NASH, and 13% and 61% lower odds of fibrosis, in heterozygous and homozygous TA carriers, respectively; odds ratios were calculated using logistic regression, with adjustment for age, age^2, sex, BMI, and principal components of ancestry; genotypic odds ratios for heterozygous (Het OR) and homozygous (Hom OR) carriers are also shown.

Figures 4A, 4B, 4C, 4D, 4E, and 4F show expression, subcellular localization, and enzymatic activity of a novel *HSD17B13* transcript; 4A: expression of *HSD17B13* transcripts A and D in homozygous reference (T/T), heterozygous (T/TA), and homozygous alternate (TA/TA) carriers of the *HSD17B13* rs72613567 splice variant; coding regions in HSD17B13 gene are indicated in vertical rectangles, untranslated regions as thick lines, and introns as thin lines; the asterisk in transcript D indicates the A insertion from rs72613567; mRNA expression is displayed in FPKM units (Fragments Per Kilobase of transcript per Million mapped reads); 4B: HSD17B13 Western blot from fresh frozen human liver and HEK293 cell samples; human liver samples are from homozygous reference (T/T), heterozygous (T/TA), and homozygous alternate (TA/TA) carriers of the *HSD17B13* rs72613567 splice variant; cell samples are from HEK293 cells overexpressing non-tagged *HSD17B13* transcripts A and D; *HSD17B13* transcript D was translated to a truncated protein IsoD with lower molecular weight than HSD17B13 IsoA; 4C: HSD17B13 IsoD protein levels were lower than IsoA protein levels from both human liver (left) and cell (right) samples; protein level normalized to Actin was shown in the bar columns; ** P<0.001, *P<0.05; 4D: both HSD17B13 isoforms A and D are localized on lipid droplet

membrane; HepG2 stably overexpressing *HSD17B13* transcripts A or D were labelled with BODIPY to show lipid droplets and anti-Myc to show HSD17B13 localization; all figures are magnified to the same extent; scale bar indicates 10 μm; insets represent 4x amplification of the original images; 4E: enzymatic activity of HSD17B13 isoforms A and D to 17-beta estradiol (estradiol), leukotriene B4 (LTB4), and 13-Hydroxyoctadecadienoic acid (13(S)-HODE; HSD17B13 isoform D show <10% enzymatic activity of the corresponding values for isoform A; 4F: HSD17B13 isoform D when overexpressed in HEK293 cells did not show much conversion of estradiol (substrate) to estrone (product) when measured in the culture media, while overexpressed HSD17B13 isoform A showed robust conversion.

Figures 5A and 5B show regional association plots for alanine aminotransferase (ALT; 5A) and aspartate aminotransferase (AST; 5B) levels in the GHS Discovery Cohort in the region around HSD17B13; diamonds indicate the splice variant rs72613567; each circle indicates a single nucleotide variant with the color of the circle indicating the linkage disequilibrium (r^2 calculated in the DiscovEHR cohort) between that variant and rs72613567; lines indicate estimated recombination rates in HapMap; the bottom panels show the relative position and the transcribed strand of each gene in the locus; there were no significant associations between ALT or AST and coding or splice region variants in the neighboring gene HSD17B11 (most significant P-values 1.4×10^{-1} and 4.3×10^{-2} for ALT and AST, respectively).

Figures 6A, 6B, 6C, 6D, 6E, 6F, 6G, and 6H show mRNA expression of eight HSD17B13 Transcripts (A-H) in homozygous reference (T/T), heterozygous (T/TA), and homozygous alternate (TA/TA) allele carriers of the HSD17B13 splice variant; each transcript is illustrated with a corresponding gene model; coding regions in gene models are indicated in vertical rectangles, untranslated regions as thick lines, and introns as thin lines; the asterisks indicate the A insertion from rs72613567; transcripts are differentially expressed according to HSD17B13 genotype, as shown in the box plots; mRNA expression is displayed in FPKM units (Fragments Per Kilobase of transcript per Million mapped reads).

Figures 7A and 7B show a protein sequence alignment of all identified HSD17B13 isoforms (A-H).

Figures 8A, 8B, and 8C show that HSD17B13 isoform D protein has lower molecular weight and is unstable when overexpressed in HEK 293 cells; 8A: RT-PCR of HSD17B13 from HEK 293 cells overexpressing HSD17B13 transcripts A (IsoA) and D (IsoD) indicated that HSD17B13 IsoD RNA level was higher than IsoA RNA level; 8B: Western blot from the same cell lines indicated that HSD17B13 transcript D was translated to a truncated protein with lower molecular weight compared to HSD17B13 transcript A; 8C: HSD17B13 IsoD protein levels

were lower than IsoA protein levels although the RNA level was higher; HSD17B13 protein level was normalized to actin; *P<0.05.

Figure 9 shows similar localization patterns of HSD17B13 isoform A and isoform D to isolated lipid droplets (LD) derived from HepG2 stable cell lines; ADRP and TIP47 were used as lipid droplet markers; LAMP1, calreticulin, and COX IV were used as markers for the lysosomal, endoplasmic reticulum, and mitochondrial compartments, respectively; GAPDH was included as a cytosolic marker; and actin was used as a cytoskeletal marker; this experiment was repeated twice in HepG2 cells, with the above being representative of both runs; PNS=Post-nuclear fraction; TM=total membrane.

Figures 10A, 10B, 10C, 10D, and 10E show oleic acid increased triglyceride content in HepG2 cells overexpressing HSD17B13 Transcript A or D; 10A: oleic acid treatment in lipid droplets in both HSD17B13 transcript A and D expressing HepG2 cell lines to a similar degree; cells were labeled with BODIPY to show lipid droplets and anti-Myc to show HSD17B13 localization; scale bar indicates 10 μm and for all images; insets represent 4x amplification of the original images; 10B: treatment with increasing concentrations of oleic acid increased triglyceride (TG) content to a similar extent in control (GFP overexpressing cells), HSD17B13 transcript A and D cell lines; 10C: HSD17B13 transcripts A and D RNA levels were similar in the cell lines; RNA levels are shown reads per kilobase of transcript per million mapped reads (RPKM); 10D: Western blot from HepG2 cells overexpressing HSD17B13 transcripts A and D; HSD17B13 transcript D was translated to a truncated protein with lower molecular weight compared to HSD17B13 transcript A; 10E: HSD17B13 IsoD protein levels were lower than IsoA protein levels; protein level normalized to actin; **P<0.01.

Figure 11 shows K_m and V_{max} values for estradiol using purified recombinant HSD17B13 protein; for K_m and V_{max} determinations, assays were performed with a dose range of 17β-estradiol between 0.2 μM to 200 μM and time points from 5 minutes to 180 minutes, with 500 μM NAD^+ and 228 nM HSD17B13; V_{max} and K_m were then determined using the Michaelis-Menten model and Prism software (GraphPad Software, USA).

Additional advantages of the present disclosure will be set forth in part in the description which follows, and in part will be apparent from the description, or can be learned by practice of the embodiments disclosed herein. Advantages of the present disclosure will be realized and attained by means of the elements and combinations particularly pointed out in the appended claims. It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the embodiments, as claimed.

Description

Various terms relating to aspects of disclosure are used throughout the specification and claims. Such terms are to be given their ordinary meaning in the art, unless otherwise indicated. Other specifically defined terms are to be construed in a manner consistent with the definition provided herein.

Unless otherwise expressly stated, it is in no way intended that any method or aspect set forth herein be construed as requiring that its steps be performed in a specific order. Accordingly, where a method claim does not specifically state in the claims or descriptions that the steps are to be limited to a specific order, it is in no way intended that an order be inferred, in any respect. This holds for any possible non-expressed basis for interpretation, including matters of logic with respect to arrangement of steps or operational flow, plain meaning derived from grammatical organization or punctuation, or the number or type of aspects described in the specification.

As used herein, the singular forms “a,” “an” and “the” include plural referents unless the context clearly dictates otherwise.

As used herein, the terms “subject” and “patient” are used interchangeably. A subject may include any animal, including mammals. Mammals include, without limitation, farm animals (*e.g.*, horse, cow, pig, sheep, etc.), companion animals (*e.g.*, dog, cat, etc.), laboratory animals (*e.g.*, mouse, rat, rabbits, etc.), and non-human primates (*e.g.*, monkeys, apes, etc.). In some embodiments, the subject is a human.

As used herein, a “nucleic acid,” a “nucleic acid molecule,” a “nucleotide sequence,” “polynucleotide,” or “oligonucleotide” can comprise a polymeric form of nucleotides of any length, may comprise DNA and/or RNA, and can be single-stranded, double-stranded, or multiple stranded. One strand of a nucleic acid also refers to its complement.

The term “compsising” throughout the present disclosure may be replaced with “consisting” or “consisting essentially of” in particular embodiments as desired.

As used herein, the phrase “corresponding to” or grammatical variations thereof when used in the context of the numbering of a particular amino acid or nucleotide sequence or position refers to the numbering of a specified reference sequence when the particular amino acid or nucleotide sequence is compared to the reference sequence (*e.g.*, with the reference sequence herein being the nucleic acid molecule or polypeptide of (wild type or full length) HSD17B13). In other words, the residue (*e.g.*, amino acid or nucleotide) number or residue (*e.g.*, amino acid or nucleotide) position of a particular polymer is designated with respect to the

reference sequence rather than by the actual numerical position of the residue within the particular amino acid or nucleotide sequence. For example, a particular amino acid sequence can be aligned to a reference sequence by introducing gaps to optimize residue matches between the two sequences. In these cases, although the gaps are present, the numbering of the residue in the particular amino acid or nucleotide sequence is made with respect to the reference sequence to which it has been aligned.

For example, a nucleic acid molecule having a thymine inserted between positions of the *HSD17B13* gene corresponding to positions 12665 and 12666 of SEQ ID NO:1 (or position 12666 of SEQ ID NO:2) can be identified by performing a sequence alignment between the particular HSD17B13 nucleic acid of interest and the nucleotide sequence of SEQ ID NO:1 and/or SEQ ID NO:2. A variety of computational algorithms exist that can be used for performing sequence alignments. For example, by using the NCBI BLAST algorithm (Altschul et al., 1997, Nucleic acid molecules Res., 25, 3389-3402) or CLUSTALW software (Sievers et al., 2014, Methods Mol. Biol., 1079, 105-116) sequence alignments may be performed. However, sequences can also be aligned manually.

It has been observed in accordance with the present disclosure that certain variations in HSD17B13 associate with a decreased risk of developing a liver disease. A variant in the *HSD17B13* gene segregating with the phenotype of a decreased risk of developing a liver disease in affected family members has been identified in accordance with the present disclosure. For example, a genetic alteration that results in a thymine inserted between positions of the *HSD17B13* gene corresponding to positions 12665 and 12666 of SEQ ID NO:1 (or position 12666 of SEQ ID NO:2) has been observed to indicate that the human having such an alteration may have a decreased risk of developing a liver disease. Therefore, human subjects that do not possess this thymine insertion and, therefore may have an increased risk of developing a liver disease, or have a liver disease, may be treated such that liver disease is inhibited, the symptoms thereof are reduced, and/or development of symptoms is repressed. Accordingly, the present disclosure provides recombinant variant *HSD17B13* nucleic acid molecules, including RNA, or cDNA derived therefrom, or mRNA, or cDNA derived therefrom, as well as recombinant variant HSD17B13 polypeptides. Additionally, the disclosure provides methods for leveraging the identification of such variants in subjects to identify or stratify risk in such subjects of developing liver disease, or to diagnose subjects as having liver disease, such that subjects at risk or subjects with active disease may be treated.

Provided herein is an *HSD17B13* variant discovered to be associated with reduced alanine and aspartate transaminase levels; a reduced risk of chronic liver diseases including

nonalcoholic and alcoholic liver fatty liver disease, cirrhosis, and hepatocellular carcinoma; and reduced progression from simple steatosis to more clinically advanced stages of chronic liver disease. Also provided herein are previously unidentified transcripts of the *HSD17B13* gene associated with the variant.

Nucleic acid molecules and polypeptides related to variants of *HSD17B13*, and cells comprising those nucleic acid molecules and polypeptides are provided herein. Also provided are methods for detecting the presence of the *HSD17B13* rs72613567 variant gene in a biological sample comprising genomic DNA, for detecting the presence or levels of any one of *HSD17B13* transcripts C, D, E, F, F', G, and H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom), and particularly Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom), in a biological sample comprising RNA or cDNA derived therefrom, mRNA or cDNA derived therefrom, and for detecting the presence or levels of any one of HSD17B13 protein isoforms C, D, E, F, F', G, or H, and particularly D, in a biological sample comprising protein. Also provided are methods for determining a subject's susceptibility to or risk of developing a liver disease. Also provided are methods for diagnosing a subject with liver disease or at risk for developing a liver disease. Also provided are methods for determining a subject's risk for progression to more clinically advanced stages of fatty liver disease. Also provided are methods for modifying a cell through use of expression vectors for expressing a recombinant *HSD17B13* gene or a nucleic acid encoding an HSD17B13 protein.

The present disclosure provides nucleic acid molecules and polypeptides related to variants of *HSD17B13* (also known as hydroxysteroid 17-beta dehydrogenase 13, 17-beta-hydroxysteroid dehydrogenase 13, 17β-hydroxysteroid dehydrogenase-13, 17β-HSD13, short-chain dehydrogenase/reductase 9, *SCDR9, HMFN0376*, *NIIL497*, and *SDR16C3*). The human *HSD17B13* gene is approximately 19 kb in length and includes seven exons and six introns located at 4q22.1 in the genome. Exemplary human HSD17B13 protein sequences are assigned UniProt Accession No. Q7Z5P4 (Q7Z5P4-1 and Q7Z5P4-2, respectively) and NCBI Reference Sequence Nos. NP_835236 and NP_001129702. Exemplary human HSD17B13 nucleic acid molecules are assigned NCBI Reference Sequence Nos. NM_178135 and NM_001136230.

In particular, provided herein is a splice variant of *HSD17B13* (rs72613567) having an insertion of an adenine adjacent to the donor splice site in intron 6. The adenine is an insertion on the forward (plus) strand of the chromosome, which corresponds to an inserted thymine on the reverse (minus) strand of the chromosome. Because the human *HSD17B13* gene is transcribed in the reverse direction, this nucleotide insertion is reflected as an inserted thymine in the exemplary variant *HSD17B13* rs72613567 sequence provided in SEQ ID NO:2 relative to the

exemplary wild type *HSD17B13* gene sequence provided in SEQ ID NO:1. The insertion will therefore be referred to herein as a thymine inserted between positions 12665 and 12666 in SEQ ID NO:1 or at position 12666 in SEQ ID NO:2.

Two transcripts (A and B; SEQ ID NOs:21 and 22, respectively) were previously identified to be expressed in subjects with the wild type *HSD17B13* gene. Transcript A includes all seven exons of the *HSD17B13* gene, whereas exon 2 is skipped in Transcript B. Transcript A is the dominant transcript in wild type subjects. Provided herein, however, are six additional, previously unidentified, *HSD17B13* transcripts that are expressed (C, D, E, F, G, and H, SEQ ID NOs: 23, 24, 25, 26, 28, and 29, respectively). In Transcript C, exon 6 is skipped compared to Transcript A. In Transcript D, there is an insertion of a guanine 3' of exon 6, resulting in a frameshift in and premature truncation of exon 7 compared to Transcript A. In Transcript E, there is an additional exon between exons 3 and 4 compared to Transcript A. In Transcript F, which is expressed only in *HSD17B13* rs72613567 variant carriers, there is read-through from exon 6 into intron 6 compared to Transcript A. In Transcript G, exon 2 is skipped, and there is an insertion of a guanine 3' of exon 6, resulting in a frameshift in and premature truncation of exon 7 compared to Transcript A. In Transcript H, there is an additional exon between exons 3 and 4, and there is an insertion of a guanine 3' of exon 6, resulting in a frameshift in and premature truncation of exon 7 compared to Transcript A. Transcripts C, D, F, G, and H are dominant in *HSD17B13* rs72613567 variant carriers, with Transcript D being the most abundant transcript in carriers of the *HSD17B13* rs72613567 variant. Also provided herein is one additional, previously unidentified, *HSD17B13* transcript that is expressed at low levels (F', SEQ ID NO:27). Like Transcript F, Transcript F' also includes a read-through from exon 6 into intron 6 compared to Transcript A, but, in contrast to Transcript F, the read-through does not include the inserted thymine present in the *HSD17B13* rs72613567 variant gene.

The nucleotide and amino acid sequences listed in the accompanying sequence listing are shown using standard letter abbreviations for nucleotide bases, and three-letter code for amino acids. The nucleotide sequences follow the standard convention of beginning at the 5' end of the sequence and proceeding forward (i.e., from left to right in each line) to the 3' end. Only one strand of each nucleotide sequence is shown, but the complementary strand is understood to be included by any reference to the displayed strand. The amino acid sequences follow the standard convention of beginning at the amino terminus of the sequence and proceeding forward (i.e., from left to right in each line) to the carboxy terminus.

SEQ ID NO:1 is the wild type HSD17B13 genomic sequence (Human Genome Assembly GRCh38). Transcripts more prevalent in subjects with wild type HSD17B13 gene include Transcript A, Transcript B, Transcript E, and Transcript F'.

SEQ ID NO:2 is an HSD17B13 genomic sequence variant (Human Genome Assembly GRCh38; rs72613567; insertion of T at chr4: 87310241-87310240; insertion of T at position 12666). Transcripts more prevalent in subjects with the variant rs72613567 HSD17B13 gene include Transcript C, Transcript D, Transcript F, Transcript G, and Transcript H.

SEQ ID NO:53 is an endogenous HSD17B13 Promoter (-499 to 100 relative to transcription start site (TSS)).

As used herein, the term "transcript" means any one or more of the RNA or mRNA molecule, or the corresponding cDNA molecule derived therefrom, disclosed in the table below, unless otherwise indicated by the context of its use. The sequence identifier nomenclature for the various transcripts is listed in the table below. The RNA transcript is shown along with its cDNA counterpart, and the mRNA transcript is shown along with its cDNA counterpart.

Transcript	**RNA**	**cDNA (RNA)**	**mRNA**	**cDNA (mRNA)**	**Protein**
A	SEQ ID NO:3	SEQ ID NO:21	SEQ ID NO:12	SEQ ID NO:30	SEQ ID NO:39
B	SEQ ID NO:4	SEQ ID NO:22	SEQ ID NO:13	SEQ ID NO:31	SEQ ID NO:40
C	SEQ ID NO:5	SEQ ID NO:23	SEQ ID NO:14	SEQ ID NO:32	SEQ ID NO:41
D	SEQ ID NO:6	SEQ ID NO:24	SEQ ID NO:15	SEQ ID NO:33	SEQ ID NO:42
E	SEQ ID NO:7	SEQ ID NO:25	SEQ ID NO:16	SEQ ID NO:34	SEQ ID NO:43
F	SEQ ID NO:8	SEQ ID NO:26	SEQ ID NO:17	SEQ ID NO:35	SEQ ID NO:44
F'	SEQ ID NO:9	SEQ ID NO:27	SEQ ID NO:18	SEQ ID NO:36	SEQ ID NO:45
G	SEQ ID NO:10	SEQ ID NO:28	SEQ ID NO:19	SEQ ID NO:37	SEQ ID NO:46
H	SEQ ID	SEQ ID	SEQ ID	SEQ ID	SEQ ID NO:47

	NO:11	NO:29	NO:20	NO:38	

Accordingly, as used herein, the term "Transcript A" means any one or more of SEQ ID NO:3, SEQ ID NO:12, SEQ ID NO:21, and/or SEQ ID NO:30; "Transcript B" means any one or more of SEQ ID NO:4, SEQ ID NO:13, SEQ ID NO:22, and/or SEQ ID NO:31; "Transcript C" means any one or more of SEQ ID NO:5, SEQ ID NO:14, SEQ ID NO:23, and/or SEQ ID NO:32; "Transcript D" means any one or more of ii) SEQ ID NO:6, SEQ ID NO:15, SEQ ID NO:24, and/or SEQ ID NO:33; "Transcript E" means any one or more of SEQ ID NO:7, SEQ ID NO:16, SEQ ID NO:25, and/or SEQ ID NO:34; "Transcript F" means any one or more of SEQ ID NO:8, SEQ ID NO:17, SEQ ID NO:26, and/or SEQ ID NO:35; "Transcript F'" means any one or more of SEQ ID NO:9, SEQ ID NO:18, SEQ ID NO:27, and/or SEQ ID NO:36; "Transcript G" means any one or more of SEQ ID NO:10, SEQ ID NO:19, SEQ ID NO:28, and/or SEQ ID NO:37; and "Transcript H" means any one or more of SEQ ID NO:11, SEQ ID NO:20, SEQ ID NO:29, and/or SEQ ID NO:38, unless otherwise indicated b the context.

The nucleotide positions of the exons within the *HSD17B13* genes for each Transcript are provided below.

Nucleotide Positions in SEQ ID NO:1 for Exons of HSD17B13 Transcripts More Prevalent in Subjects Homozygous for Wild Type *HSD17B13* Gene.

	Transcript A	**Transcript B**	**Transcript E**	**Transcript F'**
Exon 1	1-275	1-275	1-275	1-275
Exon 2	4471-4578	skipped	4471-4578	4471-4578
Exon 3	5684-5815	5684-5815	5684-5815	5684-5815
Exon 3'	not present	not present	6210-6281	not present
Exon 4	7308-7414	7308-7414	7308-7414	7308-7414
Exon 5	8947-9084	8947-9084	8947-9084	8947-9084
Exon 6	12548-12664	12548-12664	12548-12664	12548-13501*
Exon 7	17599-19118	17599-19118	17599-19118	skipped

*Includes read-through from exon 6 into intron 6; read-through = positions 12665-13501

Nucleotide Positions in SEQ ID NO:2 for Exons of HSD17B13 Transcripts More Prevalent in Subjects Homozygous for rs72613567 *HSD17B13* Variant Gene (Insertion of T at Position 12666).

	Transcript C	**Transcript D**	**Transcript F**	**Transcript G**	**Transcript H**

Exon 1	1-275	1-275	1-275	1-275	1-275
Exon 2	4471-4578	4471-4578	4471-4578	skipped	4471-4578
Exon 3	5684-5815	5684-5815	5684-5815	5684-5815	5684-5815
Exon 3'	not present	not present	not present	not present	6210-6281
Exon 4	7308-7414	7308-7414	7308-7414	7308-7414	7308-7414
Exon 5	8947-9084	8947-9084	8947-9084	8947-9084	8947-9084
Exon 6	Skipped	12548-12665^	12548-13502*	12548-12665^	12548-12665^
Exon 7	17600-19119	17600-19119	skipped	17600-19119	17600-19119

^Includes additional residue 12665 at 3' end compared to Transcript A

*Includes read-through from exon 6 into intron 6; read-through = positions 12665-13502.

The corresponding HSD17B13 isoform proteins include: i) Isoform A (SEQ ID NO:39; region encoded by Exon 1 = 1-70, region encoded by Exon 2 = 71-106, region encoded by Exon 3 = 107-150, region encoded by Exon 4 = 151-185, region encoded by Exon 5 = 186-232, region encoded by Exon 6v1 = 233-271, and region encoded by Exon 7 = 272-300); ii) Protein Isoform B (SEQ ID NO:40; region encoded by Exon 1 = 1-70, Exon 2 = skipped, region encoded by Exon 3 = 71-114, region encoded by Exon 4 = 115-149, region encoded by Exon 5 = 150-196, region encoded by Exon 6v1 = 197-235, and region encoded by Exon 7 = 236-264); iii) Protein Isoform C (SEQ ID NO:41; region encoded by Exon 1 = 1-70, region encoded by Exon 2 = 71-106, region encoded by Exon 3 = 107-150, region encoded by Exon 4 = 151-185, region encoded by Exon 5 = 186-232, Exon 6 = skipped, and region encoded by Exon 7 = 233-261); iv) Protein Isoform D (SEQ ID NO:42; region encoded by Exon 1 = 1-70, region encoded by Exon 2 = 71-106, region encoded by Exon 3 = 107-150, region encoded by Exon 4 = 151-185, region encoded by Exon 5 = 186-232, region encoded by Exon 6v2 = 233-271, and region encoded by Exon 7 = 272-274); v) Protein Isoform E (SEQ ID NO:43; region encoded by Exon 1 = 1-70, region encoded by Exon 2 = 71-106, region encoded by Exon 3 = 107-150, region encoded by Exon 3' = 151-174, region encoded by Exon 4 = 175-209, region encoded by Exon 5 = 210-256, region encoded by Exon 6v1 = 257-295, and region encoded by Exon 7 = 296-324); vi) Protein Isoform F (SEQ ID NO:44; region encoded by Exon 1 = 1-70, region encoded by Exon 2 = 71-106, region encoded by Exon 3 = 107-150, region encoded by Exon 4 = 151-185, region encoded by Exon 5 = 186-232, region encoded by Exon 6v3 = 233-284, and region encoded by read-through into Intron 6 = 272-284); vii) Protein Isoform F' (SEQ ID NO:45; region encoded by Exon 1 = 1-70, region encoded by Exon 2 = 71-106, region encoded by Exon 3 = 107-150, region encoded by Exon 4 = 151-185, region encoded by Exon 5 = 186-232, and region encoded by Exon 6v4 = 233-271); viii) Protein Isoform G (SEQ ID NO:46; region encoded by Exon 1 = 1-70, Exon 2 =

skipped, region encoded by Exon 3 = 71-114, region encoded by Exon 4 = 115-149, region encoded by Exon 5 = 150-196, region encoded by Exon 6v2 = 197-235, and region encoded by Exon 7 = 236-238); and ix) Protein Isoform H (SEQ ID NO:47; region encoded by Exon 1 = 1-70, region encoded by Exon 2 = 71-106, region encoded by Exon 3 = 107-150, region encoded by Exon 3' = 151-174, region encoded by Exon 4 = 175-209, region encoded by Exon 5 = 210-256, region encoded by Exon 6v2 = 257-295, and region encoded by Exon 7 = 296-298).

As explained in more detail elsewhere herein, the variant *HSD17B13* rs72613567 is associated with reduced alanine and aspartate transaminase levels and a reduced risk of chronic liver diseases including nonalcoholic and alcoholic liver fatty liver disease, cirrhosis, and hepatocellular carcinoma. The variant *HSD17B13* rs72613567 is also associated with reduced progression from simple steatosis to more clinically advanced stages of chronic liver disease.

Disclosed herein are variant *HSD17B13* nucleic acid molecules, including a variant *HSD17B13* gene and variant HSD17B13 transcripts. Also disclosed are nucleic acid molecules that hybridize under stringent or moderate conditions with any of the nucleic acid molecules disclosed herein. Such nucleic acid molecules can be useful, for example, to express *HSD17B13* variant proteins or as primers, probes, antisense RNAs, shRNAs, and siRNAs, each of which is described in more detail elsewhere herein. In any of the embodiments described herein, the nucleic acid molecules and/or polypeptides can be isolated nucleic acid molecules or isolated polypeptides.

The present disclosure provides nucleic acid molecules comprising or consisting of at least 15 contiguous nucleotides of an HSD17B13 gene, wherein the contiguous nucleotides are at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to a corresponding sequence in SEQ ID NO:2, and having a thymine at a position corresponding to position 12666 of SEQ ID NO:2. In some embodiments, the nucleic acid molecules comprise or consist of at least 15 contiguous nucleotides of an HSD17B13 gene, wherein the contiguous nucleotides are at least about 90% identical to a corresponding sequence in SEQ ID NO:2, and having a thymine at a position corresponding to position 12666 of SEQ ID NO:2.

The present disclosure provides nucleic acid molecules comprising or consisting of a nucleotide sequence that encodes a polypeptide having an amino acid sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to the amino acid sequence of HSD17B13 Isoform D (SEQ ID NO:42). In some embodiments, the nucleic acid molecules comprise or consist of a nucleotide sequence that encodes a polypeptide having an amino acid sequence that is at least

about 90% identical to the amino acid sequence of HSD17B13 Isoform D (SEQ ID NO:42). In some embodiments, the nucleic acid molecule comprises or consists of a nucleotide sequence that encodes a polypeptide having the amino acid sequence of HSD17B13 Isoform D (SEQ ID NO:42). In some embodiments, these nucleic acid molecules encode a polypeptide having 274 amino acids. In some embodiments, these nucleic acid molecules encode a polypeptide having a C-terminal Val-Ser-Ser. In some embodiments, these nucleic acid molecules encode a polypeptide associated with a decreased risk of developing any of the liver diseases described herein or decreased risk of progression to more clinically advanced stages of fatty liver disease.

The present disclosure provides nucleic acid molecules comprising at least 15 contiguous nucleotides of an *HSD17B13* gene and having a thymine at a position corresponding to position 12666 (or thymines at positions corresponding to positions 12666 and 12667) of the *HSD17B13* rs72613567 variant gene (SEQ ID NO:2). That is, disclosed herein are nucleic acid molecules comprising at least 15 contiguous nucleotides of an *HSD17B13* gene and having a thymine inserted between nucleotides corresponding to positions 12665 and 12666 of the wild type *HSD17B13* gene (SEQ ID NO:1). Such nucleic acid molecules can be useful, for example, to express *HSD17B13* variant transcripts and isoform proteins.

The *HSD17B13* gene can be an *HSD17B13* gene from any organism. For example, the *HSD17B13* gene can be a human *HSD17B13* gene or an ortholog from another organism, such as a non-human mammal, a rodent, a mouse, or a rat. It is understood that gene sequences within a population can vary due to polymorphisms such as single-nucleotide polymorphisms. The examples provided herein are only exemplary sequences. Other sequences are also possible. As one example, the at least 15 contiguous nucleotides can be at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to a corresponding sequence in the *HSD17B13* rs72613567 variant (SEQ ID NO:2) including position 12666 or positions 12666 and 12667 of SEQ ID NO:2. In some embodiments, the at least 15 contiguous nucleotides can be at least about 90% identical to a corresponding sequence in the *HSD17B13* rs72613567 variant (SEQ ID NO:2) including position 12666 or positions 12666 and 12667 of SEQ ID NO:2. In some embodiments, the nucleic acid molecule comprises at least 15 contiguous nucleotides of SEQ ID NO:2 including position 12666 or positions 12666 and 12667 of SEQ ID NO:2. In some embodiments, the at least 15 contiguous nucleotides can be at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to a corresponding sequence in the wild type *HSD17B13* gene (SEQ ID NO:1) including positions 12665 and 12666 of SEQ ID NO:1, wherein a thymine is present between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1.

In some embodiments, the at least 15 contiguous nucleotides can be at least about 90% identical to a corresponding sequence in the wild type *HSD17B13* gene (SEQ ID NO:1) including positions 12665 and 12666 of SEQ ID NO:1, wherein a thymine is present between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1. In some embodiments, the nucleic acid molecule comprises at least 15 contiguous nucleotides of SEQ ID NO:1 including positions 12665 and 12666 of SEQ ID NO:1, wherein a thymine is present between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1.

In some cases, the isolated nucleic acid molecule can comprise an *HSD17B13* minigene in which one or more nonessential segments of the gene have been deleted with respect to a corresponding wild type *HSD17B13* gene. As one example, the deleted segments comprise one or more intronic sequences. In some embodiments, the *HSD17B13* minigenes can comprise, for example, exons corresponding to exons 1-7 from *HSD17B13* Transcript D and an intron corresponding to intron 6 in SEQ ID NO:2. In some embodiments, an *HSD17B13* minigene may comprise exons 1-7 and intron 6 from SEQ ID NO:2. Minigenes are described in more detail elsewhere herein.

The present disclosure provides nucleic acid molecules corresponding to all or part of an RNA transcript, such as Transcript A, Transcript B, Transcript C, Transcript D, Transcript E, Transcript F, Transcript F', Transcript G, and Transcript H, or a corresponding cDNA thereof, or an mRNA transcript, such as Transcript A, Transcript B, Transcript C, Transcript D, Transcript E, Transcript F, Transcript F', Transcript G, and Transcript H, or a corresponding cDNA thereof.

The present disclosure provides nucleic acid molecules corresponding to all or part of an RNA transcript, such as Transcript C, Transcript D, Transcript E, Transcript F, Transcript F', Transcript G, and Transcript H, or a corresponding cDNA thereof, or an mRNA transcript, such as Transcript C, Transcript D, Transcript E, Transcript F, Transcript F', Transcript G, and Transcript H, or a corresponding cDNA thereof.

The present disclosure provides nucleic acid molecules corresponding to all or part of an RNA transcript, such as Transcript C, Transcript D, Transcript F, Transcript G, and Transcript H, or a corresponding cDNA thereof, or an mRNA transcript, such as Transcript C, Transcript D, Transcript F, Transcript G, and Transcript H, or a corresponding cDNA thereof.

The present disclosure provides nucleic acid molecules corresponding to all or part of RNA Transcript D, or a corresponding cDNA, or an mRNA Transcript D, or a corresponding cDNA.

Such isolated nucleic acid molecules can be useful, for example, to express *HSD17B13* variant transcripts and proteins.

In some embodiments, the nucleic acid molecule comprises or consists of a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a nucleotide sequence of an HSD17B13 Transcript D (SEQ ID NO:6, 15, 24, or 33). In some embodiments, the nucleic acid molecule is RNA and comprises or consists of SEQ ID NO:6, or a cDNA thereof comprising or consisting of SEQ ID NO:24, or wherein the nucleic acid molecule is mRNA and comprises or consists of SEQ ID NO:15, or a cDNA thereof comprising or consisting of SEQ ID NO:33.

HSD17B13 Transcript D, Transcript G, and Transcript H include an insertion of a guanine at the 3' end of exon 6, resulting in a frameshift in exon 7 and premature truncation of the region of the HSD17B13 protein encoded by exon 7 compared to Transcript A. Accordingly, provided herein are nucleic acid molecules comprising a segment (e.g., at least 15 contiguous nucleotides) present in Transcripts D, G, and H (or fragments or homologs thereof) that is not present in Transcript A (or a fragment or homolog thereof). Also provided herein are nucleic acid molecules comprising a segment (e.g., at least 15 contiguous nucleotides) present in Transcripts D (or fragments or homologs thereof) that is not present in Transcript A (or a fragment or homolog thereof). Such regions can be readily identified by comparing the sequences of the Transcripts. For example, provided herein are nucleic acid molecules comprising or consisting of at least 15 contiguous nucleotides (e.g., at least 20 contiguous nucleotides or at least 30 contiguous nucleotides) encoding all or part of an HSD17B13 protein, wherein a segment of the contiguous nucleotides (e.g., at least 5 contiguous nucleotides, at least 10 contiguous nucleotides or at least 15 contiguous nucleotides) is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to a region spanning the exon 6-exon 7 boundary in Transcript D, Transcript G , or Transcript H, and the segment includes a guanine at a residue corresponding to residue 878 at the 3' end of exon 6 in Transcript D (i.e., an insertion of a guanine at the 3' end of exon 6 relative to Transcript A in addition to the guanine at the start of exon 7), a residue corresponding to residue 770 at the 3' end of exon 6 in Transcript G (i.e., an insertion of a guanine at the 3' end of exon 6 relative to Transcript B in addition to the guanine at the start of exon 7), or a residue corresponding to residue 950 at the 3' end of exon 6 in Transcript H (i.e., an insertion of a guanine at the 3' end of exon 6 relative to Transcript E in addition to the guanine at the start of exon 7). It is understood that such a nucleic acid would include a sufficient number of nucleotides in each of exons 6 and 7 to distinguish the inserted guanine from other features in the *HSD17B13* Transcripts (e.g., from the guanine at the start of exon 7, from the read-through into intron 6 in Transcript F, or from the deleted exon 6 in Transcript C).

As one example, the nucleic acid molecule can comprise or consist of at least 15 contiguous nucleotides (e.g., at least 20 contiguous nucleotides or at least 30 contiguous nucleotides) of Transcript D spanning the exon 6-exon 7 boundary, optionally comprising exons 6 and 7 of Transcript D, and optionally comprising the entire sequence of Transcript D.

In some embodiments, the nucleic acid molecule further comprises a segment present in Transcript D (or a fragment or homolog thereof) that is not present in Transcript G (or a fragment or homolog thereof), and the nucleic acid molecule further comprises a segment present in Transcript D (or a fragment or homolog thereof) that is not present in Transcript H (or a fragment or homolog thereof). Such regions can be readily identified by comparing the sequences of the Transcripts. For example, such nucleic acid molecules can comprise or consist of a segment of the contiguous nucleotides (e.g., at least 5 contiguous nucleotides, at least 10 contiguous nucleotides or at least 15 contiguous nucleotides) that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to a region spanning the boundary of exons 3 and 4 of Transcript D to distinguish from Transcript H. Likewise, such nucleic acid molecules can comprise or consist of a segment of the contiguous nucleotides (e.g., at least 5 contiguous nucleotides, at least 10 contiguous nucleotides or at least 15 contiguous nucleotides) that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to a region within exon 2 of Transcript D, a region spanning the exon 1-exon 2 boundary of Transcript D, or a region spanning the exon 2-exon 3 boundary of Transcript D to distinguish from Transcript G. In some embodiments, the nucleic acid molecules can comprise or consist of a sequence at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to the sequence set forth in Transcript D and encodes an HSD17B13 isoform protein comprising the sequence set forth in Isoform D. Like Transcript D, Transcript H includes an insertion of a guanine 3' of exon 6 compared to Transcript A. Transcript H further includes an additional exon (exon 3') between exons 3 and 4 compared to Transcript A and Transcript D. Accordingly, provided herein are nucleic acid molecules as described above comprising a segment present in Transcripts D, G, and H (or fragments or homologs thereof) that is not present in Transcript A (or a fragment or homolog thereof) but further comprising a segment (e.g., at least 15 contiguous nucleotides) of Transcript H (or a fragment or homolog thereof) that is not present in Transcript D (or a fragment or homolog thereof). Such regions can be readily identified by comparing the sequences of the Transcripts. For example, provided herein are nucleic acid molecules as described for Transcript D, wherein a segment of the contiguous nucleotides (e.g., at least 5 contiguous nucleotides, at least 10 contiguous nucleotides

or at least 15 contiguous nucleotides) is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to a region within exon 3' of Transcript H, a region spanning the exon 3-exon 3' boundary of Transcript H, or a region spanning the exon 3'-exon 4 boundary of Transcript H. It is understood that such nucleic acid molecules would include a sufficient number of nucleotides in each of exons 3 and 3' or each of exons 3' and 4 to distinguish from other features in the *HSD17B13* transcripts (e.g., from the boundary of exons 3 and 4). For example, the region of exon 3' can comprise the entire exon 3'. Optionally, the nucleic acid molecules can comprise or consist of a sequence at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to the sequence set forth in Transcript H and encodes an HSD17B13 protein comprising Isoform H.

As one example, the nucleic acid molecules can comprise or consist of at least 15 contiguous nucleotides (e.g., at least 20 contiguous nucleotides or at least 30 contiguous nucleotides) of Transcript H including a region within exon 3', a region spanning the exon 3-exon 3' boundary, or a region spanning the exon 3'-exon 4 boundary, optionally comprising the entire exon 3' of Transcript H, and optionally comprising the entire sequence of Transcript H.

Like Transcript D, Transcript G includes an insertion of a guanine 3' of exon 6 compared to Transcript A. In addition, however, Transcript G is missing exon 2 compared to Transcript A and Transcript D (i.e., Transcript G includes an exon 1-exon 3 boundary not present in Transcripts A and D). Accordingly, provided herein are nucleic acid molecules as described above comprising or consisting of a segment present in Transcripts D, G, and H (or fragments or homologs thereof) that is not present in Transcript A (or a fragment or homolog thereof) but further comprising a segment (e.g., at least 15 contiguous nucleotides) from Transcript G (or a fragment or homolog thereof) that is not present in Transcript D (or a fragment or homolog thereof). Such regions can be readily identified by comparing the sequences of the Transcripts. For example, provided herein are nucleic acid molecules as described for Transcript D, wherein a segment of the contiguous nucleotides (e.g., at least 5 contiguous nucleotides, at least 10 contiguous nucleotides or at least 15 contiguous nucleotides) is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to a region spanning the exon 1-exon 3 boundary in Transcript G. It is understood that such nucleic acid molecules would include a sufficient number of nucleotides in each of exons 1 and 3 to distinguish from other features in the *HSD17B13* Transcripts (e.g., the boundary of exons 1 and 2 or the boundary of exons 2 and 3). For example, the region can comprise the entirety of exons 1 and 3 in Transcript G. Optionally, the nucleic acid molecules comprise or consist of a

sequence at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to the sequence set forth in Transcript G and encodes an HSD17B13 protein comprising the sequence set forth in Isoform G.

As one example, the nucleic acid molecules can comprise or consist of at least 15 contiguous nucleotides (e.g., at least 20 contiguous nucleotides or at least 30 contiguous nucleotides) of Transcript G including a region spanning the exon 1-exon 3 boundary, optionally comprising the exons 1 and 3 of Transcript G, and optionally comprising the entire sequence of Transcript G.

Also provided herein are nucleic acid molecules comprising or consisting of a segment (e.g., at least 15 contiguous nucleotides) present in Transcript E (or a fragment or homolog thereof) that is not present in Transcript A (or a fragment or homolog thereof). Such regions can be readily identified by comparing the sequences of the Transcripts. Transcript E includes an additional exon between exons 3 and 4 compared to Transcript A. Accordingly, provided herein are nucleic acid molecules comprising or consisting of at least 15 contiguous nucleotides (e.g., at least 20 contiguous nucleotides or at least 30 contiguous nucleotides) encoding all or part of an HSD17B13 protein, wherein a segment of the contiguous nucleotides (e.g., at least 5 contiguous nucleotides, at least 10 contiguous nucleotides or at least 15 contiguous nucleotides) is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to a region within exon 3' of Transcript E, a region spanning the exon 3-exon 3' boundary of Transcript E, or a region spanning the exon 3'-exon 4 boundary of Transcript E. It is understood that such nucleic acid molecules would include a sufficient number of nucleotides in each of exons 3 and 3' or each of exons 3' and 4 to distinguish from other features in the *HSD17B13* transcripts (e.g., from the boundary of exons 3 and 4). For example, the region of exon 3' can comprise the entire exon 3'. Optionally, the nucleic acid molecules further comprise a segment (e.g., at least 15 contiguous nucleotides) from Transcript E (or a fragment or homolog thereof) that is not present in Transcript H (or a fragment or homolog thereof). Such regions can be readily identified by comparing the sequences of the Transcripts. For example, provided herein are nucleic acid molecules as described above, wherein a segment of the contiguous nucleotides (e.g., at least 5 contiguous nucleotides, at least 10 contiguous nucleotides or at least 15 contiguous nucleotides) is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to a region spanning the exon 6-exon 7 boundary in Transcript E. It is understood that such nucleic acid molecules would include a sufficient number of nucleotides in each of exons 6 and 7 to distinguish from other features in the *HSD17B13* Transcripts (particularly the additional

guanine at the 3' end of exon 6 in Transcript H)). For example, the region can comprise the entirety of exons 6 and 7 in Transcript E. Optionally, the isolated nucleic acid comprises or consist of a sequence at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to the sequence set forth in Transcript E and encodes an HSD17B13 protein comprising the sequence set forth in Isoform E.

As one example, the nucleic acid molecules can comprise or consist of at least 15 contiguous nucleotides (e.g., at least 20 contiguous nucleotides or at least 30 contiguous nucleotides) of Transcript E including a region within exon 3', a region spanning the exon 3-exon 3' boundary, or a region spanning the exon 3'-exon 4 boundary, optionally comprising the entire exon 3' of Transcript E, and optionally comprising the entire sequence of Transcript E.

Also provided herein are nucleic acid molecules comprising or consisting of a segment (e.g., at least 15 contiguous nucleotides) present in Transcript F (or a fragment or homolog thereof) that is not present in Transcript A (or a fragment or homolog thereof). Such regions can be readily identified by comparing the sequences of the Transcripts. Transcript F includes a read-through from exon 6 into intron 6 compared to Transcript A, and the read-through includes the inserted thymine present in the *HSD17B13* rs72613567 variant gene. Accordingly, provided herein are nucleic acid molecules comprising or consisting of at least 15 contiguous nucleotides (e.g., at least 20 contiguous nucleotides or at least 30 contiguous nucleotides) encoding all or part of an HSD17B13 protein, wherein a segment of the contiguous nucleotides (e.g., at least 5 contiguous nucleotides, at least 10 contiguous nucleotides or at least 15 contiguous nucleotides) is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to a region within the read-through into intron 6 in Transcript F or a region spanning the boundary between the read-through into intron 6 and the rest of exon 6 in Transcript F. It is understood that such nucleic acid molecules would include a sufficient number of nucleotides in the read-through to distinguish the read-through from other features in the *HSD17B13* Transcripts (e.g., from boundary of exons 6 and 7 in other *HSD17B13* Transcripts). Optionally, the contiguous nucleotides comprise a sequence present in Transcript F (i.e., the inserted thymine) that is not present in Transcript F'. Transcript F' also includes a read-through from exon 6 into intron 6 compared to Transcript A, but the read-through does not include the inserted thymine present in the *HSD17B13* rs72613567 variant gene. For example, the region can be the entire read-through into intron 6 in Transcript F. Optionally, the isolated nucleic acid molecules comprise or consist of a sequence at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to the sequence set forth in Transcript F and encodes an HSD17B13 protein comprising the sequence

set forth in Protein Isoform F.

As one example, the nucleic acid molecules can comprise or consist of at least 15 contiguous nucleotides (e.g., at least 20 contiguous nucleotides or at least 30 contiguous nucleotides) of Transcript F including a region within the read-through into intron 6 or a region spanning the boundary between the read-through into intron 6 and the rest of exon 6, optionally comprising the entire read-through into intron 6, and optionally comprising the entire sequence of Transcript F.

Also provided herein are nucleic acid molecules comprising or consisting of a segment (e.g., at least 15 contiguous nucleotides) present in Transcript F' (or a fragment or homolog thereof) that is not present in Transcript A (or a fragment or homolog thereof). Such regions can be readily identified by comparing the sequences of the Transcripts. Transcript F' includes a read-through from exon 6 into intron 6 compared to Transcript A, and the read-through does not include the inserted thymine present in the *HSD17B13* rs72613567 variant gene. Accordingly, provided herein are nucleic acid molecules comprising or consisting of at least 15 contiguous nucleotides (e.g., at least 20 contiguous nucleotides or at least 30 contiguous nucleotides) encoding all or part of an HSD17B13 protein, wherein a segment of the contiguous nucleotides (e.g., at least 5 contiguous nucleotides, at least 10 contiguous nucleotides or at least 15 contiguous nucleotides) is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to a region within the read-through into intron 6 in Transcript F' or a region spanning the boundary between the read-through into intron 6 and the rest of exon 6 in Transcript F'. It is understood that such nucleic acid molecules would include a sufficient number of nucleotides in the read-through to distinguish the read-through from other features in the *HSD17B13* Transcripts (e.g., from boundary of exons 6 and 7 in other *HSD17B13* Transcripts). Optionally, the contiguous nucleotides comprise a sequence present in Transcript F' that is not present in Transcript F. The read-through in Transcript F includes the inserted thymine present in the *HSD17B13* rs72613567 variant gene, whereas the read-through in Transcript F' does not. For example, the region can be the entire read-through into intron 6 in Transcript F'. Optionally, the isolated nucleic acid molecules comprise or consist of a sequence at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to the sequence set forth in Transcript F' and encodes an HSD17B13 protein comprising, consisting essentially of, or consisting of the sequence set forth in Isoform F'.

As one example, the nucleic acid molecules can comprise or consist of at least 15 contiguous nucleotides (e.g., at least 20 contiguous nucleotides or at least 30 contiguous

nucleotides) of Transcript F' including a region within the read-through into intron 6 or a region spanning the boundary between the read-through into intron 6 and the rest of exon 6, optionally comprising the entire read-through into intron 6, and optionally comprising the entire sequence of Transcript F'.

Also provided herein are nucleic acid molecules comprising or consisting of a segment (e.g., at least 15 contiguous nucleotides) present in Transcript C (or a fragment or homolog thereof) that is not present in Transcript A (or a fragment or homolog thereof). Such regions can be readily identified by comparing the sequences of the Transcripts. Transcript C is missing exon 6 compared to Transcript A (i.e., Transcript C includes an exon 5-exon 7 boundary not present in Transcript A). Accordingly, provided herein are nucleic acid molecules comprising or consisting of at least 15 contiguous nucleotides (e.g., at least 20 contiguous nucleotides or at least 30 contiguous nucleotides) encoding all or part of an HSD17B13 protein, wherein a segment of the contiguous nucleotides (e.g., at least 5 contiguous nucleotides, at least 10 contiguous nucleotides or at least 15 contiguous nucleotides) is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to a region spanning the exon 5-exon 7 boundary in Transcript C. It is understood that such nucleic acid molecules would include a sufficient number of nucleotides in each of exons 5 and 7 to distinguish from other features in the *HSD17B13* Transcripts (e.g., from boundary of exons 5 and 6 or of exons 6 and 7 in other *HSD17B13* Transcripts). For example, the region can comprise the entirety of exons 5 and 7 in Transcript C. Optionally, the nucleic acid molecules comprise or consist of a sequence at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to the sequence set forth in Transcript C and encodes an HSD17B13 protein comprising the sequence set forth in Isoform C.

As one example, the nucleic acid molecules can comprise or consist of at least 15 contiguous nucleotides (e.g., at least 20 contiguous nucleotides or at least 30 contiguous nucleotides) of Transcript C including a region spanning the exon 5-exon 7 boundary, optionally comprising the entirety of exons 5 and 7 in Transcript C, and optionally comprising the entire sequence of Transcript C.

In some embodiments, the nucleic acid molecule comprises less nucleotides than the entire *HSD17B13* transcript sequence. In some embodiments, the nucleic acid molecules comprise or consist of at least about 5, at least about 8, at least about 10, at least about 12, at least about 15, at least about 20, at least about 25, at least about 30, at least about 35, at least about 40, at least about 45, at least about 50, at least about 60, at least about 70, at least about 80, at least about 90, at least about 100, at least about 200, at least about 300, at least about 400, at

least about 500, or at least about 600 contiguous nucleotides of a particular transcript. In some embodiments, the nucleic acid molecules comprise or consist of at least about 200 to at least about 500 contiguous nucleotides of a particular transcript. In this regard, the longer nucleic acid molecules are preferred over the shorter ones. In some embodiments, the nucleic acid molecules comprise or consist of at least about 50, at least about 60, at least about 70, at least about 80, at least about 90, at least about 100, at least about 200, at least about 300, at least about 400, or at least about 500 contiguous nucleotides of a particular transcript. In this regard, the longer nucleic acid molecules are preferred over the shorter ones.

In some embodiments, the nucleic acid molecules can be useful, for example, as primers, and probes.

The present disclosure provides nucleic acid molecules comprising or consisting of from about 5 nucleotides up to about 50 nucleotides that specifically hybridizes to an HSD17B13 gene at a region that includes a position corresponding to position 12666 in SEQ ID NO:2, or the complement thereof, and wherein the nucleic acid molecule specifically hybridizes to an HSD17B13 gene having a thymine at a position corresponding to position 12666 of SEQ ID NO:2, or to the complement thereof.

The present disclosure provides nucleic acid molecules comprising or consisting of from about 5 nucleotides up to about 50 nucleotides that specifically hybridizes to variant HSD17B13 Transcript D, wherein the nucleic acid molecule specifically hybridizes to: i) a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a nucleotide sequence of a SEQ ID NO:6, 15, 24, or 33, or ii) to the complement of the nucleotide sequence of i). In some embodiments, the nucleic acid molecules comprise or consist of from about 5 nucleotides up to about 50 nucleotides that specifically hybridizes to variant HSD17B13 Transcript D, wherein the nucleic acid molecule specifically hybridizes to: i) a nucleotide sequence that is at least about 90% identical to a nucleotide sequence of a SEQ ID NO:6, 15, 24, or 33, or ii) to the complement of the nucleotide sequence of i).

The present disclosure provides nucleic acid molecules comprising or consisting of from about 5 nucleotides up to about 50 nucleotides comprising or consisting of: i) a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90% identical) to a nucleotide sequence of a SEQ ID NO:6, 15, 24, or 33, or to the complement thereof; ii) a nucleic acid molecule which specifically hybridizes to exon 2 of Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived

therefrom); and/or iii) a nucleic acid molecule which specifically hybridizes to the region which bridges exons 3 and 4 of Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). In some embodiments, the nucleic acid molecule specifically hybridizes to an RNA molecule comprising or consisting of SEQ ID NO:6, or a cDNA thereof comprising or consisting of SEQ ID NO:24, or wherein the nucleic acid molecule specifically hybridizes to an mRNA comprising or consisting of SEQ ID NO:15, or a cDNA thereof comprising or consisting of SEQ ID NO:33, or the complement thereof. In some embodiments, the nucleic acid molecule is linked to a heterologous nucleic acid or comprises a heterologous label.

In some embodiments, such nucleic acid molecules comprise or consist of at least about 5, at least about 8, at least about 10, at least about 11, at least about 12, at least about 13, at least about 14, at least about 15, at least about 16, at least about 17, at least about 18, at least about 19, at least about 20, at least about 21, at least about 22, at least about 23, at least about 24, at least about 25, at least about 30, at least about 35, at least about 40, at least about 45, at least about 50, at least about 55, at least about 60, at least about 65, at least about 70, at least about 75, at least about 80, at least about 85, at least about 90, at least about 95, at least about 100, at least about 200, at least about 300, at least about 400, at least about 500, at least about 600, at least about 700, at least about 800, at least about 900, at least about 1000, at least about 2000, at least about 3000, at least about 4000, at least about 5000, at least about 6000, at least about 7000, at least about 8000, at least about 9000, at least about 10000, at least about 11000, or at least about 11500. In some embodiments, the nucleic acid molecule comprises or consists of at least 15 nucleotides. In some embodiments, the nucleic acid molecule comprises or consists of at least 15 nucleotides to at least about 35 nucleotides. In some embodiments, such nucleic acid molecules hybridize to variant *HSD17B13* genomic DNA, variant *HSD17B13* minigenes, variant *HSD17B13* RNA (or cDNA derived therefrom), or variant *HSD17B13* mRNA (or cDNA derived therefrom) under stringent conditions. Such nucleic acid molecules may be used, for example, as probes, as primers, or as alteration-specific probes or primers as described or exemplified herein.

Also disclosed herein are nucleic acid molecules comprising or consisting of at least 15 contiguous nucleotides that hybridize to an *HSD17B13* gene (e.g., an *HSD17B13* minigene) at a segment that includes or is within 1000, 500, 400, 300, 200, 100, 50, 45, 40, 35, 30, 25, 20, 15, 10, or 5 nucleotides of a position corresponding to position 12666 or positions 12666 and 12667 of the *HSD17B13* rs72613567 variant (SEQ ID NO:2). Such nucleic acid molecules can be useful, for example, as primers or probes.

In some embodiments, the at least 15 contiguous nucleotides can hybridize to a segment

of the *HSD17B13* gene or *HSD17B13* minigene that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% (or at least about 90% identical) identical to a corresponding sequence in the *HSD17B13* rs72613567 variant (SEQ ID NO:2), and having a thymine at a position corresponding to position 12666 of SEQ ID NO:2. In some embodiments, the nucleic acid molecule can hybridize to at least 15 contiguous nucleotides of SEQ ID NO:2. In some embodiments, the nucleic acid molecule hybridizes to a segment including position 12666 or positions 12666 and 12667 in SEQ ID NO:2 or a position corresponding to position 12666 or positions 12666 and 12667 in SEQ ID NO:2.

In some embodiments, the segment to which the nucleic acid molecules can hybridize can comprise, for example, at least 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 75, 90, 95, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, or 2000 contiguous nucleotides of nucleic acid molecule encoding a variant HSD17B13 protein isoform. In some embodiments, the segment to which the nucleic acid molecules can hybridize can be, for example, up to 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 75, 90, 95, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 contiguous nucleotides of a nucleic acid encoding a variant HSD17B13 protein isoform. In some embodiments, the nucleic acid molecules can comprise, for example, at least 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 11000, 12000, 13000, 14000, 15000, 16000, 17000, 18000, or 19000 contiguous nucleotides of a variant *HSD17B13* gene. In some embodiments, the segment to which the nucleic acid molecules can hybridize can be, for example, up to 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 75, 90, 95, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 contiguous nucleotides of a variant *HSD17B13* gene. In some embodiments, the segment can be about 15 to 100 nucleotides in length, or about 15 to 35 nucleotides in length.

Also provided are nucleic acid molecules hybridizing to segments of an RNA transcript, such as Transcript A, Transcript B, Transcript C, Transcript D, Transcript E, Transcript F, Transcript F', Transcript G, and Transcript H, or a corresponding cDNA thereof, or an mRNA transcript, such as Transcript A, Transcript B, Transcript C, Transcript D, Transcript E, Transcript F, Transcript F', Transcript G, and Transcript H, or a corresponding cDNA thereof.

Also provided are nucleic acid molecules hybridizing to segments of an RNA transcript, such as Transcript C, Transcript D, Transcript E, Transcript F, Transcript F', Transcript G, and Transcript H, or a corresponding cDNA thereof, or an mRNA transcript, such as Transcript C, Transcript D, Transcript E, Transcript F, Transcript F', Transcript G, and Transcript H, or a corresponding cDNA thereof.

Also provided are nucleic acid molecules hybridizing to segments of an RNA transcript,

such as Transcript C, Transcript D, Transcript F, Transcript G, and Transcript H, or a corresponding cDNA thereof, or an mRNA transcript, such as Transcript C, Transcript D, Transcript F, Transcript G, and Transcript H, or a corresponding cDNA thereof.

Also provided are nucleic acid molecules ybridizing to segments of RNA Transcript D, or a corresponding cDNA, or an mRNA Transcript D, or a corresponding cDNA.

Provided herein are nucleic acid molecules comprising or consisting of a region (e.g., at least 15 contiguous nucleotides) that hybridizes to a segment present in Transcripts D, G, and H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or fragments or homologs thereof, that is not present in Transcript A (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof. Such regions can be readily identified by comparing the sequences of the Transcripts. For example, provided herein are nucleic acid molecules that hybridize to at least 15 contiguous nucleotides of a nucleic acid encoding an HSD17B13 protein, wherein contiguous nucleotides comprise or consist of a segment (e.g., at least 5 contiguous nucleotides, at least 10 contiguous nucleotides or at least 15 contiguous nucleotides) that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90% identical) to a region spanning the exon 6–exon 7 boundary in Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), and the segment includes a guanine at a residue corresponding to residue 878 at the 3' end of exon 6 in Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) (i.e., an insertion of a guanine at the 3' end of exon 6 relative to Transcript A in addition to the guanine at the start of exon 7). Alternatively, provided herein are nucleic acid molecules that hybridize to at least 15 contiguous nucleotides of a segment of a nucleic acid encoding an HSD17B13 protein, wherein the contiguous nucleotides comprise or consist of a segment (e.g., at least 5 contiguous nucleotides, at least 10 contiguous nucleotides or at least 15 contiguous nucleotides) that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90% identical) to a region spanning the exon 6-exon 7 boundary in Transcript G, and the segment includes a guanine at a residue corresponding to residue 770 at the 3' end of exon 6 in Transcript G (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) (i.e., an insertion of a guanine at the 3' end of exon 6 relative to Transcript B in addition to the guanine at the start of exon 7). Alternatively, provided herein are nucleic acid molecules that hybridize to

at least 15 contiguous nucleotides of a nucleic acid encoding an HSD17B13 protein, wherein the contiguous nucleotides comprise or consist of a segment (e.g., at least 5 contiguous nucleotides, at least 10 contiguous nucleotides or at least 15 contiguous nucleotides) that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90% identical) to a region spanning the exon 6-exon 7 boundary in Transcript H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), and the segment includes a guanine at a residue corresponding to residue 950 at the 3' end of exon 6 in Transcript H (i.e., an insertion of a guanine at the 3' end of exon 6 relative to Transcript E in addition to the guanine at the start of exon 7). It is understood that such nucleic acid molecules would be designed to hybridize to a sufficient number of nucleotides in each of exons 6 and 7 to distinguish the inserted guanine from other features in the *HSD17B13* Transcripts (e.g., from the read-through into intron 6 in Transcript F or from the deleted exon 6 in Transcript C).

As one example, the segment can comprise or consist of a region of Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) spanning the exon 6-exon 7 boundary (i.e., including the guanine at residue 878 of Transcript D). As another example, the segment can comprise or consist of a region of Transcript G (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) spanning the exon 6-exon 7 boundary (i.e., including the guanine at residue 770 of Transcript G). As another example, the segment can comprise or consist of a region of Transcript H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) spanning the exon 6-exon 7 boundary (i.e., including the guanine at residue 950 of Transcript H).

In some embodiments, the nucleic acid molecules further comprise or consist of a region (e.g., 15 contiguous nucleotides) that hybridizes to a segment present in Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof, that is not present in Transcript G (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof, and the nucleic acid molecules further comprise or consist of a region that hybridizes to a segment present in Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof, that is not present in Transcript H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof.

Such segments can be readily identified by comparing the sequences of the Transcripts. For example, the segment (e.g., at least 5 contiguous nucleotides, at least 10 contiguous nucleotides or at least 15 contiguous nucleotides) present in Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof, that is not present in Transcript H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof, can be at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical (or at least about 90% identical) to a region spanning the boundary of exons 3 and 4 of Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) to distinguish from Transcript H. Likewise, the segment (e.g., at least 5 contiguous nucleotides, at least 10 contiguous nucleotides or at least 15 contiguous nucleotides) present in Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof, that is not present in Transcript G (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof, can be at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical (or at least about 90% identical) to a region within exon 2 of Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), a region spanning the exon 1-exon 2 boundary of Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a region spanning the exon 2-exon 3 boundary of Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) to distinguish from Transcript G.

Provided herein are nucleic acid molecules as described above comprising or consisting of a region that hybridizes to a segment present in Transcripts D, G, and H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or fragments or homologs thereof, that is not present in Transcript A (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof, but further comprising a region (e.g., at least 15 contiguous nucleotides) that hybridizes to a segment that is present in Transcript H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof, but not in Transcript D (RNA or cDNA

derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof. Such regions can be readily identified by comparing the sequences of the Transcripts. For example, the segment can be at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90% identical) to a region (e.g., at least 5 contiguous nucleotides, at least 10 contiguous nucleotides or at least 15 contiguous nucleotides) within exon 3' of Transcript H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), a region spanning the exon 3-exon 3' boundary of Transcript H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a region spanning the exon 3'-exon 4 boundary of Transcript H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). It is understood that such nucleic acid molecules would be designed to hybridize to a sufficient number of nucleotides in each of exons 3 and 3' or each of exons 3' and 4 to distinguish from other features in the *HSD17B13* transcripts (e.g., from the boundary of exons 3 and 4). As one example, the segment can comprise or consist of a region of Transcript H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) within exon 3', spanning the exon 3-exon 3' boundary, or spanning the exon 3'-exon 4 boundary.

Provided herein are nucleic acid molecules as described above comprising or consisting of a region that hybridizes to a segment present in Transcripts D, G, and H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or fragments or homologs thereof, that is not present in Transcript A (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof, but further comprising a region (e.g., at least 15 contiguous nucleotides) that hybridizes to a segment present in Transcript G (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof, but not in Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof. Such regions can be readily identified by comparing the sequences of the Transcripts. For example, the segment can be at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90% identical) to a region (e.g., at least 5 contiguous nucleotides, at least 10 contiguous nucleotides or at least 15 contiguous nucleotides) spanning the exon 1-exon 3 boundary in Transcript G (RNA or cDNA derived therefrom, and/or

mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). It is understood that such nucleic acid molecules would be designed to hybridize to a sufficient number of nucleotides in each of exons 1 and 3 to distinguish from other features in the *HSD17B13* Transcripts (e.g., the boundary of exons 1 and 2 or the boundary of exons 2 and 3). As one example, the segment can comprise or consist of a region of Transcript G (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) spanning the exon 1-exon 3 boundary.

Also provided are nucleic acid molecules comprising or consisting of a region (e.g., at least 15 contiguous nucleotides) that hybridizes to a segment of a nucleic acid encoding an HSD17B13 protein that is present in Transcript E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof, but not in Transcript A (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof. Such regions can be readily identified by comparing the sequences of the Transcripts. Transcript E includes an additional exon between exons 3 and 4 compared to Transcript A. Accordingly, provided herein are nucleic acid molecules that hybridize to at least 15 contiguous nucleotides of a nucleic acid encoding an HSD17B13 protein, wherein the contiguous nucleotides comprise or consist of a segment that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90% identical) to a region (e.g., at least 5 contiguous nucleotides, at least 10 contiguous nucleotides or at least 15 contiguous nucleotides) within exon 3' of Transcript E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), a region spanning the exon 3-exon 3' boundary of Transcript E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a region spanning the exon 3'-exon 4 boundary of Transcript E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). It is understood that such nucleic acid molecules would be designed to hybridize to a sufficient number of nucleotides in each of exons 3 and 3' or each of exons 3' and 4 to distinguish from other features in the *HSD17B13* transcripts (e.g., from the boundary of exons 3 and 4). As one example, the segment can comprise or consist of a region of Transcript E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) within exon 3', spanning the exon 3-exon 3' boundary of Transcript E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or spanning the exon 3'-exon 4

boundary (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom).

In some embodiments, the nucleic acid molecules further comprise or consist of a region (e.g., 15 contiguous nucleotides) that hybridizes to a segment present in Transcript E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof, that is not present in Transcript H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof. Such segments can be readily identified by comparing the sequences of the Transcripts. For example, the segment (e.g., at least 5 contiguous nucleotides, at least 10 contiguous nucleotides or at least 15 contiguous nucleotides) present in Transcript E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof, that is not present in Transcript H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof, can be at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical (or at least about 90% identical) to a region spanning the boundary of exons 6 and 7 of Transcript E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) to distinguish from Transcript G. It is understood that such nucleic acid molecules would be designed to hybridize to a sufficient number of nucleotides in each of exons 6 and 7 to distinguish from other features in the *HSD17B13* Transcripts (particularly the additional guanine at the 3' end of exon 6 in Transcript H).

Also provided are nucleic acid molecules comprising or consisting of a region (e.g., at least 15 contiguous nucleotides) that hybridizes to a segment of a nucleic acid encoding an HSD17B13 protein that is present in Transcript F (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof, but not in Transcript A (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof. Such regions can be readily identified by comparing the sequences of the Transcripts. Transcript F includes a read-through from exon 6 to intron 6 compared to Transcript A. Accordingly, provided herein are nucleic acid molecules that hybridize to at least 15 contiguous nucleotides of a nucleic acid encoding an HSD17B13 protein, wherein the contiguous nucleotides comprise or consist of a segment (e.g., at least 5 contiguous nucleotides, at least 10 contiguous nucleotides or at least 15 contiguous nucleotides) that is at least about

90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90% identical) to a region within the read-through into intron 6 in Transcript F (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or a region spanning the boundary between the read-through into intron 6 and the rest of exon 6 in Transcript F (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). It is understood that such nucleic acid molecules would be designed to hybridize to a sufficient number of nucleotides in the read-through to distinguish the read-through from other features in the *HSD17B13* Transcripts (e.g., from boundary of exons 6 and 7 in other *HSD17B13* Transcripts). Optionally, the contiguous nucleotides comprise or consist of a sequence present in Transcript F (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) (i.e., the inserted thymine) that is not present in Transcript F' (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). Transcript F' also includes a read-through from exon 6 into intron 6 compared to Transcript A, but the read-through does not include the inserted thymine present in the *HSD17B13* rs72613567 variant gene. As one example, the segment can comprise or consist of a region of Transcript F (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) within the read-through into intron 6 or spanning the boundary between the read-through into intron 6 and the rest of exon 6.

Also provided are nucleic acid molecules comprising or consisting of a region (e.g., at least 15 contiguous nucleotides) that hybridizes to a segment of a nucleic acid encoding an HSD17B13 protein that is present in Transcript F' (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof, but not in Transcript A (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof. Such regions can be readily identified by comparing the sequences of the Transcripts. Transcript F' includes a read-through from exon 6 to intron 6 compared to Transcript A. Accordingly, provided herein are nucleic acid molecules that hybridize to at least 15 contiguous nucleotides of a nucleic acid encoding an HSD17B13 protein, wherein the contiguous nucleotides comprise or consist of a segment (e.g., at least 5 contiguous nucleotides, at least 10 contiguous nucleotides or at least 15 contiguous nucleotides) that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90% identical) to a region within the read-

through into intron 6 in Transcript F' (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or a region spanning the boundary between the read-through into intron 6 and the rest of exon 6 in Transcript F' (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). It is understood that such nucleic acid molecules would be designed to hybridize to a sufficient number of nucleotides in the read-through to distinguish the read-through from other features in the *HSD17B13* Transcripts (e.g., from boundary of exons 6 and 7 in other *HSD17B13* Transcripts). Optionally, the contiguous nucleotides comprise or consist of a sequence present in Transcript F' (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) that is not present in Transcript F (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). The read-through in Transcript F includes the inserted thymine present in the *HSD17B13* rs72613567 variant gene, whereas the read-through in Transcript F' does not. As one example, the segment can comprise or consist of a region of Transcript F' (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) within the read-through into intron 6 or spanning the boundary between the read-through into intron 6 and the rest of exon 6.

Also provided are nucleic acid molecules comprising or consisting of a region (e.g., at least 15 contiguous nucleotides) that hybridizes to a segment of a nucleic acid encoding an HSD17B13 protein that is present in Transcript C (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof, but not in Transcript A (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a fragment or homolog thereof. Such regions can be readily identified by comparing the sequences of the Transcripts. Transcript C is missing exon 6 compared to Transcript A (i.e., Transcript C includes an exon 5-exon 7 boundary not present in Transcript A). Accordingly, provided herein are nucleic acid molecules that hybridize to at least 15 contiguous nucleotides of a nucleic acid encoding an HSD17B13 protein, wherein the contiguous nucleotides comprise or consist of a segment (e.g., at least 5 contiguous nucleotides, at least 10 contiguous nucleotides or at least 15 contiguous nucleotides) that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90% identical) to a region spanning the exon 5-exon 7 boundary in Transcript C (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). It is understood that such nucleic acid molecules would be designed to hybridize to a

sufficient number of nucleotides in exons 5 and 7 to distinguish from other features in the *HSD17B13* Transcripts (e.g., from boundary of exons 5 and 6 or of exons 6 and 7 in other *HSD17B13* Transcripts). As one example, the segment can comprise or consist of a region from Transcript C (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) spanning the exon 5-exon 7 boundary.

The present disclosure also provides supports comprising a substrate to which any one or more of the probes disclosed herein is attached. Solid supports are solid-state substrates or supports with which molecules, such as any of the probes disclosed herein, can be associated. A form of solid support is an array. Another form of solid support is an array detector. An array detector is a solid support to which multiple different probes have been coupled in an array, grid, or other organized pattern.

Solid-state substrates for use in solid supports can include any solid material to which molecules can be coupled. This includes materials such as acrylamide, agarose, cellulose, nitrocellulose, glass, polystyrene, polyethylene vinyl acetate, polypropylene, polymethacrylate, polyethylene, polyethylene oxide, polysilicates, polycarbonates, teflon, fluorocarbons, nylon, silicon rubber, polyanhydrides, polyglycolic acid, polylactic acid, polyorthoesters, polypropylfumerate, collagen, glycosaminoglycans, and polyamino acids. Solid-state substrates can have any useful form including thin film, membrane, bottles, dishes, fibers, woven fibers, shaped polymers, particles, beads, microparticles, or a combination. Solid-state substrates and solid supports can be porous or non-porous. A form for a solid-state substrate is a microtiter dish, such as a standard 96-well type. In some embodiments, a multiwell glass slide can be employed that normally contain one array per well. In some embodiments, the support is a microarray.

The nucleic acid molecules disclosed herein can comprise RNA, DNA, or both RNA and DNA. The nucleic acid molecules can also be linked or fused to a heterologous nucleotide sequence, such as in a vector, or a heterologous label. For example, the nucleic acid molecules disclosed herein can be in a vector or exogenous donor sequence comprising the nucleic acid molecule and a heterologous nucleotide sequence. The nucleic acid molecules can also be linked or fused to a heterologous label, such as a fluorescent label. Other examples of labels are disclosed elsewhere herein.

The label can be directly detectable (*e.g.*, fluorophore) or indirectly detectable (*e.g.*, hapten, enzyme, or fluorophore quencher). Such labels can be detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. Such labels include, for example, radiolabels that can be measured with radiation-counting devices; pigments, dyes or

other chromogens that can be visually observed or measured with a spectrophotometer; spin labels that can be measured with a spin label analyzer; and fluorescent labels (*e.g.*, fluorophores), where the output signal is generated by the excitation of a suitable molecular adduct and that can be visualized by excitation with light that is absorbed by the dye or can be measured with standard fluorometers or imaging systems. The label can also be, for example, a chemiluminescent substance, where the output signal is generated by chemical modification of the signal compound; a metal-containing substance; or an enzyme, where there occurs an enzyme-dependent secondary generation of signal, such as the formation of a colored product from a colorless substrate. The term "label" can also refer to a "tag" or hapten that can bind selectively to a conjugated molecule such that the conjugated molecule, when added subsequently along with a substrate, is used to generate a detectable signal. For example, one can use biotin as a tag and then use an avidin or streptavidin conjugate of horseradish peroxidate (HRP) to bind to the tag, and then use a calorimetric substrate (*e.g.*, tetramethylbenzidine (TMB)) or a fluorogenic substrate to detect the presence of HRP. Exemplary labels that can be used as tags to facilitate purification include, but are not limited to, myc, HA, FLAG or 3XFLAG, 6XHis or polyhistidine, glutathione-S-transferase (GST), maltose binding protein, an epitope tag, or the Fc portion of immunoglobulin. Numerous labels are known and include, for example, particles, fluorophores, haptens, enzymes and their calorimetric, fluorogenic and chemiluminescent substrates and other labels.

The nucleic acid molecules can be modified nucleic acid molecules and comprise, for example, nucleotides or non-natural or modified nucleotides, such as nucleotide analogs or nucleotide substitutes. Such nucleotides include a nucleotide that contains a modified base, sugar, or phosphate group, or that incorporates a non-natural moiety in its structure. Examples of non-natural nucleotides include, but are not limited to, dideoxynucleotides, biotinylated, aminated, deaminated, alkylated, benzylated, and fluorophor-labeled nucleotides.

The nucleic acid molecules disclosed herein can also comprise one or more nucleotide analogs or substitutions. A nucleotide analog is a nucleotide which contains a modification to either the base, sugar, or phosphate moieties. Modifications to the base moiety include, but are not limited to, natural and synthetic modifications of A, C, G, and T/U, as well as different purine or pyrimidine bases such as, for example, pseudouridine, uracil-5-yl, hypoxanthin-9-yl (I), and 2-aminoadenin-9-yl. Modified bases include, but are not limited to, 5-methylcytosine (5-me-C), 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiouracil, 2-thiothymine and 2-thiocytosine, 5-halouracil and cytosine, 5-propynyl

uracil and cytosine, 6-azo uracil, cytosine and thymine, 5-uracil (pseudouracil), 4-thiouracil, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxyl and other 8-substituted adenines and guanines, 5-halo particularly 5-bromo, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine and 7-methyladenine, 8-azaguanine and 8-azaadenine, 7-deazaguanine and 7-deazaadenine and 3-deazaguanine and 3-deazaadenine. Certain nucleotide analogs such as, for example, 5-substituted pyrimidines, 6-azapyrimidines, and N-2, N-6 and O-6 substituted purines including, but not limited to, 2-aminopropyladenine, 5-propynyluracil, 5-propynylcytosine, and 5-methylcytosine can increase the stability of duplex formation. Often, base modifications can be combined with, for example, a sugar modification, such as 2'-O-methoxyethyl, to achieve unique properties such as increased duplex stability.

Nucleotide analogs can also include modifications of the sugar moiety. Modifications to the sugar moiety include, but are not limited to, natural modifications of the ribose and deoxy ribose as well as synthetic modifications. Sugar modifications include, but are not limited to, the following modifications at the 2' position: OH; F; O-, S-, or N-alkyl; O-, S-, or N-alkenyl; O-, S- or N-alkynyl; or O-alkyl-O-alkyl, wherein the alkyl, alkenyl, and alkynyl may be substituted or unsubstituted C_{1-10}alkyl or C_{2-10}alkenyl, and C_{2-10}alkynyl. Exemplary 2' sugar modifications also include, but are not limited to, $-O[(CH_2)_nO]_mCH_3$, $-O(CH_2)_nOCH_3$, $-O(CH_2)_nNH_2$, $-O(CH_2)_nCH_3$, $-O(CH_2)_n-ONH_2$, and $-O(CH_2)_nON[(CH_2)_nCH_3)]_2$, where n and m are from 1 to about 10.

Other modifications at the 2' position include, but are not limited to, C_{1-10}alkyl, substituted lower alkyl, alkaryl, aralkyl, O-alkaryl or O-aralkyl, SH, SCH_3, OCN, Cl, Br, CN, CF_3, OCF_3, $SOCH_3$, SO_2CH_3, ONO_2, NO_2, N_3, NH_2, heterocycloalkyl, heterocycloalkaryl, aminoalkylamino, polyalkylamino, substituted silyl, an RNA cleaving group, a reporter group, an intercalator, a group for improving the pharmacokinetic properties of an oligonucleotide, or a group for improving the pharmacodynamic properties of an oligonucleotide, and other substituents having similar properties. Similar modifications may also be made at other positions on the sugar, particularly the 3' position of the sugar on the 3' terminal nucleotide or in 2'-5' linked oligonucleotides and the 5' position of 5' terminal nucleotide. Modified sugars can also include those that contain modifications at the bridging ring oxygen, such as CH_2 and S. Nucleotide sugar analogs can also have sugar mimetics, such as cyclobutyl moieties in place of the pentofuranosyl sugar.

Nucleotide analogs can also be modified at the phosphate moiety. Modified phosphate moieties include, but are not limited to, those that can be modified so that the linkage between two nucleotides contains a phosphorothioate, chiral phosphorothioate, phosphorodithioate, phosphotriester, aminoalkylphosphotriester, methyl and other alkyl phosphonates including 3'-

alkylene phosphonate and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates. These phosphate or modified phosphate linkage between two nucleotides can be through a 3'-5' linkage or a 2'-5' linkage, and the linkage can contain inverted polarity such as 3'-5' to 5'-3' or 2'-5' to 5'-2'. Various salts, mixed salts, and free acid forms are also included.

Nucleotide substitutes also include nucleotides or nucleotide analogs that have had the phosphate moiety or sugar moieties replaced. In some embodiments, nucleotide substitutes may not contain a standard phosphorus atom. Substitutes for the phosphate can be, for example, short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S, and CH_2 component parts.

It is also understood in a nucleotide substitute that both the sugar and the phosphate moieties of the nucleotide can be replaced by, for example, an amide type linkage (aminoethylglycine) (PNA).

It is also possible to link other types of molecules (conjugates) to nucleotides or nucleotide analogs to enhance, for example, cellular uptake. Conjugates can be chemically linked to the nucleotide or nucleotide analogs. Such conjugates include, for example, lipid moieties such as a cholesterol moiety, cholic acid, a thioether such as hexyl-S-tritylthiol, a thiocholesterol, an aliphatic chain such as dodecandiol or undecyl residues, a phospholipid such as di-hexadecyl-rac-glycerol or triethylammonium 1,2-di-O-hexadecyl-rac-glycero-3-H-phosphonate, a polyamine or a polyethylene glycol chain, adamantane acetic acid, a palmityl moiety, or an octadecylamine or hexylamino-carbonyl-oxycholesterol moiety.

Also disclosed herein are polypeptides encoded by the nucleic acid molecules disclosed herein and compositions comprising a nucleic acid or polypeptide disclosed herein and a carrier increasing the stability of the isolated nucleic acid or protein (e.g., prolonging the period under given conditions of storage (e.g., -,20°C, 4°C, or ambient temperature) for which degradation products remain below a threshold, such below 0.5% by weight of the starting nucleic acid or protein; or increasing the stability *in vivo*). Non-limiting examples of such carriers include

poly(lactic acid) (PLA) microspheres, poly(D,L-lactic-coglycolic-acid) (PLGA) microspheres, liposomes, micelles, inverse micelles, lipid cochleates, and lipid microtubules.

Also provided herein are functional polynucleotides that can interact with the disclosed nucleic acid molecules. Functional polynucleotides are nucleic acid molecules that have a specific function, such as binding a target molecule or catalyzing a specific reaction. Examples of functional polynucleotides include, but are not limited to, antisense molecules, aptamers, ribozymes, triplex forming molecules, and external guide sequences. The functional polynucleotides can act as effectors, inhibitors, modulators, and stimulators of a specific activity possessed by a target molecule, or the functional polynucleotides can possess a *de novo* activity independent of any other molecules.

Antisense molecules are designed to interact with a target nucleic acid molecule through either canonical or non-canonical base pairing. The interaction of the antisense molecule and the target molecule is designed to promote the destruction of the target molecule through, for example, RNase-H-mediated RNA-DNA hybrid degradation. Alternately, the antisense molecule is designed to interrupt a processing function that normally would take place on the target molecule, such as transcription or replication. Antisense molecules can be designed based on the sequence of the target molecule. Numerous methods for optimization of antisense efficiency by identifying the most accessible regions of the target molecule exist. Exemplary methods include, but are not limited to, *in vitro* selection experiments and DNA modification studies using DMS and DEPC. Antisense molecules generally bind the target molecule with a dissociation constant (k_d) less than or equal to about 10^{-6}, less than or equal to about 10^{-8}, less than or equal to about 10^{-10}, or less than or equal to about 10^{-12}. Examples of antisense molecules include, but are not limited to, antisense RNAs, small interfering RNAs (siRNAs), and short hairpin RNAs (shRNAs).

In some embodiments, any of the nucleic acid molecules, genomic DNA molecules, minigenes, RNA molecules, mRNA molecules, or cDNA molecules disclosed herein can be purified, e.g., are at least about 90% pure. In some embodiments, any of the nucleic acid molecules disclosed herein can be purified, e.g., are at least about 95% pure. In some embodiments, any of the nucleic acid molecules disclosed herein can be purified, e.g., are at least about 99% pure. Purification is according to the hands of a human, with human-made purification techniques.

The present disclosure also provides vectors comprising any one or more of the nucleic acid molecules disclosed herein. In some embodiments, the vectors comprise any one or more of the nucleic acid molecules disclosed herein and a heterologous nucleic acid. The vectors can be

viral or nonviral vectors capable of transporting a nucleic acid molecule. In some embodiments, the vector is a plasmid or cosmid. In some embodiments, the vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. In some embodiments, the vector can autonomously replicate in a host cell into which it is introduced. In some embodiments, the vector can be integrated into the genome of a host cell upon introduction into the host cell and thereby are replicated along with the host genome. Moreover, particular vectors can direct the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" or "expression vectors." Such vectors can also be targeting vectors.

In some embodiments, the proteins encoded by the various genetic variants disclosed herein are expressed by inserting nucleic acid molecules encoding the disclosed genetic variants into expression vectors, such that the genes are operatively linked to expression control sequences, such as transcriptional and translational control sequences. Expression vectors include, but are not limited to, plasmids, cosmids, retroviruses, adenoviruses, adeno-associated viruses (AAV), plant viruses such as cauliflower mosaic virus and tobacco mosaic virus, yeast artificial chromosomes (YACs), Epstein-Barr (EBV)-derived episomes, and other expression vectors known in the art. In some embodiments, nucleic acid molecules comprising the disclosed genetic variants can be ligated into a vector such that transcriptional and translational control sequences within the vector serve their intended function of regulating the transcription and translation of the genetic variant.

In addition to a nucleotide sequence comprising the disclosed genetic variants, the recombinant expression vectors can carry regulatory sequences that control the expression of the genetic variant in a host cell such as, for example, promoters and/or enhancers derived from retroviral LTRs, cytomegalovirus (CMV) (such as the CMV promoter/enhancer), Simian Virus 40 (SV40) (such as the SV40 promoter/enhancer), adenovirus, (*e.g.*, the adenovirus major late promoter (AdMLP)), polyoma and strong mammalian promoters such as native immunoglobulin and actin promoters. Methods of expressing polypeptides in bacterial cells or fungal cells (*e.g.*, yeast cells) are also well known.

A promoter can be, for example, a constitutively active promoter, a conditional promoter, an inducible promoter, a temporally restricted promoter (*e.g.*, a developmentally regulated promoter), or a spatially restricted promoter (*e.g.*, a cell-specific or tissue-specific promoter).

In addition to a nucleotide sequence comprising the disclosed genetic variants and regulatory sequences, the recombinant expression vectors can carry additional sequences, such as

sequences that regulate replication of the vector in host cells and selectable marker genes. Exemplary selectable marker genes include, but are not limited to, the dihydrofolate reductase (DHFR) gene (for use in dhfr-host cells with methotrexate selection / amplification), the neo gene (for G418 selection), and the glutamate synthetase (GS) gene.

The present disclosure also provides cells (*e.g.*, recombinant host cells) comprising any one or more of the nucleic acid molecules, including vectors comprising the nucleic acid molecules, and/or any one or more of the polypeptides disclosed herein. The cells can be *in vitro*, *ex vivo*, or *in vivo*. Nucleic acid molecules can be linked to a promoter and other regulatory sequences so they are expressed to produce an encoded protein. Cell lines of such cells are further provided.

In some embodiments, the cell is a totipotent cell or a pluripotent cell (*e.g.*, an embryonic stem (ES) cell such as a rodent ES cell, a mouse ES cell, or a rat ES cell). Pluripotent and/or totipotent cells can be, for example, ES cells or ES-like cells, such as an induced pluripotent stem (iPS) cells. In accordance with the present disclosure, the embryonic stem cells may be non-human embryonic stem cells. In some embodiments, the cell is a primary somatic cell, or a cell that is not a primary somatic cell. Such cells can be isolated by conventional techniques and include, for example, somatic cells, hematopoietic cells, endothelial cells, epithelial cells, fibroblasts, mesenchymal cells, keratinocytes, melanocytes, monocytes, mononuclear cells, adipocytes, preadipocytes, neurons, glial cells, hepatocytes, skeletal myoblasts, and smooth muscle cells. For example, primary cells can be derived from connective tissues, muscle tissues, nervous system tissues, or epithelial tissues.

In some embodiments, the cells may normally not proliferate indefinitely but, due to mutation or alteration, have evaded normal cellular senescence and instead can keep undergoing division. Such mutations or alterations can occur naturally or be intentionally induced. Examples of immortalized cells include, but are not limited to, Chinese hamster ovary (CHO) cells, human embryonic kidney cells (*e.g.*, HEK 293 cells), and mouse embryonic fibroblast cells (*e.g.*, 3T3 cells). In some embodiments, the cell is a differentiated cell, such as a liver cell (*e.g.*, a human liver cell).

The cell can be from any source. For example, the cell can be a eukaryotic cell, an animal cell, a plant cell, or a fungal (*e.g.*, yeast) cell. Such cells can be fish cells or bird cells, or such cells can be mammalian cells, such as human cells, non-human mammalian cells, rodent cells, mouse cells or rat cells. Mammals include, but are not limited to, humans, non-human primates, monkeys, apes, cats dogs, horses, bulls, deer, bison, sheep, rodents (*e.g.*, mice, rats, hamsters, guinea pigs), livestock (*e.g.*, bovine species such as cows, steer, etc.; ovine species

such as sheep, goats, etc.; and porcine species such as pigs and boars). Birds include, but are not limited to, chickens, turkeys, ostrich, geese, ducks, etc. Domesticated animals and agricultural animals are also included. The term "non-human animal" excludes humans. In some embodiments, the cell is a human cell.

The present disclosure provides use of any of the nucleic acid molecules described herein as probes or primers for detecting a variant HSD17B13 gene or variant HSD17B13 transcript, for determining a human subject's susceptibility or risk of developing a liver disease, or for determining a human subject's risk for progression to more clinically advanced stages of fatty liver disease.

The present disclosure provides *HSD17B13* isoform polypeptides and fragments thereof, and particularly *HSD17B13* isoform polypeptides and fragments thereof produced by the *HSD17B13* rs72613567 variant.

The present disclosure provides polypeptides comprising or consisting of an amino acid sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to the amino acid sequence of HSD17B13 Isoform D (SEQ ID NO:42). In some embodiments, the polypeptides comprise or consist of an amino acid sequence that is at least about 90% identical to the amino acid sequence of HSD17B13 Isoform D (SEQ ID NO:42). In some embodiments, the polypeptide comprises or consists of the amino acid sequence of SEQ ID NO:42. In some embodiments, these polypeptides have 274 amino acids. In some embodiments, these polypeptides have a C-terminal Val-Ser-Ser. In some embodiments, these polypeptides are associated with a decreased risk of developing any of the liver diseases described herein or decreased risk of progression to more clinically advanced stages of fatty liver disease.

In some embodiments, the polypeptides comprise or consist of an amino acid sequence at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90%) to Isoform A, Isoform B, Isoform C, Isoform D, Isoform E, Isoform F, Isoform F', Isoform G, or Isoform H. In some embodiments, the HSD17B13 protein is Isoform A, Isoform B, Isoform C, Isoform D, Isoform E, Isoform F, Isoform F', Isoform G, or Isoform H.

In some embodiments, the polypeptides comprise or consist of an amino acid sequence at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90%) to Isoform C, Isoform D, Isoform E, Isoform F, Isoform F', Isoform G, or Isoform H. In some embodiments, the polypeptide is Isoform C, Isoform D, Isoform E, Isoform F, Isoform F', Isoform G, or Isoform

H.

In some embodiments, the polypeptides comprise or consist of an amino acid sequence at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90%) to Isoform C, Isoform D, Isoform F, Isoform G, or Isoform H. In some embodiments, the polypeptide is Isoform C, Isoform D, Isoform F, Isoform G, or Isoform H.

In some embodiments, the polypeptides comprise or consist of an amino acid sequence at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to Isoform D. In some embodiments, the polypeptides comprise or consist of an amino acid sequence at least about 90% identical to Isoform D. In some embodiments, the polypeptide is Isoform D.

In some embodiments, the polypeptides comprise or consist of, for example, at least 5, 6, 8, 10, 12, 14, 15, 16, 18, 20, 22, 24, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, or 300 contiguous amino acids of: i) HSD17B13 Isoform A, B, C, D, E, F, F', G, or H or a fragment thereof; ii) HSD17B13 Isoform C, D, E, F, F', G, or H or a fragment thereof; iii) HSD17B13 Isoform C, D, F, G, or H or a fragment thereof; or iv) HSD17B13 Isoform D or a fragment thereof. It is understood that gene sequences within a population and proteins encoded by such genes can vary due to polymorphisms such as single-nucleotide polymorphisms. The sequences provided herein for each HSD17B13 isoform are only exemplary sequences. Other sequences are also possible.

As one example, the polypeptides can comprise or consist of a segment (e.g., at least 8 contiguous amino acids) that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90%) to a segment including at least a portion of the region encoded by exon 7 in Isoform D, Isoform G, or Isoform H, (or fragments or homologs thereof) that is not present in Isoform A (or a fragment or homolog thereof). Such regions can be readily identified by comparing the sequences of the Isoforms. The region encoded by exon 7 in Isoforms D, G, and H is frameshifted and truncated compared to the region encoded by exon 7 in Isoform A.

Such polypeptides can further comprise or consist of a segment present in Isoform D (or a fragment or homolog thereof) that is not present in Isoform G (or a fragment or homolog thereof), and can further comprise a segment present in Isoform D (or a fragment or homolog thereof) that is not present in Isoform H (or a fragment or homolog thereof). Such regions can be readily identified by comparing the sequences of the Isoforms. For example, such polypeptides can comprise or consist of a segment of the contiguous amino acids (e.g., at least 3 contiguous

amino acids, at least 5 contiguous amino acids, at least 8 contiguous amino acids, at least 10 contiguous amino acids, or at least 15 contiguous amino acids) that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical (or at least about 90%) to a segment spanning the boundary of the regions encoded by exons 3 and 4 of Isoform D to distinguish from Isoform H. Likewise, such polypeptides can comprise or consist of a segment of the contiguous amino acids (e.g., at least 3 contiguous amino acids, at least 5 contiguous amino acids, at least 8 contiguous amino acids, at least 10 contiguous amino acids, or at least 15 contiguous amino acids) that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical (or at least about 90%) to a segment within the region encoded by exon 2 in Isoform D, a segment spanning the boundary of the regions encoded by exons 1 and 2 in Isoform D, or a segment spanning the boundary of the regions encoded by exons 2 and 3 in Isoform D to distinguish from Isoform G.

Like Isoform D, the region encoded by exon 7 in Isoform H is frameshifted and truncated compared to Isoform A. In addition, however, Isoform H includes a region encoded by an additional exon (exon 3') between exons 3 and 4 compared to Isoforms A and D. Accordingly, such polypeptides can be as described above comprising or consisting of a segment that is present in Isoforms D, G, and H (or fragments or homologs thereof) that is not present in Isoform A (or a fragment or homolog thereof) but further comprising a segment (e.g., at least 8 contiguous amino acids) from Isoform H (or a fragment or homolog thereof) that is not present in Isoform D (or a fragment or homolog thereof). Such regions can be readily identified by comparing the sequences of the Isoforms. For example, such polypeptides can further comprise or consist of a segment of the contiguous amino acids (e.g., at least 3 contiguous amino acids, at least 5 contiguous amino acids, at least 8 contiguous amino acids, at least 10 contiguous amino acids, or at least 15 contiguous amino acids) that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90%) to a segment including at least a portion of the region encoded by exon 3' in Isoform H.

Like Isoform D, the region encoded by exon 7 in Isoform G is frameshifted and truncated compared to Isoform A. In addition, however, Isoform G is missing the region encoded by exon 2 compared to Isoforms A and D and thus includes an exon 1-exon 3 boundary not present in Isoforms A and D. Accordingly, such polypeptides can be as described above comprising or consisting of a segment that is present in Isoforms D, G, and H (or fragments or homologs thereof) that is not present in Isoform A (or a fragment or homolog thereof) but further

comprising a segment (e.g., at least 8 contiguous amino acids) from Isoform G (or a fragment or homolog thereof) that is not present in Isoform D (or a fragment or homolog thereof). Such regions can be readily identified by comparing the sequences of the Isoforms. For example, such polypeptides can further comprise or consist of a segment of the contiguous amino acids (e.g., at least 3 contiguous amino acids, at least 5 contiguous amino acids, at least 8 contiguous amino acids, at least 10 contiguous amino acids, or at least 15 contiguous amino acids) that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90%) to a segment spanning the boundary of the regions encoded by exons 1 and 3 in Isoform G.

Also provided herein are polypeptides comprising or consisting of a segment (e.g., at least 8 contiguous amino acids) that is present in Isoform E (or a fragment or homolog thereof) that is not present in Isoform A (or a fragment or homolog thereof). Isoform E includes a region encoded by an additional exon (exon 3') between exons 3 and 4 that is not present in Isoform A. Such regions can be readily identified by comparing the sequences of the Isoforms. Accordingly, the polypeptides can comprise or consist of at least 5, 6, 8, 10, 12, 14, 15, 16, 18, 20, 22, 24, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, or 200 contiguous amino acids of an HSD17B13 isoform protein (e.g., at least 8 contiguous amino acids, at least 10 contiguous amino acids, or at least 15 contiguous amino acids of an HSD17B13 protein), wherein a segment of the contiguous amino acids (e.g., at least 3 contiguous amino acids, at least 5 contiguous amino acids, at least 8 contiguous amino acids, at least 10 contiguous amino acids, or at least 15 contiguous amino acids) is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90%) to a segment including at least a portion of the region encoded by exon 3' in Isoform E or Isoform H. Optionally, such polypeptides can further comprise or consist of a segment (e.g., at least 8 contiguous amino acids) from Isoform E (or a fragment or homolog thereof) that is not present in Isoform H (or a fragment or homolog thereof). Such regions can be readily identified by comparing the sequences of the Isoforms. For example, such polypeptides can further comprise or consist of a segment of the contiguous amino acids (e.g., at least 3 contiguous amino acids, at least 5 contiguous amino acids, at least 8 contiguous amino acids, at least 10 contiguous amino acids, or at least 15 contiguous amino acids) that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90%) a segment spanning the boundary of the regions encoded by exons 6 and 7 in Isoform E.

Also provided herein are polypeptides comprising or consisting of a segment (e.g., at least 8 contiguous amino acids) present in Isoform F (or a fragment or homolog thereof) that is not present in Isoform A (or a fragment or homolog thereof). Isoform F includes a region encoded by read-through from exon 6 into intron 6 that is not present in Isoform A. Such regions can be readily identified by comparing the sequences of the Isoforms. Accordingly, the polypeptides can comprise or consist of at least 5, 6, 8, 10, 12, 14, 15, 16, 18, 20, 22, 24, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, or 200 contiguous amino acids of an HSD17B13 isoform protein (e.g., at least 8 contiguous amino acids, at least 10 contiguous amino acids, or at least 15 contiguous amino acids of an HSD17B13 protein), wherein a segment of the contiguous amino acids (e.g., at least 3 contiguous amino acids, at least 5 contiguous amino acids, at least 8 contiguous amino acids, at least 10 contiguous amino acids, or at least 15 contiguous amino acids) is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90%) to a segment including at least a portion of the region encoded by the read-through into intron 6 in Isoform F.

Also provided herein are polypeptides comprising or consisting of a segment (e.g., at least 8 contiguous amino acids) present in Isoform C (or a fragment or homolog thereof) that is not present in Isoform A (or a fragment or homolog thereof). Isoform C is missing the region encoded by exon 6 compared to Isoform A and includes an exon 5-exon 7 boundary not present in Isoform A. Such regions can be readily identified by comparing the sequences of the Isoforms. Accordingly, the polypeptides can comprise at least 5, 6, 8, 10, 12, 14, 15, 16, 18, 20, 22, 24, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, or 200 contiguous amino acids of an HSD17B13 protein isoform (e.g., at least 8 contiguous amino acids, at least 10 contiguous amino acids, or at least 15 contiguous amino acids of an HSD17B13 protein), wherein a segment of the contiguous amino acids (e.g., at least 3 contiguous amino acids, at least 5 contiguous amino acids, at least 8 contiguous amino acids, at least 10 contiguous amino acids, or at least 15 contiguous amino acids) is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90%) to a segment spanning the boundary of the regions encoded by exons 5 and 7 in Isoform C.

Any of the isolated polypeptides disclosed herein can be linked to a heterologous molecule or heterologous label. Examples of such heterologous molecules or labels are disclosed elsewhere herein. For example, the heterologous molecule can be an immunoglobulin Fc domain, a peptide tag as disclosed elsewhere herein, poly(ethylene glycol), polysialic acid, or glycolic acid.

The present disclosure also provides methods of producing any of the polypeptides or fragments thereof disclosed herein. For example, polypeptides or fragments thereof can be produced from host cells comprising nucleic acid molecules (*e.g.*, recombinant expression vectors) encoding such polypeptides or fragments thereof. Such methods can comprise culturing a host cell comprising a nucleic acid molecule (*e.g.*, recombinant expression vector) encoding a polypeptide or fragment thereof under conditions sufficient to produce the polypeptide or fragment thereof, thereby producing the polypeptide or fragment thereof. The nucleic acid can be operably linked to a promoter active in the host cell, and the culturing can be carried out under conditions whereby the nucleic acid is expressed. Such methods can further comprise recovering the expressed polypeptide or fragment thereof. The recovering can further comprise purifying the polypeptide or fragment thereof.

Examples of suitable systems for protein expression include host cells such as, for example: bacterial cell expression systems (*e.g.*, *Escherichia coli*, *Lactococcus lactis*), yeast cell expression systems (*e.g.*, *Saccharomyces cerevisiae*, *Pichia pastoris*), insect cell expression systems (*e.g.*, baculovirus-mediated protein expression), and mammalian cell expression systems.

In some embodiments, the nucleic acid molecules encode a tag in frame with the polypeptide or fragment thereof to facilitate protein purification. Examples of tags are disclosed elsewhere herein. Such tags can, for example, bind to a partner ligand (*e.g.*, immobilized on a resin) such that the tagged protein can be isolated from all other proteins (*e.g.*, host cell proteins).

Other methods can also be used to produce polypeptides or fragments thereof. For example, two or more peptides or polypeptides can be linked together by protein chemistry techniques. For example, peptides or polypeptides can be chemically synthesized using either Fmoc (9-fluorenylmethyloxycarbonyl) or Boc (t*ert* -butyloxycarbonoyl) chemistry. Alternately, the peptide or polypeptide can be independently synthesized *in vivo* as described herein. Once isolated, these independent peptides or polypeptides may be linked to form a peptide or fragment thereof via similar peptide condensation reactions.

In some embodiments, the polypeptides can possess post-expression modifications such as, for example, glycosylations, acetylations, and phosphorylations, as well as other modifications known in the art, both naturally occurring and non-naturally occurring. A polypeptide may be an entire protein, or a subsequence thereof.

The present disclosure also provides methods of producing any of the polypeptides disclosed herein, comprising culturing a host cell comprising a recombinant expression vectors

comprising nucleic acid molecules comprising a polynucleotide capable of encoding one or more of the polypeptides disclosed herein, or its complement, thereby producing the polypeptide.

The polypeptides disclosed herein can comprise an amino acid sequence of a naturally occurring HSD17B13 isoform protein, or can comprise a non-naturally occurring sequence. In one example, the non-naturally occurring sequence can differ from the non-naturally occurring sequence due to conservative amino acid substitutions. For example, the sequence can be identical with the exception of conservative amino acid substitutions.

Any of the polypeptides disclosed herein can further have one or more substitutions (such as conservative amino acid substitutions), insertions, or deletions. Insertions include, for example, amino or carboxyl terminal fusions as well as intrasequence insertions of single or multiple amino acid residues. Techniques for making substitutions at predetermined sites in DNA having a known sequence are well known, for example M13 primer mutagenesis and PCR mutagenesis. Amino acid substitutions are typically of single residues, but can occur at a number of different locations at once; insertions usually will be on the order of about from 1 to 10 amino acid residues; and deletions will range about from 1 to 30 residues. Deletions or insertions can be made in adjacent pairs, i.e. a deletion of 2 residues or insertion of 2 residues. Substitutions, deletions, insertions or any combination thereof may be combined to arrive at a final construct. In some embodiments, the mutations do not place the sequence out of reading frame and do not create complementary regions that could produce secondary mRNA structure.

In some embodiments, the polypeptides disclosed herein are linked or fused to heterologous polypeptides or heterologous molecules or labels, numerous examples of which are disclosed elsewhere herein. For example, the proteins can be fused to a heterologous polypeptide providing increased or decreased stability. The fused domain or heterologous polypeptide can be located at the N-terminus, the C-terminus, or internally within the polypeptide. A fusion partner may, for example, assist in providing T helper epitopes (an immunological fusion partner), or may assist in expressing the polypeptide (an expression enhancer) at higher yields than the native recombinant polypeptide. Certain fusion partners are both immunological and expression enhancing fusion partners. Other fusion partners may be selected to increase the solubility of the polypeptide or to facilitate targeting the polypeptide to desired intracellular compartments. Some fusion partners include affinity tags, which facilitate purification of the polypeptide.

In some embodiments, a fusion protein is directly fused to the heterologous molecule or is linked to the heterologous molecule via a linker, such as a peptide linker. For example, peptide linker sequences may contain Gly, Asn and Ser residues. Other near neutral amino acids, such as Thr and Ala may also be used in the linker sequence. A linker sequence may generally be, for

example, from 1 to about 50 amino acids in length. Linker sequences are generally not required when the first and second polypeptides have non-essential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric interference.

In some embodiments, the polypeptides are operably linked to a cell-penetrating domain. For example, the cell-penetrating domain can be derived from the HIV-1 TAT protein, the TLM cell-penetrating motif from human hepatitis B virus, MPG, Pep-1, VP22, a cell-penetrating peptide from Herpes simplex virus, or a polyarginine peptide sequence. The cell-penetrating domain can be located at the N-terminus, the C-terminus, or anywhere within the protein.

In some embodiments, the polypeptides are operably linked to a heterologous polypeptide for ease of tracking or purification, such as a fluorescent protein, a purification tag, or an epitope tag. Examples of fluorescent proteins include, but are not limited to, green fluorescent proteins (*e.g.*, GFP, GFP-2, tagGFP, turboGFP, eGFP, Emerald, Azami Green, Monomeric Azami Green, CopGFP, AceGFP, ZsGreenl), yellow fluorescent proteins (*e.g.*, YFP, eYFP, Citrine, Venus, YPet, PhiYFP, ZsYellowl), blue fluorescent proteins (*e.g.*, eBFP, eBFP2, Azurite, mKalamal, GFPuv, Sapphire, T-sapphire), cyan fluorescent proteins (*e.g.*, eCFP, Cerulean, CyPet, AmCyanl, Midoriishi-Cyan), red fluorescent proteins (*e.g.*, mKate, mKate2, mPlum, DsRed monomer, mCherry, mRFP1, DsRed-Express, DsRed2, DsRed-Monomer, HcRed-Tandem, HcRedl, AsRed2, eqFP611, mRaspberry, mStrawberry, Jred), orange fluorescent proteins (*e.g.*, mOrange, mKO, Kusabira-Orange, Monomeric Kusabira-Orange, mTangerine, tdTomato), and any other suitable fluorescent protein. Examples of tags include, but are not limited to, glutathione-S-transferase (GST), chitin binding protein (CBP), maltose binding protein, thioredoxin (TRX), poly(NANP), tandem affinity purification (TAP) tag, myc, AcV5, AU1, AU5, E, ECS, E2, FLAG, hemagglutinin (HA), nus, Softag 1, Softag 3, Strep, SBP, Glu-Glu, HSV, KT3, S, S1, T7, V5, VSV-G, histidine (His), biotin carboxyl carrier protein (BCCP), and calmodulin. In some embodiments, the heterologous molecule is an immunoglobulin Fc domain, a peptide tag, a transduction domain, poly(ethylene glycol), polysialic acid, or glycolic acid.

In some embodiments, isolated polypeptides comprise non-natural or modified amino acids or peptide analogs. For example, there are numerous D-amino acids or amino acids which have a different functional substituent than the naturally occurring amino acids. The opposite stereo isomers of naturally occurring peptides are disclosed, as well as the stereo isomers of peptide analogs.

In some embodiments, the isolated polypeptides are peptide mimetics, which can be produced to resemble peptides, but which are not connected via a natural peptide linkage. For example, linkages for amino acids or amino acid analogs include, but are not limited to, -CH_2NH-, -CH_2S-, -CH_2-, -CH=CH- (cis and trans), -$COCH_2$-, -$CH(OH)CH_2$-, and -CHH_2SO-. Peptide analogs can have more than one atom between the bond atoms, such as b-alanine, gaminobutyric acid, and the like.

In some embodiments, the polypeptides comprise D-amino acids, which can be used to generate more stable peptides because D amino acids are not recognized by peptidases. Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (*e.g.*, D-lysine in place of L-lysine) can be used to generate more stable peptides. Cysteine residues can be used to cyclize or attach two or more peptides together.

The present disclosure also provides nucleic acid molecules encoding any of the polypeptides disclosed herein. This includes all degenerate sequences related to a specific polypeptide sequence (all nucleic acid molecules having a sequence that encodes one particular polypeptide sequence as well as all nucleic acids, including degenerate nucleic acids, encoding the disclosed variants and derivatives of the protein sequences). Thus, while each particular nucleotide sequence may not be written out herein, each and every sequence is in fact disclosed and described herein through the disclosed polypeptide sequences.

Percent identity (or percent complementarity) between particular stretches of nucleotide sequences within nucleic acid molecules or amino acid sequences within polypeptides can be determined routinely using BLAST programs (basic local alignment search tools) and PowerBLAST programs (Altschul *et al.*, J. Mol. Biol., 1990, 215, 403-410; Zhang and Madden, Genome Res., 1997, 7, 649-656) or by using the Gap program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, Madison Wis.), using default settings, which uses the algorithm of Smith and Waterman (Adv. Appl. Math., 1981, 2, 482-489). Herein, if reference is made to percent sequence identity, the higher percentages of sequence identity are preferred over the lower ones.

The present disclosure also provides compositions comprising any one or more of the nucleic acid molecules and/or any one or more of the polypeptides disclosed herein and a carrier and/or excipient. In some embodiments, the carrier increases the stability of the nucleic acid molecule and/or polypeptide (*e.g.*, prolonging the period under given conditions of storage (*e.g.*, -20°C, 4°C, or ambient temperature) for which degradation products remain below a threshold, such as below 0.5% by weight of the starting nucleic acid or protein; or increasing the stability *in vivo*). Examples of carriers include, but are not limited to, poly(lactic acid) (PLA) microspheres,

poly(D,L-lactic-coglycolic-acid) (PLGA) microspheres, liposomes, micelles, inverse micelles, lipid cochleates, and lipid microtubules. A carrier may comprise a buffered salt solution such as PBS, HBSS, etc.

The nucleic acid molecules and polypeptides disclosed herein can be introduced into a cell by any means. Non-limiting transfection methods include chemical-based transfection methods using liposomes, nanoparticles, calcium, dendrimers, and cationic polymers such as DEAE-dextran or polyethylenimine. Viral methods can also be used for transfection by, for example, adenovirus, adeno-associated virus, lentivirus, retrovirus, by transfection, by lipid-mediated transfection, or by nucleofection. In some embodiments, nucleofection is performed using the LONZA® NUCLEOFECTOR™ system. Introduction of nucleic acid molecules or proteins into a cell can also be accomplished by microinjection. Non-chemical methods include electroporation, sono-poration, optical transfection, particle-based transfection including use of a gene gun, or magnet-assisted transfection, intracytoplasmic injection. Introduction of nucleic acid molecules and proteins into cells can also be accomplished by hydrodynamic delivery (HDD). In some embodiments, a nucleic acid or protein can be introduced into a cell in a carrier such as a poly(lactic acid) (PLA) microsphere, a poly(D,L-lactic-coglycolic-acid) (PLGA) microsphere, a liposome, a micelle, an inverse micelle, a lipid cochleate, or a lipid microtubule.

The introduction of nucleic acid molecules or proteins into the cell can be performed one time or multiple times over a period of time. For example, the introduction can be performed at least two times over a period of time, at least three times over a period of time, at least four times over a period of time, at least five times over a period of time, at least six times over a period of time, at least seven times over a period of time, at least eight times over a period of time, at least nine times over a period of times, at least ten times over a period of time, at least eleven times, at least twelve times over a period of time, at least thirteen times over a period of time, at least fourteen times over a period of time, at least fifteen times over a period of time, at least sixteen times over a period of time, at least seventeen times over a period of time, at least eighteen times over a period of time, at least nineteen times over a period of time, or at least twenty times over a period of time.

The present disclosure provides methods for detecting the presence of the variant *HSD17B13* rs72613567 gene in a biological sample comprising genomic DNA, for detecting the presence or levels of any one of or a combination of *HSD17B13* Transcripts C, D, E, F, F', G, and H, and particularly D, in a biological sample comprising RNA, or cDNA derived therefrom, or comprising mRNA, or cDNA derived therefrom, or for detecting the presence or levels of any one of or a combination of HSD17B13 protein Isoforms C, D, E, F, F', G, or H, and particularly

D, in a biological sample comprising protein. It is understood that gene sequences within a population and RNAs, mRNAs, and proteins encoded by such genes can vary due to polymorphisms such as single-nucleotide polymorphisms. The sequences provided herein for the *HSD17B13* gene and for each *HSD17B13* Transcript and HSD17B13 Isoform are only exemplary sequences for the *HSD17B13* gene and for each *HSD17B13* Transcript (RNA, mRNA, and cDNA derived therefrom) and HSD17B13 isoform. Other sequences for the *HSD17B13* gene and for each *HSD17B13* Transcript and HSD17B13 Isoform are also possible.

The biological sample can be derived from any cell, tissue, or biological fluid from the subject. The sample may comprise any clinically relevant tissue, such as a bone marrow sample, a tumor biopsy, a fine needle aspirate, or a sample of bodily fluid, such as blood, gingival crevicular fluid, plasma, serum, lymph, ascitic fluid, cystic fluid, or urine. In some cases, the sample comprises a buccal swab. The sample used in the methods disclosed herein will vary based on the assay format, nature of the detection method, and the tissues, cells, or extracts that are used as the sample.

A biological sample can be processed differently depending on the assay being employed. For example, when detecting the *HSD17B13* rs72613567 variant nucleic acid molecule, preliminary processing designed to isolate or enrich the sample for the genomic DNA can be employed. A variety of known techniques may be used for this purpose. When detecting the level of *HSD17B13* Transcript C, D, E, F, G, or H mRNA, different techniques can be used to enrich the biological sample with mRNA. Various methods to detect the presence or level of a particular *HSD17B13* rs72613567 variant nucleic acid molecule can be used.

The present disclosure provides methods for detecting a variant *HSD17B13* rs72613567 gene in a cell or in a subject such as a human subject.

The present disclosure provides methods of detecting a variant HSD17B13 gene in a human subject, comprising or consisting of performing an assay on a biological sample obtained from the human subject, wherein the assay determines whether a thymine is inserted between positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene, or whether a thymine is present at a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene, wherein the presence of said thymine is indicative for a variant HSD17B13 gene. In some embodiments, the assay comprises or consists of sequencing a portion of the HSD17B13 gene including positions corresponding to positions 12665 and 12666 of SEQ ID NO:1, or including a position corresponding to position 12666 of SEQ ID NO:2. In some embodiments, the assay comprises or consists of: i) contacting the biological sample with a primer hybridizing to a region of the HSD17B13 gene that is within 50

nucleotides of a position of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1, or within 50 nucleotides of a position of the HSD17B13 gene corresponding to position 12666 of SEQ ID NO:2; ii) extending the primer at least through the position of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1, or corresponding to position 12666 of SEQ ID NO:2; and iii) determining whether a thymine is inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene, or whether a thymine is present at a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene, in an extension product of the primer. In some embodiments, the method further comprises determining whether the human subject is homozygous for the variant HSD17B13 gene.

In some embodiments, the methods comprise or consist of, for example, obtaining a biological sample from the subject comprising an *HSD17B13* gene, and performing an assay on the biological sample that determines that a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene is occupied by a thymine or that a thymine is inserted between positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene. It is understood that determining that a position of the *HSD17B13* gene corresponding to position 12666 of SEQ ID NO:2 is occupied by a thymine means that the identity of a sufficient number of nucleotides is determined in the positions flanking the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 that it can be determined that a thymine is inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1. Such assays can comprise, for example determining the identity of positions corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene (or positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene) and one or more surrounding positions (e.g., at least 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 positions flanking one side or each side of position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene or positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene).

The assay in such a method can comprise, for example, sequencing a portion of the *HSD17B13* gene including a position corresponding to position 12666 or positions 12666 and 12667 of SEQ ID NO:2. Likewise, the assay can comprise sequencing a portion of the *HSD17B13* gene including positions corresponding to positions 12665 and 12666 of SEQ ID NO:1. As an example, this method can comprise: i) contacting the biological sample with a primer, such as an alteration-specific primer, hybridizing to a segment of the *HSD17B13* gene that is proximate to a position of the *HSD17B13* gene corresponding to position 12666 or positions 12666 and 12667 of SEQ ID NO:2; ii) extending the primer at least through the

position of the *HSD17B13* gene corresponding to position 12666 or positions 12666 and 12667 of SEQ ID NO:2; and iii) determining the identity of the position of the *HSD17B13* gene corresponding to position 12666 or positions 12666 and 12667 of SEQ ID NO:2 in an extension product of the primer. As another example, this method can comprise: i) contacting the biological sample with a primer, such as an alteration-specific primer, hybridizing to a segment of the *HSD17B13* gene that is proximate to positions of the *HSD17B13* gene corresponding 12665 and 12666 of SEQ ID NO:1; ii) extending the primer at least through the positions of the *HSD17B13* gene corresponding to 12665 and 12666 of SEQ ID NO:2; and iii) determining the whether a thymine is present between the positions of the *HSD17B13* gene corresponding to positions 12665 and 12666 of SEQ ID NO:1 in an extension product of the primer. In some embodiments, the alteration-specific probe or alteration-specific primer comprises or consists of a nucleotide sequence which is complementary to and/or hybridizes, or specifically hybridizes, to a particular *HSD17B13* gene or transcript, such as Transcript D, but which does not hybridize, or specifically hybridize, to a wild type *HSD17B13* gene (SEQ ID NO:1). As used herein, "proximate" means within about 50, within about 45, within about 40, within about 35, within about 30, within about 25, within about 20, within about 15, within about 10, or within about 5, nucleotides of the particular stated position.

Alternatively, the assay in such a method can comprise contacting the biological sample with a primer or probe that specifically hybridizes to the *HSD17B13* rs72613567 variant and not the corresponding wild type *HSD17B13* sequence (e.g., under stringent conditions), and determining whether hybridization has occurred.

The present disclosure provides methods of detecting the presence of an HSD17B13 Transcript in a human subject.

The present disclosure provides methods of detecting the presence of an HSD17B13 Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) in a human subject, comprising or consisting of performing an assay on a biological sample obtained from the subject, wherein the assay determines the presence of an HSD17B13 Transcript D in the biological sample. In some embodiments, the assay comprises or consists of contacting the biological sample with one or more primers or probes that specifically hybridize to a nucleic acid sequence of an HSD17B13 Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or to the complement thereof, and determining whether hybridization has occurred. In some embodiments, the method further comprises specifically detecting Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA

derived therefrom; preferably RNA or cDNA derived therefrom) by using a nucleic acid molecule comprising or consisting of from about 5 nucleotides up to about 50 nucleotides comprising or consisting of: i) a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90%) to a nucleotide sequence of a SEQ ID NO:6, 15, 24, or 33, or to the complement thereof; ii) a nucleic acid molecule which specifically hybridizes to exon 2 of Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom); and/or iii) a nucleic acid molecule which specifically hybridizes to the region which bridges exons 3 and 4 of Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). In some embodiments, the HSD17B13 Transcript D comprises or consists of a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90%) to SEQ ID NO:6, 15, 24, or 33. In some embodiments, the one or more primers or probes specifically hybridize to SEQ ID NO:6, SEQ ID NO:15, SEQ ID NO:24, and/or SEQ ID NO:33. In some embodiments, the assay comprises reverse transcription polymerase chain reaction (RT-PCR). In some embodiments, the assay comprises sequencing.

The present disclosure provides methods for detecting the presence of one or a combination of *HSD17B13* Transcripts C, D, E, F, F', G, or H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) in a cell or in a subject such as a human subject. Such methods can comprise or consist of, for example, obtaining a biological sample from the subject comprising RNA, or cDNA derived therefrom, or comprising mRNA, or cDNA derived therefrom, and performing an assay on the sample that determines the presence of Transcript C, D, E, F, F', G, or H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) in the subject. For example, such an assay can detect a region or combination of regions that is present in (e.g., is unique to) one or more of Transcripts C, D, E, F, F', G, or H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) that is not present in Transcripts A and B (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). Such a region may be unique to a particular Transcript (e.g., unique to Transcript C (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom)) or unique to a combination of Transcripts (e.g., unique to Transcripts D, G, and H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived

therefrom; preferably RNA or cDNA derived therefrom)). Such regions can be readily identified by comparing the sequences of Transcripts A-H and are described in further detail elsewhere herein.

As one example, the assay can comprise RNA sequencing (RNA-Seq). As another example, the assay can comprise or consist of contacting the biological sample with one or more primers or probes that specifically hybridize to one or more sequences, the combination of which is unique to one or a combination of Transcripts C, D, E, F, F', G, or H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) among *HSD17B13* Transcripts A, B, C, D, E, F, F', G, and H (i.e., that is not present in Transcripts A and B(RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom)), and determining whether hybridization has occurred. Optionally, the assay can comprise reverse transcription polymerase chain reaction (RT-PCR). Such assays can be specific for a particular *HSD17B13* Transcript or for a particular combination of *HSD17B13* Transcripts. For example, Transcripts D, G, and H each include an additional guanine inserted at the 3' end of exon 6 compared to Transcripts A, B, and E (Transcript C does not include exon 6), and Transcripts D, G, and H each include exon 7, whereas Transcript F reads through from exon 6 to intron 6. Thus, a primer or probe hybridizing to the region spanning the boundary of exon 6 and exon 7 in Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) can specifically detect the combination of Transcripts D, G, and H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). It is understood that such a primer or probe would be designed to hybridize to a sufficient number of nucleotides in each of exons 6 and 7 to distinguish the inserted guanine from other features in the *HSD17B13* Transcripts (e.g., from the read-through into intron 6 in Transcript F (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or from the deleted exon 6 in Transcript C (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom)). Likewise, Transcripts E and H each include exon 3' compared to all of the other transcripts. Thus, a primer or probe specifically hybridizing to a region within exon 3' or its boundary with exon 3 or exon 4 can specifically detect the combination of Transcripts E and H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). It is understood that such a primer or probe would be designed to hybridize to a sufficient number of nucleotides in each of exons 3 and 3' or each of exons 3' and 4 to distinguish from other features in the *HSD17B13* transcripts (e.g., from the

boundary of exons 3 and 4). Likewise, Transcripts B and G are each missing exon 2. Thus, a primer or probe specifically hybridizing to a region spanning the boundary of exons 1 and 3 can specifically detect the combination of Transcripts B and G (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). It is understood that a primer or probe would be designed to hybridize to a sufficient number of nucleotides in each of exons 1 and 3 to distinguish from other features in the *HSD17B13* Transcripts (e.g., the boundary of exons 1 and 2 or the boundary of exons 2 and 3).

In one specific example, the one or more primers or probes specifically hybridize to a region spanning the boundary of exons 6 and 7 in Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), Transcript G (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or Transcript H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) (i.e., including the additional guanine at the 3' end of exon 6 that is not present in exon 6 in other *HSD17B13* Transcripts). Optionally, the one or more primers or probes further specifically hybridize to a region within exon 3' of Transcript H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), a region spanning the exon 3-exon 3' boundary of Transcript H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a region spanning the exon 3'-exon 4 boundary of Transcript H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or further specifically hybridize to a region spanning the exon 1-exon 3 boundary in Transcript G (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). Optionally, the one or more primers or probes further specifically hybridize to a region within exon 1 of any of Transcripts A-H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) (common to Transcripts A-H). For example, a primer specifically hybridizing to a region within exon 1 of any of Transcripts A-H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) and a primer specifically hybridizing to a region spanning the boundary of exons 6 and 7 in Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), Transcript G (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or Transcript H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived

therefrom; preferably RNA or cDNA derived therefrom) can be used to amplify the intervening sequence and distinguish between Transcripts D, G, and H based on the size of the amplified product, as Transcript G is missing exon 2 compared to Transcript D, and Transcript H includes an additional exon between exons 3 and 4 compared to Transcript D.

In another specific example, the one or more primers or probes specifically hybridize to a region within exon 3' of Transcript E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or Transcript H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), a region spanning the exon 3-exon 3' boundary of Transcript E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or Transcript H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a region spanning the exon 3'-exon 4 boundary of Transcript E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or Transcript H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). Optionally, the one or more primers or probes can further specifically hybridize to a region spanning the boundary of exons 6 and 7 in Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), Transcript G (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or Transcript H(RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). Alternatively, the one or more primers or probes can further specifically hybridize to a region spanning the boundary of exons 6 and 7 in Transcript E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). For example, a primer specifically hybridizing to a region within exon 3' of Transcript E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or Transcript H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), a region spanning the exon 3-exon 3' boundary of Transcript E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or Transcript H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or a region spanning the exon 3'-exon 4 boundary of Transcript E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or Transcript H (RNA or cDNA derived therefrom,

and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) and a primer specifically hybridizing to a region spanning the boundary of exons 6 and 7 in Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), Transcript G (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or Transcript H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) can be used to amplify the intervening sequence and distinguish between Transcripts E and H, because only Transcript H and not Transcript E includes the additional guanine at the 3' end of exon 6.

In another specific example, the one or more primers or probes specifically hybridize to a region within the read-through into intron 6 in Transcript F (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or a region spanning the boundary between the read-through into intron 6 and the rest of exon 6 in Transcript F (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). It is understood that such a primer or probe would be designed to hybridize to a sufficient number of nucleotides in the read-through to distinguish the read-through from other features in the *HSD17B13* Transcripts (e.g., from boundary of exons 6 and 7 in other *HSD17B13* Transcripts). Optionally, the contiguous nucleotides comprise a sequence present in Transcript F (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) (i.e., the inserted thymine) that is not present in Transcript F' (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). Transcript F' also includes a read-through from exon 6 into intron 6 compared to Transcript A, but the read-through does not include the inserted thymine present in the *HSD17B13* rs72613567 variant gene.

In another specific example, the one or more primers or probes specifically hybridize to a region within the read-through into intron 6 in Transcript F' (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or a region spanning the boundary between the read-through into intron 6 and the rest of exon 6 in Transcript F' (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). It is understood that such a primer or probe would be designed to hybridize to a sufficient number of nucleotides in the read-through to distinguish the read-through from other features in the *HSD17B13* Transcripts (e.g., from boundary of exons 6 and 7 in other *HSD17B13* Transcripts). Optionally, the contiguous nucleotides comprise a sequence present in Transcript F' (RNA or cDNA derived therefrom,

and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) that is not present in Transcript F (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). The read-through in Transcript F includes the inserted thymine present in the *HSD17B13* rs72613567 variant gene, whereas the read-through in Transcript F' does not.

In yet another specific example, the one or more primers or probes specifically hybridize to a region spanning the exon 5-exon 7 boundary in Transcript C (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). It is understood that such a primer or probe would be designed to hybridize to a sufficient number of nucleotides in exons 5 and 7 to distinguish from other features in the *HSD17B13* Transcripts (e.g., from boundary of exons 5 and 6 or of exons 6 and 7 in other *HSD17B13* Transcripts).

Certain methods utilize probes and primers (described in more detail elsewhere herein) of sufficient nucleotide length to bind to the target DNA sequence and specifically detect and/or identify a polynucleotide comprising the *HSD17B13* rs72613567 variant or specific HSD17B13 RNA or mRNA transcripts, or cDNA derived therefrom. The hybridization conditions or reaction conditions can be determined by the operator to achieve this result. This length may be any length that is sufficient to be useful in a detection method of choice. Such probes and primers can hybridize specifically to a target sequence under high stringency hybridization conditions. Probes and primers may have complete DNA sequence identity of contiguous nucleotides with the target sequence, although probes differing from the target DNA sequence and that retain the ability to specifically detect and/or identify a target DNA sequence may be designed by conventional methods. Accordingly, probes and primers can share about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, or 100% sequence identity or complementarity to the target polynucleotide. In some embodiments, the probes and primers can share about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, or 100% sequence identity or complementarity to the target polynucleotide. In some embodiments, the probes and primers can share about 95%, about 96%, about 97%, about 98%, about 99%, or 100% sequence identity or complementarity to the target polynucleotide.

Specific primers can be used to amplify the *HSD17B13* rs72613567 variant gene and/or specific HSD17B13 RNA or mRNA transcripts to produce an amplicon that can be used as a "specific probe" or can itself be detected for identifying the *HSD17B13* rs72613567 variant gene or for determining the level of specific HSD17B13 RNA or mRNA transcripts in a biological

sample. The *HSD17B13* variant gene can be used to denote a genomic nucleic acid sequence including a position corresponding to residue 12666 in SEQ ID NO:2 (insertion of a thymine relative to the wild type genomic locus set forth in SEQ ID NO:1 (i.e., inserted between positions 12665 and 12666 in SEQ ID NO:1)). When the probe is hybridized with the polynucleotides of a biological sample under conditions that allow for the binding of the probe to the sample, this binding can be detected and thus allow for an indication of the presence of the *HSD17B13* rs72613567 variant gene or the presence or the level of specific HSD17B13 RNA or mRNA transcripts in the biological sample. Such identification of a bound probe has been described. The specific probe may comprise a sequence of at least about 80%, from about 80% to about 85%, from about 85% to about 90%, from about 90% to about 95%, and from about 95% to about 100% identical (from about 90% to about 95% or from about 95% to about 100% identical) (or complementary) to a specific region of the *HSD17B13* gene, an HSD17B13 RNA or mRNA transcript, or an HSD17B13 cDNA derived therefrom.

To determine whether a nucleic acid molecule within a biological sample comprises the inserted thymine at residue 12666 in the *HSD17B13* rs72613567 variant gene (e.g., SEQ ID NO:2) (i.e., an inserted thymine between residues 12665 and 12666 in the wild type *HSD17B13* locus (SEQ ID NO:1)), the biological sample may be subjected to a polynucleotide amplification method using a primer pair that includes a first primer derived from the 5' flanking sequence adjacent to the inserted thymine and a second primer derived from the 3' flanking sequence adjacent to the inserted thymine to produce an amplicon that is diagnostic for the presence of the inserted thymine at residue 12666 in the *HSD17B13* rs72613567 variant gene (SEQ ID NO:2) (i.e., an inserted thymine between residues 12665 and 12666 in the wild type *HSD17B13* gene (SEQ ID NO:1)). In some cases, the amplicon may range in length from the combined length of the primer pairs plus one nucleotide base pair to any length of amplicon producible by a DNA amplification protocol. This distance can range from one nucleotide base pair up to the limits of the amplification reaction, or about twenty thousand nucleotide base pairs. Optionally, the primer pair flanks a region including the inserted thymine and at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more nucleotides on each side of the thymine.

PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as the PCR primer analysis tool in Vector NTI version 10 (Informax Inc., Bethesda Md.); PrimerSelect (DNASTAR Inc., Madison, Wis.); and Primer3 (Version 0.4.0.COPYRGT., 1991, Whitehead Institute for Biomedical Research, Cambridge, Mass.). Additionally, the sequence can be visually scanned and primers manually identified using known guidelines.

As outlined in further detail below, any conventional nucleic acid hybridization or amplification or sequencing method can be used to specifically detect the presence of the *HSD17B13* rs72613567 variant locus and/or the level of specific HSD17B13 RNA or mRNA transcripts. By "specifically detect" is intended that the polynucleotide can be used either as a primer to amplify a region of the *HSD17B13* polynucleotide or the polynucleotide can be used as a probe that hybridizes under stringent conditions to a polynucleotide comprising the *HSD17B13* rs72613567 variant gene or a polynucleotide comprising a specific HSD17B13 transcript, particularly Transcript C, D, E, F, G, or H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom).

A variety of techniques are available in the art including, for example, nucleic acid sequencing, nucleic acid hybridization, and nucleic acid amplification. Illustrative examples of nucleic acid sequencing techniques include, but are not limited to, chain terminator (Sanger) sequencing and dye terminator sequencing.

Other methods involve nucleic acid hybridization methods other than sequencing, including using labeled primers or probes directed against purified DNA, amplified DNA, and fixed cell preparations (fluorescence *in situ* hybridization (FISH)). In some methods, a target nucleic acid may be amplified prior to or simultaneous with detection. Illustrative examples of nucleic acid amplification techniques include, but are not limited to, polymerase chain reaction (PCR), ligase chain reaction (LCR), strand displacement amplification (SDA), and nucleotide sequence based amplification (NASBA). Other methods include, but are not limited to, ligase chain reaction, strand displacement amplification, and thermophilic SDA (tSDA).

Any method can be used for detecting either the non-amplified or amplified polynucleotides including, for example, Hybridization Protection Assay (HPA), quantitative evaluation of the amplification process in real-time, and determining the quantity of target sequence initially present in a sample, but which is not based on a real-time amplification.

Also provided are methods for identifying nucleic acid molecules which do not necessarily require sequence amplification and are based on, for example, the known methods of Southern (DNA:DNA) blot hybridizations, *in situ* hybridization (ISH), and fluorescence *in situ* hybridization (FISH) of chromosomal material. Southern blotting can be used to detect specific nucleotide sequences. In such methods, nucleic acid that is extracted from a sample is fragmented, electrophoretically separated on a matrix gel, and transferred to a membrane filter.

Examples of suitable quantitative assays include fluorescence-mediated in situ hybridization (FISH), comparative genomic hybridization, isothermic DNA amplification, quantitative hybridization to an immobilized probe(s), INVADER® Probes, TAQMAN®

Molecular Beacon probes, or ECLIPSE™ probe technology. Conventional assays for screening for targeted modifications, such as long-range PCR, Southern blotting, or Sanger sequencing, can also be used. Next generation sequencing (NGS) can also be used for screening. Next-generation sequencing can also be referred to as "NGS" or "massively parallel sequencing" or "high throughput sequencing."

In hybridization techniques, stringent conditions can be employed such that a probe or primer will specifically hybridize to its target. In some embodiments, a polynucleotide primer or probe under stringent conditions will hybridize to its target sequence (*e.g.*, the variant *HSD17B13* gene, variant *HSD17B13* RNA or cDNA corresponding thereto, or variant *HSD17B13* mRNA or cDNA corresponding thereto) to a detectably greater degree than to other sequences (*e.g.*, the corresponding wild type *HSD17B13* gene, wild type *HSD17B13* RNA or cDNA corresponding thereto, or wild type *HSD17B13* mRNA or cDNA corresponding thereto), such as, at least 2-fold, at least 3-fold, at least 4-fold, or more over background, including over 10-fold over background. In some embodiments, a polynucleotide primer or probe under stringent conditions will hybridize to its target sequence to a detectably greater degree than to other sequences by at least 2-fold. In some embodiments, a polynucleotide primer or probe under stringent conditions will hybridize to its target sequence to a detectably greater degree than to other sequences by at least 3-fold. In some embodiments, a polynucleotide primer or probe under stringent conditions will hybridize to its target sequence to a detectably greater degree than to other sequences by at least 4-fold. In some embodiments, a polynucleotide primer or probe under stringent conditions will hybridize to its target sequence to a detectably greater degree than to other sequences by over 10-fold over background. Stringent conditions are sequence-dependent and will be different in different circumstances.

Appropriate stringency conditions which promote DNA hybridization, for example, 6X sodium chloride/sodium citrate (SSC) at about 45°C., followed by a wash of 2X SSC at 50°C, are known or can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Typically, stringent conditions for hybridization and detection will be those in which the salt concentration is less than about 1.5 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (*e.g.*, 10 to 50 nucleotides) and at least about 60°C for longer probes (*e.g.*, greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulfate) at 37°C, and a wash in 1X to 2X SSC (20X SSC = 3.0 M NaCl/0.3 M trisodium

citrate) at 50 to 55°C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1.0 M NaCl, 1% SDS at 37°C, and a wash in 0.5X to 1X SSC at 55 to 60°C. Exemplary high stringency conditions include hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.1X SSC at 60 to 65°C. Optionally, wash buffers may comprise about 0.1% to about 1% SDS. Duration of hybridization is generally less than about 24 hours, usually about 4 to about 12 hours. The duration of the wash time will be at least a length of time sufficient to reach equilibrium.

The present disclosure provides methods of detecting the presence of HSD17B13 Isoform D in a human subject, comprising or consisting of performing an assay on a biological sample obtained from the human subject, wherein the assay determines the presence of HSD17B13 Isoform D in the biological sample. In some embodiments, the HSD17B13 Isoform D comprises or consists of an amino acid sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to SEQ ID NO:42. In some embodiments, the assay comprises sequencing.

The present disclosure provides methods for detecting the presence or quantifying the levels of variant HSD17B13 polypeptide in a biological sample, including, for example, protein sequencing and immunoassays. In some embodiments, the method of detecting the presence of variant HSD17B13 polypeptide in a human subject comprises performing an assay on a biological sample from the human subject that detects the presence of the variant HSD17B13 polypeptide in the biological sample.

Illustrative non-limiting examples of protein sequencing techniques include, but are not limited to, mass spectrometry and Edman degradation. Illustrative examples of immunoassays include, but are not limited to, immunoprecipitation, Western blot, immunohistochemistry, ELISA, immunocytochemistry, flow cytometry, and immuno-PCR. Polyclonal or monoclonal antibodies detectably labeled using various known techniques (*e.g.*, calorimetric, fluorescent, chemiluminescent, or radioactive) are suitable for use in the immunoassays. Regarding immunoassays, the variant HSD17B13 isoforms have different sizes as compared to the HSD17B13 isoforms which correspond to the wild type condition and, therefore, run at a different molecular weights on a protein gel. Thus, by using the same antibody, the HSD17B13 HSD17B13 isoforms which correspond to the wild type condition can be distinguished from the variant HSD17B13 isoforms in, for example, a Western Blot assay.

In some embodiments, the HSD17B13 isoform detected is not displaced from the cell membrane. In some embodiments, the HSD17B13 isoform is a membrane-bound protein. Such

association may aid in the processing of particular biological samples (i.e., obtaining membrane preparation samples).

The present disclosure also provides kits for making the compositions and utilizing the methods described herein. The kits described herein can comprise an assay or assays for detecting one or more genetic variants in a sample of a subject.

In some embodiments, the kits for human identification of HSD17B13 variants utilize the compositions and methods described above. In some embodiments, a basic kit can comprise a container having at least one pair of oligonucleotide primers or probes, such as alteration-specific probes or alteration-specific primers, for hybridization to any of the nucleic acid molecules disclosed herein. A kit can also optionally comprise instructions for use. A kit can also comprise other optional kit components, such as, for example, one or more of an allelic ladder directed to each of the loci amplified, a sufficient quantity of enzyme for amplification, amplification buffer to facilitate the amplification, divalent cation solution to facilitate enzyme activity, dNTPs for strand extension during amplification, loading solution for preparation of the amplified material for electrophoresis, genomic DNA as a template control, a size marker to insure that materials migrate as anticipated in the separation medium, and a protocol and manual to educate the user and limit error in use.

In some embodiments, any of the kits disclosed herein may further comprise any one or more of: a nucleotide ladder, protocol, an enzyme (such as an enzyme used for amplification, such as polymerase chain reaction (PCR)), dNTPs, a buffer, a salt or salts, and a control nucleic acid sample. In some embodiments, any of the kits disclosed herein may further comprise any one or more of: a detectable label, products and reagents required to carry out an annealing reaction, and instructions. In some embodiments, a kit can comprise one or more of the primers or probes disclosed herein. For example, a kit can comprise one or more probes that hybridize to one or more of the disclosed genetic variants. In some embodiments, a kit can comprise one of the disclosed cells or cell lines. A kit can further comprise media for cell culture.

The present disclosure provides methods for determining a subject's susceptibility or risk of developing a liver disease (e.g., a chronic liver disease) or of diagnosing a subject with liver disease (e.g., a fatty liver disease, NAFLD, or simple steatosis) or at risk of developing liver disease. The subject can be any organism, including, for example, a human, a non-human mammal, a rodent, a mouse, or a rat. Such methods can comprise or consist of, for example, detecting the presence of the *HSD17B13* rs72613567 variant gene in a biological sample comprising genomic DNA, detecting the presence or levels of any one of *HSD17B13* Transcripts C, D, F, G, and H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived

therefrom; preferably RNA or cDNA derived therefrom), and particularly D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), in a biological sample comprising RNA, or cDNA derived therefrom, or comprising mRNA, or cDNA derived therefrom, or detecting the presence or levels of any one of HSD17B13 Isoforms C, D, F, G, or H, and particularly D, in a biological sample comprising protein. It is understood that gene sequences within a population and RNAs, mRNAs, and proteins encoded by such genes can vary due to polymorphisms such as single-nucleotide polymorphisms. The sequences provided herein for the *HSD17B13* gene and for each *HSD17B13* Transcript and HSD17B13 Isoform are only exemplary sequences for the *HSD17B13* gene and for each *HSD17B13* Transcript and HSD17B13 isoform. Other sequences for the *HSD17B13* gene and for each *HSD17B13* Transcript and HSD17B13 Isoform are also possible.

In any of the methods or uses described herein, the liver disease can be a chronic liver disease, fatty liver disease, nonalcoholic fatty liver disease (NAFLD), alcoholic liver fatty liver disease, cirrhosis, viral hepatitis, hepatocellular carcinoma, simple steatosis, steatohepatitis, fibrosis, or non-alcoholic steatohepatitis (NASH). In some embodiments, the liver disease is fatty liver disease, NAFLD, or simple steatosis. In some embodiments, the liver disease is a chronic liver disease. In some embodiments, the liver disease is fatty liver disease. In some embodiments, the liver disease is NAFLD. In some embodiments, the liver disease is alcoholic liver fatty liver disease. In some embodiments, the liver disease is fibrosis. In some embodiments, the liver disease is cirrhosis. In some embodiments, the liver disease is viral hepatitis. In some embodiments, the liver disease is NASH. In some embodiments, the liver disease is hepatocellular carcinoma. In some embodiments, the liver disease is simple steatosis. In some embodiments, the liver disease is steatohepatitis. In some embodiments, the liver disease is fibrosis, NASH, or cirrhosis.

Liver diseases, such as chronic liver diseases, include diseases of the liver which last over a period of six months and can include, for example, diseases of the liver involving progressive destruction and regeneration of the liver parenchyma that can lead to fibrosis and cirrhosis. Liver pathologies encompassed by chronic liver diseases can include, for example, inflammation (e.g., chronic hepatitis), liver cirrhosis, and hepatocellular carcinoma. Types of chronic liver disease are disclosed elsewhere herein and include, for example, fatty liver disease, viral hepatitis, nonalcoholic fatty liver disease, alcoholic fatty liver disease, cirrhosis, and hepatocellular carcinoma. Symptoms and signs of chronic liver diseases are known and can include, for example, enlarged liver, fatigue, pain in the upper right abdomen, abdominal swelling (ascites), enlarged blood vessels just beneath the skin's surface, enlarged breasts in

men, enlarged spleen, red palms, and yellowing of the skin and eyes (jaundice). Testing for chronic liver diseases can involve blood tests, imaging of the liver, and biopsy of the liver. An individual is at increased risk of a chronic liver disease if the subject has at least one known risk-factor (e.g., genetic factor such as a disease-causing mutation) placing individuals with that risk factor at a statistically significant greater risk of developing the disease than individuals without the risk factor. Risk factors for chronic liver diseases are also well known and can include, for example, excessive alcohol use, obesity, high cholesterol, high levels of triglycerides in the blood, polycystic ovary syndrome, sleep apnea, type 2 diabetes, underactive thyroid (hypothyroidism), underactive pituitary gland (hypopituitarism), and metabolic syndromes including raised blood lipids.

The present disclosure provides methods of determining a human subject's susceptibility or risk of developing a liver disease, comprising or consisting of: a) performing an assay on a biological sample obtained from the human subject, wherein the assay determines whether a thymine is inserted between positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene, or whether a thymine is present at a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene; and b) classifying the human subject as being at decreased risk for developing the liver disease if a thymine is inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene or if a thymine is present at a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene, or classifying the human subject as being at increased risk for developing the liver disease if a thymine is not inserted between the positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1 or if a thymine is not present at a position of the HSD17B13 gene corresponding to position 12666 of SEQ ID NO:2. In some embodiments, the liver disease is a chronic liver disease. In some embodiments, the liver disease is selected from the group consisting of fatty liver disease, nonalcoholic fatty liver disease (NAFLD), alcoholic liver fatty liver disease, cirrhosis, viral hepatitis, hepatocellular carcinoma, simple steatosis, steatohepatitis, fibrosis, and non-alcoholic steatohepatitis (NASH). In some embodiments, the assay comprises or consists of: i) contacting the biological sample with a primer hybridizing to a region of the HSD17B13 gene that is within 50 nucleotides of positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1, or corresponding to position 12666 of SEQ ID NO:2; ii) extending the primer at least through the positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1, or corresponding to position 12666 of SEQ ID NO:2; and iii) determining whether a thymine is inserted between the positions corresponding to

positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene, or whether a thymine is present at a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene, in an extension product of the primer. In some embodiments, the assay comprises or consists of contacting the biological sample with a primer or probe that specifically hybridizes to the variant HSD17B13 gene having a thymine at a position corresponding to position 12666 of SEQ ID NO:2, and does not hybridize to the corresponding wild type HSD17B13 gene under stringent conditions, and determining whether hybridization has occurred. In some embodiments, the variant HSD17B13 gene is detected by sequencing. In some embodiments, the method further comprises determining whether the human subject is homozygous for the variant HSD17B13 gene.

In some embodiments, the methods comprise or consist of detecting the presence of the *HSD17B13* rs72613567 variant gene in a biological sample comprising genomic DNA. Such methods can comprise or consist of: a) performing an assay on a biological sample comprising an *HSD17B13* gene obtained from the subject, wherein the assay determines the identity of the nucleotide occupying a position of the *HSD17B13* gene corresponding to position 12666 or positions 12666 and 12667 of SEQ ID NO:2; and b) classifying the subject as being at decreased risk for developing the liver disease if the position corresponding to position 12666 of SEQ ID NO:2 is occupied by a thymine or positions 12666 and 12667 of SEQ ID NO:2 are occupied by thymines. Alternatively, the subject can be classified as being at increased risk for developing the liver disease if the position is not occupied by a thymine. Likewise, such methods can comprise performing an assay on the biological sample that determines the identity of the nucleotides occupying positions of the *HSD17B13* gene corresponding to positions 12665 and 12666 of SEQ ID NO:1. The subject can be classified as being at decreased risk for developing the liver disease if a thymine is inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene. Alternatively, the subject can be classified as being at increased risk for developing the chronic liver disease if there is no thymine inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene.

Any assay for determining the identity of the nucleotide occupying a position of the *HSD17B13* gene corresponding to position 12666 or positions 12666 and 12667 of SEQ ID NO:2 (or corresponding to positions 12665 and 12666 of SEQ ID NO:1) can be used. As one example, the assay can comprise or consist of sequencing a portion of the *HSD17B13* gene including a position corresponding to position 12666 or positions 12666 and 12667 of SEQ ID NO:2. Sequencing can comprise or consist of: i) contacting the biological sample with a primer

hybridizing to a segment of the *HSD17B13* gene that is proximate to a position of the *HSD17B13* gene corresponding to position 12666 or positions 12666 and 12667 of SEQ ID NO:2; ii) extending the primer at least through the position of the *HSD17B13* gene corresponding to position 12666 or positions 12666 and 12667 of SEQ ID NO:2; and iii) determining the identity of the position of the *HSD17B13* gene corresponding to position 12666 or positions 12666 and 12667 of SEQ ID NO:2 in an extension product of the primer. As used herein, "proximate" means within about 50, within about 45, within about 40, within about 35, within about 30, within about 25, within about 20, within about 15, within about 10, or within about 5, nucleotides of the particular stated position.

As another example, the assay can comprise or consist of contacting the biological sample with a primer or probe, such as an alteration-specific primer or an alteration-specific probe, that specifically hybridizes to the variant *HSD17B13* rs72613567 gene and not the corresponding wild type *HSD17B13* gene under stringent conditions, and determining whether hybridization has occurred.

The present disclosure provides methods of determining a human subject's susceptibility or risk for developing a liver disease, comprising or consisting of: a) performing an assay on a biological sample obtained from the human subject, wherein the assay determines the presence of an HSD17B13 Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) in the biological sample; and b) classifying the human subject as being at decreased risk for developing the liver disease if an HSD17B13 Transcript D is present in the biological sample, or classifying the human subject as being at increased risk for developing the liver disease if an HSD17B13 Transcript D is not present in the biological sample. In some embodiments, the HSD17B13 Transcript D comprises or consists of a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (at least about 90% identical) to SEQ ID NO:6, 15, 24, or 33. In some embodiments, the HSD17B13 Transcript D is RNA and comprises or consists of SEQ ID NO:6, or a cDNA thereof comprising or consisting of SEQ ID NO:24, or wherein the HSD17B13 Transcript D is mRNA and comprises or consists of SEQ ID NO:15, or a cDNA thereof comprising or consisting of SEQ ID NO:33. In some embodiments, the assay determines the expression level of HSD17B13 Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) in the biological sample, wherein an increased expression level of HSD17B13 Transcript D compared to a control sample from a control human subject homozygous for a wild type HSD17B13 allele indicates a decreased risk for developing the liver

disease, and wherein the same or a decreased expression level of HSD17B13 Transcript D compared to the control sample indicates an increased risk for developing the liver disease. In some embodiments, the liver disease is a chronic liver disease. In some embodiments, the liver disease is selected from the group consisting of fatty liver disease, nonalcoholic fatty liver disease (NAFLD), alcoholic liver fatty liver disease, cirrhosis, viral hepatitis, hepatocellular carcinoma, simple steatosis, steatohepatitis, fibrosis, and non-alcoholic steatohepatitis (NASH). In some embodiments, the assay comprises or consists of contacting the biological sample with one or more primers or probes that specifically hybridize to a nucleic acid sequence of HSD17B13 Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or to the complement thereof, and determining whether hybridization has occurred. In some embodiments, the method further comprises specifically detecting Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) by using a nucleic acid molecule comprising or consisting of from about 5 nucleotides up to about 50 nucleotides comprising or consisting of: i) a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90% identical) to a nucleotide sequence of a SEQ ID NO:6, 15, 24, or 33, or to the complement thereof; ii) a nucleic acid molecule which specifically hybridizes to exon 2 of Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom); and/or iii) a nucleic acid molecule which specifically hybridizes to the region which bridges exons 3 and 4 of Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). In some embodiments, the HSD17B13 Transcript D comprises or consists of a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90% identical) to SEQ ID NO:6, 15, 24, or 33. In some embodiments, the one or more primers or probes specifically hybridize to SEQ ID NO:6, SEQ ID NO:15, SEQ ID NO:24, and/or SEQ ID NO:33. In some embodiments, the assay comprises reverse transcription polymerase chain reaction (RT-PCR) or quantitative RT-PCR (qRT-PCR). In some embodiments, the assay comprises sequencing.

In some embodiments, the methods comprise or consist of: a) performing an assay on a biological sample comprising RNA, or cDNA derived therefrom, or comprising mRNA, or cDNA derived therefrom, obtained from the subject, wherein the assay determines the presence of Transcript C, D, F, G, or H (RNA or cDNA derived therefrom, and/or mRNA or cDNA

derived therefrom; preferably RNA or cDNA derived therefrom) in the biological sample; and b) classifying the subject as being at decreased risk for developing the liver disease if Transcript C, D, F, G, or H is present in the biological sample. Such an assay can, for example, detect a region or combination of regions that is present in (e.g., is unique to) one or more of Transcripts C, D, F, G, and H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) that is not present in Transcripts A and B (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or that is not present in Transcripts A, B, and E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or that is not present in Transcripts A, B, E, and F' (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). Such regions can be readily identified by comparing the sequences of Transcripts A-H and are described in more detail elsewhere herein. Alternatively, the subject can be classified as being at increased risk for developing the liver disease if Transcript C, D, F, G, or H is not present in the biological sample. In a specific example, the assay can determine the expression level of Transcript C, D, F, G, or H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), and particularly Transcript D(RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), in the biological sample, wherein an increased expression level of Transcript C, D, F, G, or H, and particularly Transcript D, in the biological sample compared to a control sample from a control subject homozygous for a wild type *HSD17B13* allele indicates a decreased risk for developing the liver disease. Alternatively, a decreased expression level or no change in expression level of Transcript C, D, F, G, or H, and particularly Transcript D, in the biological sample compared to a control sample from a control subject homozygous for a wild type *HSD17B13* allele indicates an increased risk for developing the liver disease. In another specific example, the assay can comprise determining the expression level of Transcript C, D, F, G, or H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), and particularly Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), relative to Transcript A, B, or E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or Transcript A, B, E, or F' (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), and particularly Transcript A (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived

therefrom), in the biological sample, wherein an increased ratio of Transcript C, D, F, G, or H, and particularly Transcript D, expression relative to Transcript A, B, or E or Transcript A, B, E, or F', and particularly Transcript A, expression compared to the ratio in a control sample from a control subject homozygous for a wild type *HSD17B13* allele indicates a decreased risk for developing the liver disease. Alternatively, a decreased ratio or no change in the ratio of Transcript C, D, F, G, or H, and particularly Transcript D, expression relative to Transcript A, B, or E or Transcript A, B, E, and F', and particularly Transcript A, expression compared to the ratio in a control sample from a control subject homozygous for a wild type *HSD17B13* allele indicates an increased risk for developing the liver disease.

In some methods for detecting the presence or levels of any one of Transcripts C, D, F, G, or H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), and particularly D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), the assay can comprise or consist of contacting the biological sample with one or more primers or probes (e.g., alteration-specific primers or alteration-specific probes) that specifically hybridize to a region spanning the boundary of exons 6 and 7 in Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), Transcript G (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or Transcript H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) (i.e., including the additional guanine at the 3' end of exon 6 that is not present in exon 6 in other *HSD17B13* Transcripts), and determining whether hybridization has occurred. In addition or alternatively, the assay can comprise contacting the biological sample with one or more primers or probes that specifically hybridize to a region within the read-through into intron 6 in Transcript F (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or a region spanning the boundary between the read-through into intron 6 and the rest of exon 6 in Transcript F (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), and determining whether hybridization has occurred. In addition or alternatively, the assay can comprise contacting the biological sample with one or more primers or probes that specifically hybridize to a region spanning the exon 5-exon 7 boundary in Transcript C (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), and determining whether hybridization has occurred.

Other assays that can be used in the methods disclosed herein include, for example, reverse transcription polymerase chain reaction (RT-PCR) or quantitative RT-PCR (qRT-PCR). Yet other assays that can be used in the methods disclosed herein include, for example, RNA sequencing (RNA-Seq) followed by determination of the presence and quantity of Transcript C, D, F, G, or H, and particularly Transcript D, in the biological sample.

Other methods can comprise detecting the presence or levels of any one of *HSD17B13* Transcripts A, B, and E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or Transcripts A, B, E, and F' (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) in a biological sample. Such methods can comprise or consist of: a) performing an assay on a biological sample obtained from the subject, wherein the assay determines the presence of Transcript A, B, or E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or Transcript A, B, E, or F' (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) in the biological sample; and b) classifying the subject as being at increased risk for developing the liver disease if Transcript A, B, or E or Transcript A, B, E, or F' is present in the biological sample. Such an assay can, for example, detect a region or combination of regions that is present in (e.g., is unique to) one or more of Transcripts A, B, or E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or Transcripts A, B, E, or F' (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) that is not present in Transcripts C, D, F, G, and H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). Such regions can be readily identified by comparing the sequences of Transcripts A-H and are described in more detail elsewhere herein. Alternatively, the subject can be classified as being at decreased risk for developing the liver disease if Transcript A, B, or E or Transcript A, B, E, or F' is not present in the biological sample. In a specific example, the assay can determine the expression level of Transcript A, B, or E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or Transcript A, B, E, or F' (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) in the biological sample, wherein an increased expression level of Transcript A, B, or E or Transcript A, B, E, or F' in the biological sample compared to a control sample from a control subject homozygous for the *HSD17B13* rs72613567 variant allele indicates an increased risk for developing the liver disease.

Alternatively, a decreased expression level or no change in expression level of Transcript A, B, or E or Transcript A, B, E, or F' in the biological sample compared to a control sample from a control subject homozygous for the *HSD17B13* rs72613567 variant allele indicates a decreased risk for developing the liver disease. In another specific example, the assay can comprise determining the expression level of Transcript A, B, or E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or Transcript A, B, E, or F' (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) relative to Transcript C, D, F, G, or H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), particularly Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), in the biological sample, wherein an increased ratio of Transcript A, B, or E or Transcript A, B, E, or F' expression relative to Transcript C, D, F, G, or H, particularly Transcript D, expression compared to the ratio in a control sample from a control subject homozygous for the *HSD17B13* rs72613567 variant allele indicates an increased risk for developing the liver disease. Alternatively, a decreased ratio or no change in the ratio of Transcript A, B, or E or Transcript A, B, E, or F' expression relative to Transcript C, D, F, G, or H, particularly Transcript D, expression compared to the ratio in a control sample from a control subject homozygous for the *HSD17B13* rs72613567 variant allele indicates a decreased risk for developing the liver disease.

In some methods for detecting the presence or levels of any one of Transcripts A, B, or E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or Transcripts A, B, E, or F' (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), the assay can comprise or consist of contacting the biological sample with one or more primers or probes, such as alteration-specific primers or alteration-specific probes, that specifically hybridize to a region within exon 3', spanning the boundary of exons 3 and 3', or spanning the boundary of exons 3' and 4 in Transcript E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) to distinguish Transcript E from Transcripts A, B, C, D, F, and G, and further contacting the biological sample with one or more primers and probes, such as alteration-specific primers or alteration-specific probes, that specifically hybridize to a region spanning the boundary of exons 6 and 7 in Transcript E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) to distinguish Transcript E from Transcript H, and determining whether hybridization has occurred. In addition or alternatively,

the assay can comprise or consist of contacting the biological sample with one or more primers or probes, such as alteration-specific primers or alteration-specific probes, that specifically hybridize to a region spanning the boundaries of exon 1 and 3 in Transcript B (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) to distinguish transcript B from Transcripts A, C, D, E, F, and H, and further contacting the biological sample with one or more primers or probes, that specifically hybridize to a region spanning the boundary of exons 6 and 7 in Transcript B (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) to distinguish Transcript B from Transcript G, and determining whether hybridization has occurred. In addition or alternatively, the assay can comprise or consist of contacting the biological sample with one or more primers or probes, such as alteration-specific primers or alteration-specific probes, that specifically hybridize to a region spanning the exon 6-exon 7 boundary in Transcript A (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) to distinguish Transcript A from Transcripts D, F, G, and H, and further contacting the biological sample with one or more primers or probes, such as alteration-specific primers or alteration-specific probes, that specifically hybridize to a region within exon 6, spanning the exon 5-exon 6 boundary, or spanning the exon 6-exon 7 boundary in Transcript A (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) to distinguish Transcript A from Transcript C, and determining whether hybridization has occurred. Optionally, the assay can further comprise contacting the biological sample with one or more primers or probes, such as alteration-specific primers or alteration-specific probes, that specifically hybridize to a region within exon 2, spanning exon 1-exon 2 boundary, or spanning the exon 2-exon 3 boundary in Transcript A (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) to distinguish Transcript A from Transcripts B and H, and further contacting the biological sample with one or more primers or probes, such as alteration-specific primers or alteration-specific probes, that specifically hybridize to a region spanning the exon 3-exon 4 boundary in Transcript A (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) to distinguish Transcript A from Transcripts E and H, and determining whether hybridization has occurred.

Other assays that can be used in the methods disclosed herein include, for example, reverse transcription polymerase chain reaction (RT-PCR) or quantitative RT-PCR (qRT-PCR). Yet other assays that can be used in the methods disclosed herein include, for example, RNA

sequencing (RNA-Seq) followed by determination of the presence and quantity of Transcript A, B, or E or Transcript A, B, E, or F' in the biological sample.

The present disclosure provides methods of determining a human subject's susceptibility or risk for developing a liver disease, comprising or consisting of: a) detecting whether HSD17B13 Isoform D is present in a biological sample obtained from the human subject; and b) classifying the human subject as being at decreased risk for developing the liver disease if HSD17B13 Isoform D is detected in the biological sample, or classifying the human subject as being at decreased risk for developing the liver disease if HSD17B13 Isoform D is not detected in the biological sample. In some embodiments, the HSD17B13 Isoform D comprises or consists of an amino acid sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90% identical) to SEQ ID NO:42. In some embodiments, the liver disease is a chronic liver disease. In some embodiments, the liver disease is selected from the group consisting of fatty liver disease, nonalcoholic fatty liver disease (NAFLD), alcoholic liver fatty liver disease, cirrhosis, viral hepatitis, hepatocellular carcinoma, simple steatosis, steatohepatitis, fibrosis, and non-alcoholic steatohepatitis (NASH). In some embodiments, the detecting comprises sequencing.

In some embodiments, the methods comprise or consist of detecting the presence or levels of any one of HSD17B13 Isoforms C, D, F, G, or H, and particularly D, in a biological sample comprising protein. Such epitopes can be readily identified by comparing the sequences of Isoforms A-H and are described in more detail elsewhere herein. Alternatively, the subject can be classified as being at increased risk for developing the liver disease if Isoform C, D, F, G, or H is not present in the biological sample.

In some embodiments, the detecting determines the expression level of Isoform C, D, F, G, or H in the biological sample, wherein an increased expression level of Isoform C, D, F, G, or H compared to a control sample from a control subject homozygous for a wild type *HSD17B13* allele indicates a decreased risk for developing the liver disease. Alternatively, a decreased expression level or no change in expression level of Isoform C, D, F, G, or H compared to a control sample from a control subject homozygous for a wild type *HSD17B13* allele indicates an increased risk for developing the liver disease.

In some embodiments, the detecting determines the expression level of Isoform A, B, or E or Isoform A, B, E, or F' in the biological sample, wherein an increased expression level of Isoform A, B, or E or Isoform A, B, E, or F' compared to a control sample from a control subject homozygous for the *HSD17B13* rs72613567 variant allele indicates an increased risk for

developing the liver disease. Alternatively, a decreased expression level or no change in expression level of Isoform A, B, or E or Isoform A, B, E, or F' compared to a control sample from a control subject homozygous for the *HSD17B13* rs72613567 variant allele indicates a decreased risk for developing the liver disease.

The present disclosure provides methods of determining a human subject's risk for progression to more clinically advanced stages of fatty liver disease.

The present disclosure provides methods of determining a human subject's risk for progression to more clinically advanced stages of fatty liver disease, comprising or consisting of: a) performing an assay on a biological sample obtained from the human subject, wherein the assay determines whether a thymine is inserted between positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene, or whether a thymine is present at a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene; and b) classifying the human subject as being at decreased risk for progression to more clinically advanced stages of fatty liver disease if a thymine is inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene or if a thymine is present at a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene, or classifying the human subject as being at increased risk for progression to more clinically advanced stages of fatty liver disease if a thymine is not inserted between the positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1, or if a thymine is not present at a position of the HSD17B13 gene corresponding to position 12666 of SEQ ID NO:2. In some embodiments, the assay comprises or consists of: i) contacting the biological sample with a primer hybridizing to a region of the HSD17B13 gene that is within 50 nucleotides of positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1, or corresponding to position 12666 of SEQ ID NO:2; ii) extending the primer at least through the positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1, or corresponding to position 12666 of SEQ ID NO:2; and iii) determining whether a thymine is inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene, or whether a thymine is present at a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene, in an extension product of the primer. In some embodiments, the assay comprises or consists of contacting the biological sample with a primer or probe that specifically hybridizes to the variant HSD17B13 gene having a thymine at a position corresponding to position 12666 of SEQ ID NO:2, and not to the corresponding wild type HSD17B13 gene under stringent conditions, and determining whether hybridization has occurred. In some embodiments,

the variant HSD17B13 gene is detected by sequencing. In some embodiments, the method further comprises determining whether the human subject is homozygous for the variant HSD17B13 gene.

In some embodiments, the methods comprise or consist of a) performing an assay on a biological sample comprising an *HSD17B13* gene obtained from the subject, wherein the assay determines the identity of the nucleotide occupying a position of the *HSD17B13* gene corresponding to position 12666 or positions 12666 and 12667 of SEQ ID NO:2; and b) classifying the subject as being at decreased risk for progression to more clinically advanced stages of the liver disease (e.g., for progression from simple steatosis to one or more of steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma) if the position corresponding to position 12666 of SEQ ID NO:2 is occupied by a thymine or positions 12666 and 12667 of SEQ ID NO:2 are occupied by thymines. Alternatively, the subject can be classified as being at increased risk for progression to more clinically advanced stages of liver disease (e.g., for histopathological progression from simple steatosis to one or more of steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma) if the position is not occupied by a thymine. Likewise, such methods can comprise or consist of performing an assay on the biological sample that determines the identity of the nucleotides occupying positions of the *HSD17B13* gene corresponding to positions 12665 and 12666 of SEQ ID NO:1. The subject can be classified as being at decreased risk for progression to more clinically advanced stages of liver disease (e.g., for histopathological progression from simple steatosis to one or more of steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma) if the thymine is inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene. Alternatively, the subject can be classified as being at increased risk for progression to more clinically advanced stages of liver disease (e.g., for histopathological progression from simple steatosis to one or more of steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma) if there is no thymine inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene.

In some embodiments, if a subject is determined to have HSD17B13 protein isoforms or transcripts for A, B, E, or F', then the subject is at an increased risk for developing fibrosis, which may manifest as late-stage NASH. In contrast, if a subject is determined to have HSD17B13 protein isoforms or transcripts for C, D, F, G, or H, then the subject is at a decreased risk for developing fibrosis. In some embodiments, the histopathologic features of NASH including, for example, lobular inflammation and hepatocyte ballooning, can also be examined in subjects having or suspected of having an increased risk for developing fibrosis. Subjects having

HSD17B13 protein isoforms or transcripts for C, D, F, G, or H, are at a decreased risk for developing lobular inflammation and hepatocyte ballooning.

Any assay, such as the assays described herein, for determining the identity of the nucleotide occupying a position of the *HSD17B13* gene corresponding to position 12666 or positions 12666 and 12667 of SEQ ID NO:2 (or corresponding to positions 12665 and 12666 of SEQ ID NO:1) can be used. In addition, any assay, such as the assays described herein, comprising or consisting of contacting the biological sample with a primer or probe, such as an alteration-specific primer or an alteration-specific probe, that specifically hybridizes to the *HSD17B13* rs72613567 variant and not the corresponding wild type *HSD17B13* sequence under stringent conditions, and determining whether hybridization has occurred.

The present disclosure provides methods of determining a human subject's risk for progression to more clinically advanced stages of fatty liver disease, comprising or consisting of: a) performing an assay on a biological sample obtained from the human subject, wherein the assay determines the presence of an HSD17B13 Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) in the biological sample; and b) classifying the human subject as being at decreased risk for progression to more clinically advanced stages of fatty liver disease if an HSD17B13 Transcript D is present in the biological sample, or classifying the human subject as being at increased risk for progression to more clinically advanced stages of fatty liver disease if an HSD17B13 Transcript D is not present in the biological sample. In some embodiments, the HSD17B13 Transcript D comprises or consists of a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90% identical) to SEQ ID NO:6, 15, 24, or 33. In some embodiments, the HSD17B13 Transcript D is RNA and comprises or consists of SEQ ID NO:6, or a cDNA thereof comprising or consisting of SEQ ID NO:24, or wherein the HSD17B13 Transcript D is mRNA and comprises or consists of SEQ ID NO:15, or a cDNA thereof comprising or consisting of SEQ ID NO:33. In some embodiments, the assay determines the expression level of HSD17B13 Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) in the biological sample, wherein an increased expression level of HSD17B13 Transcript D compared to a control sample from a control human subject homozygous for a wild type HSD17B13 allele indicates a decreased risk for progression to more clinically advanced stages of fatty liver disease, and wherein the same or a decreased expression level of HSD17B13 Transcript D compared to the control sample indicates an increased risk for progression to more clinically advanced stages of

fatty liver disease. In some embodiments, the assay comprises or consists of contacting the biological sample with one or more primers or probes that specifically hybridize to a nucleic acid sequence of HSD17B13 Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or to the complement thereof, and determining whether hybridization has occurred. In some embodiments, the method further comprises specifically detecting Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) by using a nucleic acid molecule comprising from about 5 nucleotides up to about 50 nucleotides comprising or consisting of: i) a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90% identical) to a nucleotide sequence of a SEQ ID NO:6, 15, 24, or 33, or to the complement thereof; ii) a nucleic acid molecule which specifically hybridizes to exon 2 of Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom); and/or iii) a nucleic acid molecule which specifically hybridizes to the region which bridges exons 3 and 4 of Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom). In some embodiments, the one or more primers or probes specifically hybridize to SEQ ID NO:6, SEQ ID NO:15, SEQ ID NO:24, and/or SEQ ID NO:33. In some embodiments, the assay comprises reverse transcription polymerase chain reaction (RT-PCR) or quantitative RT-PCR (qRT-PCR). In some embodiments, the assay comprises sequencing.

In some embodiments, the methods comprise or consist of: performing an assay on a biological sample obtained from the subject, wherein the assay determines the presence of Transcript C, D, F, G, or H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), particularly Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), in the biological sample; and b) classifying the subject as being at decreased risk for progression to more clinically advanced stages of liver disease (e.g., for histopathological progression from simple steatosis to one or more of steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma) if Transcript C, D, F, G, or H, and particularly Transcript D, is present in the biological sample. Alternatively, the subject can be classified as being at increased risk for progression to more clinically advanced stages of liver disease (e.g., for histopathological progression from simple steatosis to one or more of steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma) if Transcript C, D, F, G, or H, particularly Transcript D, is not present in the biological sample. In a specific example, the assay can determine the

expression level of Transcript C, D, F, G, or H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), particularly Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), in the biological sample, wherein an increased expression level of Transcript C, D, F, G, or H, particularly Transcript D, in the biological sample compared to a control sample from a control subject homozygous for a wild type *HSD17B13* allele indicates a decreased risk for progression to more clinically advanced stages of liver disease (e.g., for histopathological progression from simple steatosis to one or more of steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma). Alternatively, a decreased expression level or no change in expression level of Transcript C, D, F, G, or H, particularly Transcript D, in the biological sample compared to a control sample from a control subject homozygous for a wild type *HSD17B13* allele indicates an increased risk for progression to more clinically advanced stages of liver disease (e.g., for histopathological progression from simple steatosis to one or more of steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma). In another specific example, the assay can comprise or consist of determining the expression level of Transcript C, D, F, G, or H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), particularly Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), relative to Transcript A, B, or E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or Transcript A, B, E, or F' (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), particularly Transcript A (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), in the biological sample, wherein an increased ratio of Transcript C, D, F, G, or H, particularly Transcript D, expression relative to Transcript A, B, or E or Transcript A, B, E, or F', particularly Transcript A, expression compared to the ratio in a control sample from a control subject homozygous for a wild type *HSD17B13* allele indicates a decreased risk for progression to more clinically advanced stages of liver disease (e.g., for histopathological progression from simple steatosis to one or more of steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma). Alternatively, a decreased ratio or no change in the ratio of Transcript C, D, F, G, or H, particularly Transcript D, expression relative to Transcript A, B, or E or Transcript A, B, E, or F', particularly Transcript A, expression compared to the ratio in a control sample from a control subject homozygous for a wild type *HSD17B13* allele indicates an increased risk for progression to more clinically advanced stages of liver disease

(e.g., for histopathological progression from simple steatosis to one or more of steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma).

In some embodiments, for detecting the presence or levels of any one of Transcripts C, D, F, G, or H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), and particularly D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), the assay can comprise or consist of contacting the biological sample with one or more primers or probes (e.g., alteration-specific primers or alteration-specific probes) that specifically hybridize to a region spanning the boundary of exons 6 and 7 in Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), Transcript G (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), or Transcript H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) (i.e., including the additional guanine at the 3' end of exon 6 that is not present in exon 6 in other *HSD17B13* Transcripts), and determining whether hybridization has occurred. In addition or alternatively, the assay can comprise or consist of contacting the biological sample with one or more primers or probes that specifically hybridize to a region within the read-through into intron 6 in Transcript F (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or a region spanning the boundary between the read-through into intron 6 and the rest of exon 6 in Transcript F (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), and determining whether hybridization has occurred. In addition or alternatively, the assay can comprise or consist of contacting the biological sample with one or more primers or probes that specifically hybridize to a region spanning the exon 5-exon 7 boundary in Transcript C (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), and determining whether hybridization has occurred.

Other assays that can be used in the methods disclosed herein include, for example, reverse transcription polymerase chain reaction (RT-PCR) or quantitative RT-PCR (qRT-PCR). Yet other assays that can be used in the methods disclosed herein include, for example, RNA sequencing (RNA-Seq) followed by determination of the presence and quantity of Transcript C, D, F, G, or H, and particularly Transcript D, in the biological sample.

In some embodiments, the methods comprise or consist of: a) performing an assay on a biological sample obtained from the subject, wherein the assay determines the presence of

Transcript A, B, or E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or Transcript A, B, E, or F' (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) in the biological sample; and b) classifying the subject as being at increased risk for progression to more clinically advanced stages of liver disease (e.g., for histopathological progression from simple steatosis to one or more of steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma) if Transcript A, B, or E or Transcript A, B, E, or F' is present in the biological sample. Alternatively, the subject can be classified as being at decreased risk for progression to more clinically advanced stages of liver disease (e.g., for histopathological progression from simple steatosis to one or more of steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma) if Transcript A, B, or E or Transcript A, B, E, or F' is not present in the biological sample. In a specific example, the assay can determine the expression level of Transcript A, B, or E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or Transcript A, B, E, or F' (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) in the biological sample, wherein an increased expression level of Transcript A, B, or E or Transcript A, B, E, or F' in the biological sample compared to a control sample from a control subject homozygous for the *HSD17B13* rs72613567 variant allele indicates an increased risk for progression to more clinically advanced stages of liver disease (e.g., for histopathological progression from simple steatosis to one or more of steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma). Alternatively, a decreased expression level or no change in expression level of Transcript A, B, or E or Transcript A, B, E, or F' in the biological sample compared to a control sample from a control subject homozygous for the *HSD17B13* rs72613567 variant allele indicates a decreased risk for progression to more clinically advanced stages of liver disease (e.g., for histopathological progression from simple steatosis to one or more of steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma). In another specific example, the assay can comprise determining the expression level of Transcript A, B, or E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or Transcript A, B, E, or F' (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) relative to Transcript C, D, F, G, or H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), particularly Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), in the biological sample, wherein an increased

ratio of Transcript A, B, or E or Transcript A, B, E, or F' expression relative to Transcript C, D, F, G, or H, particularly Transcript D, expression compared to the ratio in a control sample from a control subject homozygous for the *HSD17B13* rs72613567 variant allele indicates an increased risk for progression to more clinically advanced stages of liver disease (e.g., for histopathological progression from simple steatosis to one or more of steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma). Alternatively, a decreased ratio or no change in the ratio of Transcript A, B, or E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or Transcript A, B, E, or F' (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) expression relative to C, D, F, G, or H (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), particularly Transcript D (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), expression compared to the ratio in a control sample from a control subject homozygous for the *HSD17B13* rs72613567 variant allele indicates a decreased risk for progression to more clinically advanced stages of liver disease (e.g., for histopathological progression from simple steatosis to one or more of steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma).

In some embodiments, for detecting the presence or levels of any one of Transcripts A, B, or E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) or Transcripts A, B, E, or F' (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom), the assay can comprise or consist of contacting the biological sample with one or more primers or probes, such as alteration-specific primers or alteration-specific probes, that specifically hybridize to a region within exon 3', spanning the boundary of exons 3 and 3', or spanning the boundary of exons 3' and 4 in Transcript E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) to distinguish Transcript E from Transcripts A, B, C, D, F, and G, and further contacting the biological sample with one or more primers and probes, such as alteration-specific primers or alteration-specific probes, that specifically hybridize to a region spanning the boundary of exons 6 and 7 in Transcript E (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) to distinguish Transcript E from Transcript H, and determining whether hybridization has occurred. In addition or alternatively, the assay can comprise or consist of contacting the biological sample with one or more primers or probes, such as alteration-specific primers or alteration-specific probes, that specifically

hybridize to a region spanning the boundaries of exon 1 and 3 in Transcript B (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) to distinguish transcript B from Transcripts A, C, D, E, F, and H, and further contacting the biological sample with one or more primers or probes, that specifically hybridize to a region spanning the boundary of exons 6 and 7 in Transcript B (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) to distinguish Transcript B from Transcript G, and determining whether hybridization has occurred. In addition or alternatively, the assay can comprise or consist of contacting the biological sample with one or more primers or probes, such as alteration-specific primers or alteration-specific probes, that specifically hybridize to a region spanning the exon 6-exon 7 boundary in Transcript A (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) to distinguish Transcript A from Transcripts D, F, G, and H, and further contacting the biological sample with one or more primers or probes, such as alteration-specific primers or alteration-specific probes, that specifically hybridize to a region within exon 6, spanning the exon 5-exon 6 boundary, or spanning the exon 6-exon 7 boundary in Transcript A (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) to distinguish Transcript A from Transcript C, and determining whether hybridization has occurred. Optionally, the assay can further comprise or consist of contacting the biological sample with one or more primers or probes, such as alteration-specific primers or alteration-specific probes, that specifically hybridize to a region within exon 2, spanning exon 1-exon 2 boundary, or spanning the exon 2-exon 3 boundary in Transcript A (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) to distinguish Transcript A from Transcripts B and H, and further contacting the biological sample with one or more primers or probes, such as alteration-specific primers or alteration-specific probes, that specifically hybridize to a region spanning the exon 3-exon 4 boundary in Transcript A (RNA or cDNA derived therefrom, and/or mRNA or cDNA derived therefrom; preferably RNA or cDNA derived therefrom) to distinguish Transcript A from Transcripts E and H, and determining whether hybridization has occurred.

Other assays that can be used in the methods disclosed herein include, for example, reverse transcription polymerase chain reaction (RT-PCR) or quantitative RT-PCR (qRT-PCR). Yet other assays that can be used in the methods disclosed herein include, for example, RNA sequencing (RNA-Seq) followed by determination of the presence and quantity of Transcript A, B, or E or Transcript A, B, E, or F' in the biological sample.

The present disclosure provides methods of determining a human subject's risk for progression to more clinically advanced stages of fatty liver disease, comprising or consisting of: a) detecting whether HSD17B13 Isoform D is present in a biological sample obtained from the human subject; and b) classifying the human subject as being at decreased risk for progression to more clinically advanced stages of liver disease if HSD17B13 Isoform D is detected in the biological sample. In some embodiments, the HSD17B13 Isoform D comprises or consists of an amino acid sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical (or at least about 90% identical) to SEQ ID NO:42. In some embodiments, the detecting comprises sequencing.

In some embodiments, the detecting determines the expression level of Isoform C, D, F, G, or H in the biological sample, wherein an increased expression level of Isoform C, D, F, G, or H compared to a control sample from a control human subject homozygous for a wild type *HSD17B13* allele indicates a decreased risk for progression to more clinically advanced stages of liver disease (e.g., for histopathological progression from simple steatosis to one or more of steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma). Alternatively, a decreased expression level or no change in expression level of Isoform C, D, F, G, or H compared to a control sample from a control subject homozygous for a wild type *HSD17B13* allele indicates an increased risk for progression to more clinically advanced stages of liver disease (e.g., for histopathological progression from simple steatosis to one or more of steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma).

In some embodiments, the detecting determines the expression level of Isoform A, B, or E or Isoform A, B, E, or F' in the biological sample, wherein an increased expression level of Isoform A, B, or E or Isoform A, B, E, or F' compared to a control sample from a control human subject homozygous for the *HSD17B13* rs72613567 variant allele indicates an increased risk for progression to more clinically advanced stages of liver disease (e.g., for histopathological progression from simple steatosis to one or more of steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma). Alternatively, a decreased expression level or no change in expression level of Isoform A, B, or E or Isoform A, B, E, or F' compared to a control sample from a control subject homozygous for the *HSD17B13* rs72613567 variant allele indicates a decreased risk for progression to more clinically advanced stages of liver disease (e.g., for histopathological progression from simple steatosis to one or more of steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma).

Any one or more of the methods described herein can be carried out *in vitro*.

In any of the methods disclosed herein, a primer or probe may either hybridize to its

intended target nucleic acid molecule or specifically hybridize to its intended target nucleic acid molecule. In some embodiments, a primer or probe that specifically hybridizes to a particular target does not hybridize to a wild type nucleic acid molecule (such as, for example, SEQ ID NO:1, or a Transcript that has functional activity associated with wild type HSD17B13).

In any of the methods disclosed herein in which a subject is classified as being at increased risk for progression to more clinically advanced stages of liver disease (e.g., for histopathological progression from simple steatosis to one or more of steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma) or as being at increased risk for developing liver disease (e.g., chronic liver disease), the method can further comprise a therapeutic or prophylactic method. Alternatively, the method can further comprise administering a therapeutic agent to prevent or alleviate one or more symptoms associated with progression to more clinically advanced stages of liver disease (e.g., progression from simple steatosis to more clinically advanced stages of liver disease, or progression from simple steatosis to one or more of steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma). For example, such treatments could be focused on preventing or reducing inflammation or preventing or reducing fibrosis. Examples of such therapeutics in development include, but are not limited to obeticholic acid, GS-9674, Simtuzumab, GS-4997, NDI-010976, GFT505/Elafibranor, Aramchol, Cenicriviroc, GR-MD-02, TD139, SHP626, PXS4728A, and RP103 – Cysteamine bitartrate. The present disclosure provides a therapeutic agent selected from the group consisting of the therapeutic agents disclosed above for use in the treatment, prevention or alleviation of a liver disease in a human subject which does not carry an HSD17B13 variant gene. In one aspect, the human subject has been tested negative for a variant HSD17B13 gene. In one aspect, the treatment comprises the step of determining whether or not the human subject carries a variant HSD17B13 gene. In one aspect, the human patient has been determined as having a variant HSD17B13 gene in accordance with any of the methods described herein. The present disclosure also provides a therapeutic agent selected from the group consisting of the therapeutic agents disclosed above for use in the treatment, prevention or alleviation of a fatty liver disease in a human subject which has been determined as having a risk for progression to more clinically advances stages of fatty liver disease in accordance with any of the methods of the present disclosure.

Various methods are provided for modifying a cell through use of any combination of antisense molecules such as antisense RNA, siRNA, and shRNA, HSD17B13 proteins or fragments thereof, and expression vectors for expressing a recombinant *HSD17B13* gene or a nucleic acid encoding an HSD17B13 protein. The methods can occur *in vitro*, *ex vivo*, or *in vivo*. The antisense molecules such as antisense RNA, siRNA, and shRNA, HSD17B13 proteins or

fragments thereof, and expression vectors can be introduced into the cell in any form and by any means as described elsewhere herein, and all or some can be introduced simultaneously or sequentially in any combination.

Antisense molecules can be used to alter expression of an *HSD17B13* gene or a nucleic acid encoding an HSD17B13 isoform protein. Examples of antisense molecules include antisense RNAs, small interfering RNAs (siRNAs), and short hairpin RNAs (shRNAs). Such antisense RNAs, siRNAs, or shRNAs can be designed to target any region of an mRNA. For example, the antisense RNAs, siRNAs, or shRNAs can be designed to target a region unique to one or more of the HSD17B13 transcripts disclosed herein, or a region common to one or more of the HSD17B13 transcripts disclosed herein.

Subjects can be, for example, a subject (e.g., a human) who is not a carrier of the *HSD17B13* rs72613567 variant (or is only a heterozygous carrier of the *HSD17B13* rs72613567 variant) and has or is susceptible to developing a liver disease.

The following representative embodiments are presented:

Embodiment 1. A nucleic acid molecule comprising or consisting of at least 15 contiguous nucleotides of an HSD17B13 gene, and having a thymine inserted between nucleotides at positions corresponding to positions 12665 and 12666 of SEQ ID NO:1.

Embodiment 2. The nucleic acid molecule of embodiment 1, wherein the contiguous nucleotides are at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to a corresponding sequence in SEQ ID NO:2, including the position corresponding to position 12666 of SEQ ID NO:2.

Embodiment 3. The nucleic acid molecule of embodiment 1 or 2, wherein the HSD17B13 gene is a human HSD17B13 gene.

Embodiment 4. The nucleic acid molecule of any one of embodiments 1 to 3, wherein the isolated nucleic acid molecule comprises or consists of at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 200, at least 300, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, at least 1000, at least 2000, at least 3000, at least 4000, at least 5000, at least 6000, at least 7000, at least 8000, at least 9000, at least 10000, at least 11000, at least 12000, at least 13000, at least 14000, at least 15000, at least 16000, at least 17000, at least 18000, or at least 19000 nucleotides corresponding to contiguous nucleotides of SEQ ID NO:2.

Embodiment 5. The nucleic acid molecule of any one of embodiments 1 to 4, wherein the isolated nucleic acid molecule comprises or consists of an HSD17B13 minigene in which one

or more nonessential segments of the gene have been deleted with respect to a corresponding wild type HSD17B13 gene.

Embodiment 6. The nucleic acid molecule of embodiment 5, wherein the deleted segments comprise one or more intronic sequences.

Embodiment 7. The nucleic acid molecule of embodiment 5 or 6, wherein the isolated nucleic acid molecule further comprises an intron corresponding to intron 6 of SEQ ID NO:2.

Embodiment 8. The nucleic acid molecule of embodiment 7, wherein the intron is intron 6 of SEQ ID NO:2.

Embodiment 9. A nucleic acid molecule comprising or consisting of at least 15 contiguous nucleotides encoding all or part of an HSD17B13 protein, wherein the contiguous nucleic acid molecules comprise a segment that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a corresponding segment present in: i) SEQ ID NO:6, SEQ ID NO:15, SEQ ID NO:24, or SEQ ID NO:33 (Transcript D); ii) SEQ ID NO:10, SEQ ID NO:19, SEQ ID NO:28, or SEQ ID NO:37 (Transcript G); or iii) SEQ ID NO:11, SEQ ID NO:20, SEQ ID NO:29, or SEQ ID NO:38 (Transcript H), that is not present in SEQ ID NO:3, SEQ ID NO:12, SEQ ID NO:21, or SEQ ID NO:30 (Transcript A).

Embodiment 10. The nucleic acid molecule of embodiment 9, wherein the contiguous nucleotides further comprise or consists of a segment that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a corresponding segment present in SEQ ID NO:6, SEQ ID NO:15, SEQ ID NO:24, or SEQ ID NO:33 (Transcript D) that is not present in SEQ ID NO:11, SEQ ID NO:20, SEQ ID NO:29, or SEQ ID NO:38 (Transcript H), and wherein the contiguous nucleotides further comprise or consists of a segment that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a corresponding segment present in SEQ ID NO:6, SEQ ID NO:15, SEQ ID NO:24, or SEQ ID NO:33 (Transcript D) that is not present in SEQ ID NO:10, SEQ ID NO:19, SEQ ID NO:28, or SEQ ID NO:37 (Transcript G).

Embodiment 11. The nucleic acid molecule of embodiment 9, wherein the contiguous nucleotides further comprise or consists of a segment that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a corresponding segment present in SEQ ID NO:11, SEQ ID NO:20, SEQ ID NO:29, or SEQ ID NO:38 (Transcript H) that is not present in SEQ ID NO:6, SEQ ID NO:15, SEQ ID NO:24, or SEQ ID NO:33 (Transcript D).

Embodiment 12. The nucleic acid molecule of embodiment 9, wherein the contiguous nucleotides further comprise or consists of a segment that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a corresponding segment present in SEQ ID NO:10, SEQ ID NO:19, SEQ ID NO:28, or SEQ ID NO:37 (Transcript G) that is not present in SEQ ID NO:6, SEQ ID NO:15, SEQ ID NO:24, or SEQ ID NO:33 (Transcript D).

Embodiment 13. An nucleic acid molecule comprising at least 15 contiguous nucleotides encoding all or part of an HSD17B13 protein, wherein the contiguous nucleotides comprise or consists of a segment that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a corresponding segment present in SEQ ID NO:7, SEQ ID NO:16, SEQ ID NO:25, or SEQ ID NO:34 (Transcript E) that is not present in SEQ ID NO:3, SEQ ID NO:12, SEQ ID NO:21, or SEQ ID NO:30 (Transcript A), optionally wherein the contiguous nucleotides further comprise or consists of a segment that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a corresponding segment present in SEQ ID NO:7, SEQ ID NO:16, SEQ ID NO:25, or SEQ ID NO:34 (Transcript E) that is not present in SEQ ID NO:11, SEQ ID NO:20, SEQ ID NO:29, or SEQ ID NO:38 (Transcript H).

Embodiment 14. A nucleic acid molecule comprising or consisting of at least 15 contiguous nucleotides encoding all or part of an HSD17B13 protein, wherein the contiguous nucleotides comprise or consist of a segment that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a corresponding segment present in SEQ ID NO:8, SEQ ID NO:17, SEQ ID NO:26, or SEQ ID NO:35 (Transcript F) that is not present in SEQ ID NO:3, SEQ ID NO:12, SEQ ID NO:21, or SEQ ID NO:30 (Transcript A).

Embodiment 15. A nucleic acid molecule comprising or consisting of at least 15 contiguous nucleotides encoding all or part of an HSD17B13 protein, wherein the contiguous nucleotides comprise or consist of a segment that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a corresponding segment present in SEQ ID NO:5, SEQ ID NO:14, SEQ ID NO:23, or SEQ ID NO:32 (Transcript C) that is not present in SEQ ID NO:3, SEQ ID NO:12, SEQ ID NO:21, or SEQ ID NO:30 (Transcript A).

Embodiment 16. The nucleic acid molecule of any one of embodiments 9 to 15, wherein the HSD17B13 protein is a human HSD17B13 protein.

Embodiment 17. The nucleic acid molecule of any one of embodiments 9 to 16, wherein the isolated nucleic acid molecule comprises or consists of at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 200, at least 300, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, at least 1000, or at least 2000 contiguous nucleotides encoding all or part of an HSD17B13 protein.

Embodiment 18. A nucleic acid molecule comprising or consisting of a sequence at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to the sequence set forth in: i) SEQ ID NO:5, SEQ ID NO:14, SEQ ID NO:23, or SEQ ID NO:32 (Transcript C), ii) SEQ ID NO:6, SEQ ID NO:15, SEQ ID NO:24, or SEQ ID NO:33 (Transcript D), iii) SEQ ID NO:7, SEQ ID NO:16, SEQ ID NO:25, or SEQ ID NO:34 (Transcript E), iv) SEQ ID NO:8, SEQ ID NO:17, SEQ ID NO:26, or SEQ ID NO:35 (Transcript F), v) SEQ ID NO:10, SEQ ID NO:19, SEQ ID NO:28, or SEQ ID NO:37 (Transcript G), or vi) SEQ ID NO:11, SEQ ID NO:20, SEQ ID NO:29, or SEQ ID NO:38 (Transcript H) and encoding an HSD17B13 protein comprising or consisting of the sequence set forth in SEQ ID NO:41 (Isoform C), SEQ ID NO:42 (Isoform D), SEQ ID NO:43 (Isoform E), SEQ ID NO:44 (Isoform F), SEQ ID NO:46 (Isoform G), or SEQ ID NO:47 (Isoform H).

Embodiment 19. The nucleic acid molecule of any one of embodiments 9 to 18, wherein the contiguous nucleotides comprise or consist of sequences from at least two different exons of an HSD17B13 gene without an intervening intron.

Embodiment 20. A polypeptide encoded by the nucleic acid molecule of any one of embodiments 1 to 19.

Embodiment 21. A nucleic acid molecule comprising or consisting of at least 15 contiguous nucleotides that hybridize to an HSD17B13 gene at a segment that includes or is within 1000, 500, 400, 300, 200, 100, 50, 45, 40, 35, 30, 25, 20, 15, 10, or 5 nucleotides of a position corresponding to position 12666 in SEQ ID NO:2.

Embodiment 22. The nucleic acid molecule of embodiment 21, wherein the segment is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to a corresponding sequence in SEQ ID NO:2, and having a thymine at a position corresponding to position 12666 of SEQ ID NO:2.

Embodiment 23. The nucleic acid molecule of embodiment 21 or 22, wherein the segment comprises or consists of at least 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, or 2000 contiguous nucleotides of SEQ ID NO:2.

Embodiment 24. The nucleic acid molecule of any one of embodiments 21 to 23, wherein the segment includes a position corresponding to position 12666 in SEQ ID NO:2.

Embodiment 25. The nucleic acid molecule of any one of embodiments 21 to 24, wherein the HSD17B13 gene is a human HSD17B13 gene.

Embodiment 26. A nucleic acid molecule that hybridizes to at least 15 contiguous nucleotides of a nucleic acid encoding an HSD17B13 protein, wherein the contiguous nucleotides comprise or consist of a segment that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a corresponding segment present in: i) SEQ ID NO:6, SEQ ID NO:15, SEQ ID NO:24, or SEQ ID NO:33 (Transcript D), ii) SEQ ID NO:10, SEQ ID NO:19, SEQ ID NO:28, or SEQ ID NO:37 (Transcript G), or iii) SEQ ID NO:11, SEQ ID NO:20, SEQ ID NO:29, or SEQ ID NO:38 (Transcript H), that is not present within SEQ ID NO:3, SEQ ID NO:12, SEQ ID NO:21, or SEQ ID NO:30 (Transcript A).

Embodiment 27. A nucleic acid molecule that hybridizes to at least 15 contiguous nucleotides of a nucleic acid encoding an HSD17B13 protein, wherein the contiguous nucleotides comprise or consist of a segment that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to a corresponding segment present in SEQ ID NO:7, SEQ ID NO:16, SEQ ID NO:25, or SEQ ID NO:34 (Transcript E) or SEQ ID NO:11, SEQ ID NO:20, SEQ ID NO:29, or SEQ ID NO:38 (Transcript H) that is not present in SEQ ID NO:3, SEQ ID NO:12, SEQ ID NO:21, or SEQ ID NO:30 (Transcript A).

Embodiment 28. A nucleic acid molecule that hybridizes to at least 15 contiguous nucleotides of a nucleic acid encoding an HSD17B13 protein, wherein the contiguous nucleotides comprise or consist of a segment that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to a corresponding segment in SEQ ID NO:8, SEQ ID NO:17, SEQ ID NO:26, or SEQ ID NO:35 (Transcript F) that is not present in SEQ ID NO:3, SEQ ID NO:12, SEQ ID NO:21, or SEQ ID NO:30 (Transcript A).

Embodiment 29. A nucleic acid molecule that hybridizes to at least 15 contiguous nucleotides of a nucleic acid encoding an HSD17B13 protein, wherein the contiguous nucleotides comprise or consist of a segment that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to a corresponding segment present in SEQ ID NO:5, SEQ ID NO:14, SEQ ID NO:23, or SEQ ID

NO:32 (Transcript C) that is not present in SEQ ID NO:3, SEQ ID NO:12, SEQ ID NO:21, or SEQ ID NO:30 (Transcript A).

Embodiment 30. The nucleic acid molecule of any one of embodiments 26 to 29, wherein the HSD17B13 protein is a human HSD17B13 protein.

Embodiment 31. The nucleic acid molecule of any one of embodiments 26 to 29, wherein the isolated nucleic acid is an antisense RNA, a short hairpin RNA, or a small-interfering RNA.

Embodiment 32. The nucleic acid molecule of any one of embodiments 21 to 31, wherein the isolated nucleic acid comprises or consists of from 5 up to about 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 nucleotides in length.

Embodiment 33. The nucleic acid molecule of any one of embodiments 1 to 19, 21 to 30, and 32, wherein the nucleic acid comprises or consists of DNA.

Embodiment 34. The nucleic acid molecule of any one of embodiments 1 to 19 and 21 to 32, wherein the nucleic acid comprises or consists of RNA.

Embodiment 35. The nucleic acid molecule of any one of embodiments 1 to 19 and 21 to 34, wherein the nucleic acid molecule is linked to a heterologous nucleic acid or comprises a heterologous label.

Embodiment 36. The nucleic acid molecule of embodiment 35, wherein the heterologous label is a fluorescent label.

Embodiment 37. A vector or an exogenous donor sequence comprising the nucleic acid molecule of any one of embodiments 1 to 19 and 21 to 36 and a heterologous nucleic acid molecule.

Embodiment 38. The nucleic acid molecule of any one of embodiments 1 to 19 and 21 to 36, wherein the nucleic acid molecule comprises a non-natural nucleotide.

Embodiment 39. Use of a nucleic acid molecule of any one of embodiments 1 to 19 and 21 to 38 in a method of detecting an HSD17B13 rs72613567 variant in a subject, a method of detecting the presence of an HSD17B13 nucleic acid molecule (Transcript C, D, E, F, G, or H) in a subject, a method of determining a subject's susceptibility to developing a liver disease, or a method of diagnosing a subject with a liver disease or at risk of developing a liver disease.

Embodiment 40. A polypeptide comprising or consisting of at least 8 contiguous amino acids of an HSD17B13 protein, wherein the contiguous amino acids comprise or consist of a segment that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a corresponding segment present in

SEQ ID NO:42 (Isoform D), SEQ ID NO:46 (Isoform G), or SEQ ID NO:47 (Isoform H) that is not present in SEQ ID NO:39 (Isoform A).

Embodiment 41. The polypeptide of embodiment 40, wherein the contiguous amino acids further comprise or consist of a segment that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a corresponding segment present in SEQ ID NO:42 (Isoform D) that is not present in SEQ ID NO:47 (Isoform H), and wherein the contiguous amino acids further comprise or consist of a segment that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a corresponding segment present in SEQ ID NO:42 (Isoform D) that is not present in SEQ ID NO:46 (Isoform G).

Embodiment 42. The polypeptide of embodiment 40, wherein the contiguous amino acids further comprise or consist of a segment that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a corresponding segment present in SEQ ID NO:47 (Isoform H) that is not present in SEQ ID NO:42 (Isoform D).

Embodiment 43. The polypeptide of embodiment 40, wherein the contiguous amino acids further comprise or consist of a segment that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a corresponding segment present in SEQ ID NO:46 (Isoform G) that is not present in SEQ ID NO:42 (Isoform D).

Embodiment 44. A polypeptide comprising or consisting of at least 8 contiguous amino acids of an HSD17B13 protein, wherein the contiguous amino acids comprise or consist of a segment that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a corresponding segment present in SEQ ID NO:43 (Isoform E) that is not present in SEQ ID NO:39 (Isoform A), optionally wherein the contiguous amino acids further comprise or consist of a segment that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a corresponding segment present in SEQ ID NO:43 (Isoform E) that is not present in SEQ ID NO:47 (Isoform H).

Embodiment 45. A polypeptide comprising or consisting of at least 8 contiguous amino acids of an HSD17B13 protein, wherein the contiguous amino acids comprise or consist of a segment that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a corresponding segment present in SEQ ID NO:44 (HSD17B13) that is not present in SEQ ID NO:39 (Isoform A).

Embodiment 46. A polypeptide comprising or consisting of at least 8 contiguous amino acids of an HSD17B13 protein, wherein the contiguous amino acids comprise or consist of a segment that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a corresponding segment present in SEQ ID NO:41 (Isoform C) that is not present in SEQ ID NO:39 (Isoform A).

Embodiment 47. A polypeptide comprising or consisting of a sequence at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to SEQ ID NO:41 (Isoform C), SEQ ID NO:42 (Isoform D), SEQ ID NO:43 (Isoform E), SEQ ID NO:44 (Isoform F), SEQ ID NO:46 (Isoform G), or SEQ ID NO:47 (Isoform H).

Embodiment 48. The polypeptide of any one of embodiments 53 to 60 which is linked to a heterologous molecule.

Embodiment 49. The polypeptide of embodiment 48, wherein the heterologous molecule is an immunoglobulin Fc domain, a peptide tag, a transduction domain, poly(ethylene glycol), polysialic acid, or glycolic acid.

Embodiment 50. A nucleic acid molecule encoding the polypeptide of any one of embodiments 53 to 62.

Embodiment 51. A host cell comprising the nucleic acid molecule of embodiment 50 operably linked to a heterologous promoter active in the host cell.

Embodiment 52. The host cell of embodiment 51, wherein the host cell is a bacterial cell, a yeast cell, an insect cell, or a mammalian cell.

Embodiment 53. A method of producing the isolated polypeptide of any one of embodiments 53 to 62, comprising culturing the host cell of embodiment 51 or 52, whereby the nucleic acid molecule is expressed, and recovering the polypeptide.

Embodiment 54. A composition comprising the polypeptide of any one of embodiments 20 and 40 to 49, or the nucleic acid molecule of any one of embodiments 1 to 19, 21 to 36, 38, and 50, the vector of embodiment 37, or the host cell of embodiment 51 and 52, and a carrier.

Embodiment 55. The composition of embodiment 54, wherein the carrier comprises a poly(lactic acid) (PLA) microsphere, a poly(D,L-lactic-coglycolic-acid) (PLGA) microsphere, a liposome, a micelle, an inverse micelle, a lipid cochleate, or a lipid microtubule.

Embodiment 56. A cell comprising the polypeptide of any one of embodiments 20 and 40 to 49, or the nucleic acid of any one of embodiments 1 to 19, 21 to 36, 38, and 50, or the vector of embodiment 37.

Embodiment 57. The cell of embodiment 56, wherein the cell is a human cell.

Embodiment 58. The cell of embodiment 56 or 57, wherein the cell is a liver cell.

Embodiment 59. The cell of embodiment 56, wherein the cell is a rodent cell, a mouse cell, or a rat cell.

Embodiment 60. The cell of embodiment 59, wherein the cell is a pluripotent cell.

Embodiment 61. A method of detecting an HSD17B13 rs72613567 variant in a human subject, comprising performing an assay on a biological sample comprising an HSD17B13 gene obtained from the human subject, wherein the assay determines whether a thymine is inserted between positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene.

Embodiment 62. The method of embodiment 61, wherein the assay comprises or consists of sequencing a portion of the HSD17B13 gene including positions corresponding to positions 12665 and 12666 of SEQ ID NO:1.

Embodiment 63. The method of embodiment 62, wherein the assay comprises or consists of: i) contacting the biological sample with a primer hybridizing to a segment of the HSD17B13 gene that is within about 50 nucleotides of a position of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1; ii) extending the primer at least through the position of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1; and iii) determining whether a thymine is inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene in an extension product of the primer.

Embodiment 64. The method of embodiment 61, wherein the assay comprises or consists of contacting the biological sample with a primer or probe that specifically hybridizes to the HSD17B13 rs72613567 variant and not the corresponding wild type HSD17B13 sequence under stringent conditions, and determining whether hybridization has occurred.

Embodiment 65. A method of detecting the presence of one or more of HSD17B13 Transcripts C, D, E, F, G, and H in a human subject, comprising or consisting of performing an assay on a biological sample comprising mRNA or cDNA obtained from the subject, wherein the assay determines the presence of one or more of HSD17B13 Transcripts C, D, E, F, G, and H in the biological sample.

Embodiment 66. The method of embodiment 65, wherein the assay comprises or consists of contacting the biological sample with one or more primers or probes that specifically hybridize to one or more sequences, the combination of which is unique to one or more of: i) SEQ ID NO:5, SEQ ID NO:14, SEQ ID NO:23, or SEQ ID NO:32 (Transcript C), ii) SEQ ID NO:6, SEQ ID NO:15, SEQ ID NO:24, or SEQ ID NO:33 (Transcript D), iii) SEQ ID NO:7,

SEQ ID NO:16, SEQ ID NO:25, or SEQ ID NO:34 (Transcript E), iv) SEQ ID NO:8, SEQ ID NO:17, SEQ ID NO:26, or SEQ ID NO:35 (Transcript F), v) SEQ ID NO:10, SEQ ID NO:19, SEQ ID NO:28, or SEQ ID NO:37 (Transcript G), and vi) SEQ ID NO:11, SEQ ID NO:20, SEQ ID NO:29, or SEQ ID NO:38 (Transcript H), and determining whether hybridization has occurred.

Embodiment 67. The method of embodiment 66, wherein the assay comprises reverse transcription polymerase chain reaction (RT-PCR).

Embodiment 68. The method of embodiment 65 or 66, wherein the one or more primers or probes specifically hybridize to a region corresponding to a region present in: i) SEQ ID NO:6, SEQ ID NO:15, SEQ ID NO:24, or SEQ ID NO:33 (Transcript D), ii) SEQ ID NO:10, SEQ ID NO:19, SEQ ID NO:28, or SEQ ID NO:37 (Transcript G), or iii) SEQ ID NO:11, SEQ ID NO:20, SEQ ID NO:29, or SEQ ID NO:38 (Transcript H) that is not present in SEQ ID NO:3, SEQ ID NO:12, SEQ ID NO:21, or SEQ ID NO:30 (Transcript A).

Embodiment 69. The method of embodiment 68, wherein the one or more primers or probes further specifically hybridize to a region corresponding to a region present in SEQ ID NO:6, SEQ ID NO:15, SEQ ID NO:24, or SEQ ID NO:33 (Transcript D) that is not present in SEQ ID NO:11, SEQ ID NO:20, SEQ ID NO:29, or SEQ ID NO:38 (Transcript H), and a region present in SEQ ID NO:6, SEQ ID NO:15, SEQ ID NO:24, or SEQ ID NO:33 (Transcript D) that is not present in SEQ ID NO:10, SEQ ID NO:19, SEQ ID NO:28, or SEQ ID NO:37 (Transcript G).

Embodiment 70. The method of embodiment 68, wherein the one or more primers or probes further specifically hybridize to a region corresponding to a region present in SEQ ID NO:11, SEQ ID NO:20, SEQ ID NO:29, or SEQ ID NO:38 (Transcript H) that is not present in SEQ ID NO:6, SEQ ID NO:15, SEQ ID NO:24, or SEQ ID NO:33 (Transcript D).

Embodiment 71. The method of embodiment 68, wherein the one or more primers or probes further specifically hybridize to a region corresponding to a region present in SEQ ID NO:10, SEQ ID NO:19, SEQ ID NO:28, or SEQ ID NO:37 (Transcript G) that is not present in SEQ ID NO:6, SEQ ID NO:15, SEQ ID NO:24, or SEQ ID NO:33 (Transcript D).

Embodiment 72. The method of embodiment 65 or 66, wherein the one or more primers or probes specifically hybridize to a region corresponding to a region present in SEQ ID NO:7, SEQ ID NO:16, SEQ ID NO:25, or SEQ ID NO:34 (Transcript E) that is not present in SEQ ID NO:3, SEQ ID NO:12, SEQ ID NO:21, or SEQ ID NO:30 (Transcript A), optionally wherein the one or more primers or probes further specifically hybridize a region corresponding to a region present in SEQ ID NO:7, SEQ ID NO:16, SEQ ID NO:25, or SEQ ID NO:34 (Transcript E) that

is not present in SEQ ID NO:11, SEQ ID NO:20, SEQ ID NO:29, or SEQ ID NO:38 (Transcript H).

Embodiment 73. The method of embodiment 65 or 66, wherein the one or more primers or probes specifically hybridize to a region corresponding to a region present in SEQ ID NO:8, SEQ ID NO:17, SEQ ID NO:26, or SEQ ID NO:35 (Transcript F) that is not present in SEQ ID NO:3, SEQ ID NO:12, SEQ ID NO:21, or SEQ ID NO:30 (Transcript A).

Embodiment 74. The method of embodiment 65 or 66, wherein the one or more primers or probes specifically hybridize to a region corresponding to a region present in SEQ ID NO:5, SEQ ID NO:14, SEQ ID NO:23, or SEQ ID NO:32 (Transcript C) that is not present in SEQ ID NO:3, SEQ ID NO:12, SEQ ID NO:21, or SEQ ID NO:30 (Transcript A).

Embodiment 75. The method of embodiment 65, wherein the assay comprises RNA sequencing (RNA-Seq).

Embodiment 76. A method of detecting the presence of one or more of HSD17B13 Isoforms C, D, E, F, G, or H in a human subject, comprising or consisting of performing an assay on a biological sample comprising mRNA or cDNA obtained from the human subject, wherein the assay determines the presence of one or more of HSD17B13 Isoforms C, D, E, F, G, or H in the biological sample.

Embodiment 77. A method of determining a human subject's susceptibility to developing a liver disease, comprising or consisting of: a) performing an assay on a biological sample comprising an HSD17B13 gene obtained from the human subject, wherein the assay determines whether a thymine is inserted between positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene; and b) classifying the human subject as being at decreased risk for developing the liver disease if a thymine is inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene, or classifying the human subject as being at increased risk for developing the liver disease if a thymine is not inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene.

Embodiment 78. The method of embodiment 77, wherein the liver disease is selected from the group consisting of fatty liver disease, nonalcoholic fatty liver disease (NAFLD), alcoholic liver fatty liver disease, cirrhosis, viral hepatitis, hepatocellular carcinoma, simple steatosis, steatohepatitis, fibrosis, and non-alcoholic steatohepatitis (NASH).

Embodiment 79. A method of diagnosing a human subject with fatty liver disease, comprising or consisting of: a) performing an assay on a biological sample comprising an HSD17B13 gene obtained from the human subject, wherein the assay determines whether a

thymine is inserted between positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene; and b) classifying the human subject as being at decreased risk for progression to more clinically advanced stages of fatty liver disease if a thymine is inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene, or classifying the human subject as being at increased risk for progression to more clinically advanced stages of fatty liver disease if a thymine is not inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene.

Embodiment 80. The method of embodiment 79, wherein the assay comprises sequencing a portion of the HSD17B13 gene including positions corresponding to positions 12665 and 12666 of SEQ ID NO:1.

Embodiment 81. The method of any one of embodiments 77 to 80, wherein the assay comprises or consists of: i) contacting the biological sample with a primer hybridizing to a segment of the HSD17B13 gene that is within about 50 nucleotides of positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1; ii) extending the primer at least through the positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1; and iii) determining whether a thymine is inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene in an extension product of the primer.

Embodiment 82. The method of any one of embodiments 77 to 80, wherein the assay comprises or consists of contacting the biological sample with a primer or probe that specifically hybridizes to the HSD17B13 rs72613567 variant and not the corresponding wild type HSD17B13 sequence under stringent conditions, and determining whether hybridization has occurred.

Embodiment 83. A method of determining a human subject's susceptibility to developing a liver disease, comprising or consisting of: a) performing an assay on a biological sample comprising an HSD17B13 mRNA or cDNA obtained from the human subject, wherein the assay determines the presence of one or more of HSD17B13 Transcripts C, D, F, G, and H in the biological sample; and b) classifying the human subject as being at decreased risk for developing the liver disease if HSD17B13 Transcript C, D, F, G, or H is present in the biological sample, or classifying the human subject as being at increased risk for developing the liver disease if HSD17B13 Transcript C, D, F, G, or H is not present in the biological sample.

Embodiment 84. The method of embodiment 83, wherein the assay in step a) determines the presence of HSD17B13 Transcript D.

Embodiment 85. The method of embodiment 83, wherein the assay in step a) determines the expression level of one or more of HSD17B13 Transcripts C, D, F, G, and H in the biological sample, wherein an increased expression level of HSD17B13 Transcript C, D, F, G, or H compared to a control sample from a control human subject homozygous for a wild type HSD17B13 allele indicates a decreased risk for developing the liver disease, and wherein the same or a decreased expression level of HSD17B13 Transcript C, D, F, G, or H compared to the control sample indicates an increased risk for developing the liver disease.

Embodiment 86. The method of embodiment 83, wherein the assay in step a) determines the expression level of one or more of HSD17B13 Transcripts C, D, F, G, and H relative to HSD17B13 Transcript A, B, or E in the biological sample, wherein an increased ratio of HSD17B13 Transcript C, D, F, G, or H expression relative to HSD17B13 Transcript A, B, or E expression compared to the ratio in a control sample from a control human subject homozygous for a wild type HSD17B13 allele indicates a decreased risk for developing the liver disease, or wherein the same or a decreased ratio of HSD17B13 Transcript C, D, F, G, or H expression relative to HSD17B13 Transcript A, B, or E expression compared to the ratio in the control sample indicates an increased risk for developing the liver disease.

Embodiment 87. The method of embodiment 86, wherein the assay in step a) determines the expression level of HSD17B13 Transcript D relative to the expression level of HSD17B13 Transcript A in the biological sample.

Embodiment 88. The method of any one of embodiments 82 to 87, wherein the liver disease is selected from the group consisting of fatty liver disease, nonalcoholic fatty liver disease (NAFLD), alcoholic liver fatty liver disease, cirrhosis, viral hepatitis, hepatocellular carcinoma, simple steatosis, steatohepatitis, fibrosis, and non-alcoholic steatohepatitis (NASH).

Embodiment 89. A method of diagnosing a human subject with fatty liver disease, comprising or consisting of: a) performing an assay on a biological sample comprising an HSD17B13 mRNA or cDNA obtained from the human subject, wherein the assay determines the presence of one or more of HSD17B13 Transcripts C, D, F, G, and H in the biological sample; and b) classifying the human subject as being at decreased risk for progression to more clinically advanced stages of fatty liver disease if HSD17B13 Transcript C, D, F, G, or H is present in the biological sample, or classifying the human subject as being at increased risk for progression to more clinically advanced stages of fatty liver disease if HSD17B13 Transcript C, D, F, G, or H is not present in the biological sample.

Embodiment 90. The method of embodiment 89, wherein the assay in step a) determines the presence of HSD17B13 Transcript D.

Embodiment 91. The method of embodiment 89, wherein the assay in step a) determines the expression level of one or more of HSD17B13 Transcripts C, D, F, G, and H in the biological sample, wherein an increased expression level of HSD17B13 Transcript C, D, F, G, or H compared to a control sample from a control human subject homozygous for a wild type HSD17B13 allele indicates a decreased risk for progression to more clinically advanced stages of liver disease, or wherein the same or a decreased expression level of HSD17B13 Transcript C, D, F, G, or H compared to the control sample indicates an increased risk for progression to more clinically advanced stages of liver disease.

Embodiment 92. The method of embodiment 89, wherein the assay in step a) determines the expression level of one or more of HSD17B13 Transcripts C, D, F, G, and H relative to HSD17B13 Transcript A, B, or E in the biological sample, wherein an increased ratio of HSD17B13 Transcript C, D, F, G, or H expression relative to HSD17B13 Transcript A, B, or E expression compared to the ratio in a control sample from a control human subject homozygous for a wild type HSD17B13 allele indicates a decreased risk for progression to more clinically advanced stages of liver disease, or wherein the same or a decreased ratio of HSD17B13 Transcript C, D, F, G, or H expression relative to HSD17B13 Transcript A, B, or E expression compared to the ratio in the control sample indicates an increased risk for progression to more clinically advanced stages of liver disease.

Embodiment 93. The method of embodiment 92, wherein the assay in step a) determines the expression level of HSD17B13 Transcript D relative to the expression level of HSD17B13 Transcript A in the biological sample.

Embodiment 94. The method of any one of embodiments 83 to 93, wherein the assay in step a) comprises or consists of contacting the biological sample with one or more primers or probes that specifically hybridize to a segment corresponding to a segment present in: i) SEQ ID NO:6, SEQ ID NO:15, SEQ ID NO:24, or SEQ ID NO:33 (Transcript D); ii) SEQ ID NO:10, SEQ ID NO:19, SEQ ID NO:28, or SEQ ID NO:37 (Transcript G); or SEQ ID NO:11, SEQ ID NO:20, SEQ ID NO:29, or SEQ ID NO:38 (Transcript H) that is not present in SEQ ID NO:3, SEQ ID NO:12, SEQ ID NO:21, or SEQ ID NO:30 (Transcript A), and determining whether hybridization has occurred.

Embodiment 95. The method of any one of embodiments 82 to 93, wherein the assay in step a) comprises or consists of contacting the biological sample with one or more primers or probes that specifically hybridize to a segment corresponding to a segment present within SEQ ID NO:8, SEQ ID NO:17, SEQ ID NO:26, or SEQ ID NO:35 (Transcript F) that is not present in

SEQ ID NO:3, SEQ ID NO:12, SEQ ID NO:21, or SEQ ID NO:30 (Transcript A), and determining whether hybridization has occurred.

Embodiment 96. The method of any one of embodiments 82 to 93, wherein the assay in step a) comprises or consists of contacting the biological sample with one or more primers or probes that specifically hybridize to a segment corresponding to a segment present within SEQ ID NO:5, SEQ ID NO:14, SEQ ID NO:23, or SEQ ID NO:32 (Transcript C) that is not present in SEQ ID NO:3, SEQ ID NO:12, SEQ ID NO:21, or SEQ ID NO:30 (Transcript A), and determining whether hybridization has occurred.

Embodiment 97. The method of any one of embodiments 82 to 96, wherein the assay in step a) comprises reverse transcription polymerase chain reaction (RT-PCR) or quantitative RT-PCR (qRT-PCR).

Embodiment 98. The method of any one of embodiments 85 to 96, wherein the assay in step a) comprises RNA sequencing (RNA-Seq) followed by determination of the presence and quantity of one or more of HSD17B13 Transcripts C, D, F, G, and H in the biological sample.

Embodiment 99. A method of determining a human subject's susceptibility to developing a liver disease, comprising or consisting of: a) detecting whether one or more of HSD17B13 Isoforms C, D, F, G, and H is present in a biological sample comprising proteins obtained from the human subject; and b) classifying the human subject as being at decreased risk for developing the liver disease if HSD17B13 Isoform C, D, F, G, or H is detected in the biological sample.

Embodiment 100. The method of embodiment 99, wherein the liver disease is selected from the group consisting of fatty liver disease, nonalcoholic fatty liver disease (NAFLD), alcoholic liver fatty liver disease, cirrhosis, viral hepatitis, hepatocellular carcinoma, simple steatosis, steatohepatitis, fibrosis, and non-alcoholic steatohepatitis (NASH).

Embodiment 101. The method of embodiments 99 or 100, wherein the detecting in step a) determines the expression level of one or more of HSD17B13 Isoforms C, D, F, G, and H in the biological sample, wherein an increased expression level of HSD17B13 Isoform C, D, F, G, or H compared to a control sample from a control human subject homozygous for a wild type HSD17B13 allele indicates a decreased risk for developing the liver disease.

Embodiment 102. A method of diagnosing a human subject with fatty liver disease, comprising or consisting of: a) detecting whether one or more of HSD17B13 Isoforms C, D, F, G, or H is present in a biological sample comprising proteins obtained from the human subject; and b) classifying the human subject as being at decreased risk for progression to more clinically

advanced stages of liver disease if HSD17B13 Isoform C, D, F, G, or H is detected in the biological sample.

Embodiment 103. The method of embodiment 102, wherein the detecting in step a) determines the expression level of one or more of HSD17B13 Isoforms C, D, F, G, and H in the biological sample, wherein an increased expression level of HSD17B13 Isoform C, D, F, G, or H compared to a control sample from a control human subject homozygous for a wild type HSD17B13 allele indicates a decreased risk for progression to more clinically advanced stages of liver disease.

Embodiment 104. The method of any one of embodiments 77 to 103, further comprising or consisting of administering a therapeutic that prevents or alleviates symptoms associated with progression from simple steatosis to one or more of steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma.

Embodiment 105. An isolated nucleic acid comprising at least 15 contiguous nucleotides of an HSD17B13 gene and having a thymine inserted between nucleotides corresponding to positions 12665 and 12666 of SEQ ID NO:1 when optimally aligned with SEQ ID NO:1.

Embodiment 106. The isolated nucleic acid of embodiment 105, wherein the contiguous nucleotides are at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to a corresponding sequence in SEQ ID NO:2 including position 12666 of SEQ ID NO: 2 when optimally aligned with SEQ ID NO:2.

Embodiment 107. The isolated nucleic acid of embodiment 105 or 106, wherein the HSD17B13 gene is a human HSD17B13 gene.

Embodiment 108. The isolated nucleic acid of any preceding embodiment, wherein the isolated nucleic acid comprises at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 200, at least 300, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, at least 1000, at least 2000, at least 3000, at least 4000, at least 5000, at least 6000, at least 7000, at least 8000, at least 9000, at least 10000, at least 11000, at least 12000, at least 13000, at least 14000, at least 15000, at least 16000, at least 17000, at least 18000, or at least 19000 contiguous nucleotides of SEQ ID NO:2.

Embodiment 109. The isolated nucleic acid of any preceding embodiment, wherein the isolated nucleic acid comprises an HSD17B13 minigene in which one or more nonessential segments of the gene have been deleted with respect to a corresponding wild type HSD17B13 gene.

Embodiment 110. The isolated nucleic acid of embodiment 109, wherein the deleted segments comprise one or more intronic sequences.

Embodiment 111. The isolated nucleic acid of embodiment 109 or 110, wherein the isolated nucleic acid further comprises an intron corresponding to intron 6 of SEQ ID NO:2 when optimally aligned with SEQ ID NO:2.

Embodiment 112. The isolated nucleic acid of embodiment 111, wherein the intron is intron 6 of SEQ ID NO:2.

Embodiment 113. An isolated nucleic acid comprising at least 15 contiguous nucleotides encoding all or part of an HSD17B13 protein, wherein the contiguous nucleic acids comprise a segment that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to a segment present in SEQ ID NO:24 (HSD17B13 Transcript D), SEQ ID NO:28 (HSD17B13 Transcript G), and SEQ ID NO:29 (HSD17B13 Transcript H) that is not present in SEQ ID NO:21 (HSD17B13 Transcript A).

Embodiment 114. The isolated nucleic acid of embodiment 113, wherein the contiguous nucleotides further comprise a segment that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to a segment present in SEQ ID NO:24 (HSD17B13 Transcript D) that is not present SEQ ID NO:29 (HSD17B13 Transcript H), and wherein the contiguous nucleotides further comprise a segment that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to a segment present in SEQ ID NO:24 (HSD17B13 Transcript D) that is not present in SEQ ID NO:28 (HSD17B13 Transcript G).

Embodiment 115. The isolated nucleic acid of embodiment 113, wherein the contiguous nucleotides further comprise a segment that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to a segment present in SEQ ID NO:29 (HSD17B13 Transcript H) that is not present in SEQ ID NO:24 (HSD17B13 Transcript D).

Embodiment 116. The isolated nucleic acid of embodiment 113, wherein the contiguous nucleotides further comprise a segment that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to a segment present in SEQ ID NO:28 (HSD17B13 Transcript G) that is not present in SEQ ID NO:24 (HSD17B13 Transcript D).

Embodiment 117. An isolated nucleic acid comprising at least 15 contiguous nucleotides encoding all or part of an HSD17B13 protein, wherein the contiguous nucleotides comprise a segment that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to a segment present in SEQ ID NO:25 (HSD17B13 Transcript E) that is not present in SEQ ID NO:21 (HSD17B13 Transcript A), optionally wherein the

contiguous nucleotides further comprise a segment that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to a segment present in SEQ ID NO:25 (HSD17B13 Transcript E) that is not present in SEQ ID NO:29 (HSD17B13 Transcript H).

Embodiment 118. An isolated nucleic acid comprising at least 15 contiguous nucleotides encoding all or part of an HSD17B13 protein, wherein the contiguous nucleotides comprise a segment that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to a segment present in SEQ ID NO:26 (HSD17B13 Transcript F) that is not present in SEQ ID NO:21 (HSD17B13 Transcript A).

Embodiment 119. An isolated nucleic acid comprising at least 15 contiguous nucleotides encoding all or part of an HSD17B13 protein, wherein the contiguous nucleotides comprise a segment that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to a segment present in SEQ ID NO:23 (HSD17B13 Transcript C) that is not present in SEQ ID NO:21 (HSD17B13 Transcript A).

Embodiment 120. The isolated nucleic acid of any one of embodiments 113-119, wherein the HSD17B13 protein is a human HSD17B13 protein.

Embodiment 121. The isolated nucleic acid of any one of embodiments 113-120, wherein the isolated nucleic acid comprises at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 200, at least 300, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, at least 1000, or at least 2000 contiguous nucleotides encoding all or part of an HSD17B13 protein.

Embodiment 122. An isolated nucleic acid comprising a sequence at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to the sequence set forth in SEQ ID NO:23, 24, 25, 26, 28, or 29 (HSD17B13 Transcript C, D, E, F, G, or H) and encoding an HSD17B13 protein comprising the sequence set forth in SEQ ID NO:41, 42, 43, 44, 46, or 47 (HSD17B13 Isoform C, D, E, F, G, or H), respectively.

Embodiment 123. The isolated nucleic acid of any one of embodiments 113-122, wherein the contiguous nucleotides comprise sequence from at least two different exons of an HSD17B13 gene without an intervening intron.

Embodiment 124. A protein encoded by the isolated nucleic acid of any preceding embodiment.

Embodiment 125. An isolated nucleic acid comprising at least 15 contiguous nucleotides that hybridize to an HSD17B13 gene at a segment that includes or is within 1000,

500, 400, 300, 200, 100, 50, 45, 40, 35, 30, 25, 20, 15, 10, or 5 nucleotides of a position corresponding to position 12666 in SEQ ID NO:2 when optimally aligned with SEQ ID NO:2.

Embodiment 126. The isolated nucleic acid of embodiment 125, wherein the segment is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to a corresponding sequence in SEQ ID NO:2 when optimally aligned with SEQ ID NO:2.

Embodiment 127. The isolated nucleic acid of embodiment 125 or 126, wherein the segment comprises at least 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, or 2000 contiguous nucleotides of SEQ ID NO:2.

Embodiment 128. The isolated nucleic acid of any one of embodiments 125-127, wherein the segment includes position 12666 in SEQ ID NO:2 or a position corresponding to position 12666 in SEQ ID NO:2 when optimally aligned with SEQ ID NO:2.

Embodiment 129. The isolated nucleic acid of any one of embodiments 125-128, wherein the HSD17B13 gene is a human HSD17B13 gene.

Embodiment 130. An isolated nucleic acid that hybridizes to at least 15 contiguous nucleotides of a nucleic acid encoding an HSD17B13 protein, wherein the contiguous nucleotides comprise a segment that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to a segment present in SEQ ID NO:24 (HSD17B13 Transcript D), SEQ ID NO:28 (HSD17B13 Transcript G), and SEQ ID NO:29 (HSD17B13 Transcript H) that is not present within SEQ ID NO:21 (HSD17B13 Transcript A).

Embodiment 131. An isolated nucleic acid that hybridizes to at least 15 contiguous nucleotides of a nucleic acid encoding an HSD17B13 protein, wherein the contiguous nucleotides comprise a segment that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to a segment present in SEQ ID NO:25 (HSD17B13 Transcript E) and SEQ ID NO:29 (HSD17B13 Transcript H) that is not present in SEQ ID NO:21 (HSD17B13 Transcript A).

Embodiment 132. An isolated nucleic acid that hybridizes to at least 15 contiguous nucleotides of a nucleic acid encoding an HSD17B13 protein, wherein the contiguous nucleotides comprise a segment that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to a segment in SEQ ID NO:26 (HSD17B13 Transcript F) that is not present in SEQ ID NO:21 (HSD17B13 Transcript A).

Embodiment 133. An isolated nucleic acid that hybridizes to at least 15 contiguous nucleotides of a nucleic acid encoding an HSD17B13 protein, wherein the contiguous nucleotides comprise a segment that is at least 90%, at least 95%, at least 96%, at least 97%, at

least 98%, or at least 99% identical to a segment present in SEQ ID NO:23 (HSD17B13 Transcript C) that is not present in SEQ ID NO:21 (HSD17B13 Transcript A).

Embodiment 134. The isolated nucleic acid of any one of embodiments 130-133, wherein the HSD17B13 protein is a human HSD17B13 protein.

Embodiment 135. The isolated nucleic acid of any one of embodiments 130-133, wherein the isolated nucleic acid is an antisense RNA, a short hairpin RNA, or a small-interfering RNA.

Embodiment 136. The isolated nucleic acid of any one of embodiments 125-135, wherein the isolated nucleic acid is up to about 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 nucleotides in length.

Embodiment 137. The isolated nucleic acid of any one of embodiments 105-123, 125-134, and 136, wherein the isolated nucleic acid comprises DNA.

Embodiment 138. The isolated nucleic acid of any one of embodiments 105-123 and 125-136, wherein the isolated nucleic acid comprises RNA.

Embodiment 139. The isolated nucleic acid of any one of embodiments 105-123 and 125-138, wherein the isolated nucleic acid is linked to a heterologous nucleic acid or comprises a heterologous label.

Embodiment 140. The isolated nucleic acid of embodiment 139, wherein the heterologous label is a fluorescent label.

Embodiment 141. A vector comprising the isolated nucleic acid of any one of embodiments 105-123 and 125-140 and a heterologous nucleic acid sequence.

Embodiment 142. The isolated nucleic acid of any one of embodiments 105-123 and 125-140, wherein the isolated nucleic acid includes a non-natural nucleotide.

Embodiment 143. Use of the isolated nucleic acid of any one of embodiments 105-123 and 125-142 in a method of detecting an HSD17B13 rs72613567 variant in a subject, a method of detecting the presence of HSD17B13 Transcript C, D, E, F, G, or H in a subject, a method of determining a subject's susceptibility to developing a chronic liver disease, method of diagnosing a subject with fatty liver disease, or a method of modifying an HSD17B13 gene in a cell, a method for altering expression of an HSD17B13 gene in a cell.

Embodiment 144. An isolated polypeptide comprising at least 8 contiguous amino acids of an HSD17B13 protein, wherein the contiguous amino acids comprise a segment that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to a segment present in SEQ ID NO:42 (HSD17B13 Isoform D), SEQ ID NO:46 (HSD17B13

Isoform G), and SEQ ID NO:47 (HSD17B13 Isoform H) that is not present in SEQ ID NO:39 (HSD17B13 Isoform A).

Embodiment 145. The isolated polypeptide of embodiment 144, wherein the contiguous amino acids further comprise a segment that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to a segment present in SEQ ID NO:42 (HSD17B13 Isoform D) that is not present in SEQ ID NO:47 (HSD17B13 Isoform H), and wherein the contiguous amino acids further comprise a segment that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to a segment present in SEQ ID NO:42 (HSD17B13 Isoform D) that is not present in SEQ ID NO:46 (HSD17B13 Isoform G).

Embodiment 146. The isolated polypeptide of embodiment 144, wherein the contiguous amino acids further comprise a segment that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to a segment present in SEQ ID NO:47 (HSD17B13 Isoform H) that is not present in SEQ ID NO:42 (HSD17B13 Isoform D).

Embodiment 147. The isolated polypeptide of embodiment 144, wherein the contiguous amino acids further comprise a segment that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to a segment present in SEQ ID NO:46 (HSD17B13 Isoform G) that is not present in SEQ ID NO:42 (HSD17B13 Isoform D).

Embodiment 148. An isolated polypeptide comprising at least 8 contiguous amino acids of an HSD17B13 protein, wherein the contiguous amino acids comprise a segment that is least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to a segment present in SEQ ID NO:43 (HSD17B13 Isoform E) that is not present in SEQ ID NO:39 (HSD17B13 Isoform A), optionally wherein the contiguous amino acids further comprise a segment that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to a segment present in SEQ ID NO:43 (HSD17B13 Isoform E) that is not present in SEQ ID NO:47 (HSD17B13 Isoform H).

Embodiment 149. An isolated polypeptide comprising at least 8 contiguous amino acids of an HSD17B13 protein, wherein the contiguous amino acids comprise a segment that is least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to a segment present in SEQ ID NO:44 (HSD17B13 Isoform F) that is not present in SEQ ID NO:39 (HSD17B13 Isoform A).

Embodiment 150. An isolated polypeptide comprising at least 8 contiguous amino acids of an HSD17B13 protein, wherein the contiguous amino acids comprise a segment that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to a

segment present in SEQ ID NO:41 (HSD17B13 Isoform C) that is not present in SEQ ID NO:39 (HSD17B13 Isoform A).

Embodiment 151. An isolated polypeptide comprising a sequence at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:41, 42, 43, 44, 46, or 47 (HSD17B13 Isoform C, D, E, F, G, or H) when optimally aligned with SEQ ID NO:41, 42, 43, 44, 46, or 47 (HSD17B13 Isoform C, D, E, F, G, or H), respectively.

Embodiment 152. The isolated polypeptide of any one of embodiments 144-151 which is linked to a heterologous molecule.

Embodiment 153. The isolated polypeptide of embodiment 152, wherein the heterologous molecule is an immunoglobulin Fc domain, a peptide tag, a transduction domain, poly(ethylene glycol), polysialic acid, or glycolic acid.

Embodiment 154. An isolated nucleic acid encoding the isolated polypeptide of any one of embodiments 144-153.

Embodiment 155. A host cell comprising the isolated nucleic acid of embodiment 154 operably linked to a heterologous promoter active in the host cell.

Embodiment 156. The host cell of embodiment 155, wherein the host cell is a bacterial cell, a yeast cell, an insect cell, or a mammalian cell.

Embodiment 157. A method of producing the isolated polypeptide of any one of embodiments 144-153, comprising culturing the host cell of embodiment 155 or 156, whereby the nucleic acid is expressed, and recovering the isolated polypeptide.

Embodiment 158. A composition comprising the isolated polypeptide of any one of embodiments 124 and 144-153, or the isolated nucleic acid of any one of embodiments 105-123, 125-140, 142, and 154, the vector of embodiment 141, and a carrier increasing the stability of the isolated polypeptide, the isolated nucleic acid, or the vector.

Embodiment 159. The composition of embodiment 158, wherein the carrier comprises a poly(lactic acid) (PLA) microsphere, a poly(D,L-lactic-coglycolic-acid) (PLGA) microsphere, a liposome, a micelle, an inverse micelle, a lipid cochleate, or a lipid microtubule.

Embodiment 160. A cell comprising the isolated polypeptide of any one of embodiments 124 and 144-153, or the isolated nucleic acid of any one of embodiments 105-123, 125-140, 142, and 154, or the vector of embodiment 141.

Embodiment 161. The cell of embodiment 160, wherein the cell is a human cell.

Embodiment 162. The cell of embodiment 160 or 161, wherein the cell is a liver cell.

Embodiment 163. The cell of embodiment 160, wherein the cell is a rodent cell, a mouse cell, or a rat cell.

Embodiment 164. The cell of embodiment 163, wherein the cell is a pluripotent cell.

Embodiment 165. A method of detecting an HSD17B13 rs72613567 variant in a human subject, comprising: (a) obtaining a biological sample from the human subject comprising an HSD17B13 gene; and (b) performing an assay on the biological sample that determines that a thymine is inserted between positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1 when the HSD17B13 gene and SEQ ID NO:1 are optimally aligned.

Embodiment 166. The method of embodiment 165, wherein the assay comprises sequencing a portion of the HSD17B13 gene including positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 when the HSD17B13 gene and SEQ ID NO:1 are optimally aligned.

Embodiment 167. The method of embodiment 166, wherein the assay comprises: (i) contacting the biological sample with a primer hybridizing to a segment of the HSD17B13 gene that is proximate to a position of the HSD17B13 gene corresponding to positions 12665 and 12666of SEQ ID NO:1 when the HSD17B13 gene and SEQ ID NO:1 are optimally aligned; (ii) extending the primer at least through the position of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1; and (iii) determining the whether a thymine is inserted between the positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1 in an extension product of the primer.

Embodiment 168. The method of embodiment 165, wherein the assay comprises contacting the biological sample with a primer or probe that specifically hybridizes to the HSD17B13 rs72613567 variant and not the corresponding wild type HSD17B13 sequence under stringent conditions, and determining whether hybridization has occurred.

Embodiment 169. A method of detecting the presence of one or more of HSD17B13 Transcripts C, D, E, F, G, and H in a human subject, comprising: (a) obtaining a biological sample from the human subject comprising mRNA or cDNA; and (b) performing an assay on the biological sample that determines the presence of one or more of HSD17B13 Transcripts C, D, E, F, G, and H in the biological sample.

Embodiment 170. The method of embodiment 169, wherein the assay comprises contacting the biological sample with one or more primers or probes that specifically hybridize to one or more sequences, the combination of which is unique to one or more of SEQ ID NOS:23, 24, 25, 26, 28, and 29 (HSD17B13 Transcripts C-H) among SEQ ID NOS:21, 22, 23, 24, 25, 26, 28, and 29 (HSD17B13 Transcripts A-H), and determining whether hybridization has occurred.

Embodiment 171. The method of embodiment 170, wherein the assay comprises reverse transcription polymerase chain reaction (RT-PCR).

Embodiment 172. The method of embodiment 169 or 170, wherein the one or more primers or probes specifically hybridize to a region present in SEQ ID NO:24 (HSD17B13 Transcript D), SEQ ID NO:28 (HSD17B13 Transcript G), and SEQ ID NO:29 (HSD17B13 Transcript H) that is not present in SEQ ID NO:21 (HSD17B13 Transcript A).

Embodiment 173. The method of embodiment 172, wherein the one or more primers or probes further specifically hybridize to a region present in SEQ ID NO:24 (HSD17B13 Transcript D) that is not present in SEQ ID NO:29 (HSD17B13 Transcript H) and a region present in SEQ ID NO:24 (HSD17B13 Transcript D) that is not present in SEQ ID NO:28 (HSD17B13 Transcript G).

Embodiment 174. The method of embodiment 172, wherein the one or more primers or probes further specifically hybridize to a region present in SEQ ID NO:29 (HSD17B13 Transcript H) that is not present in SEQ ID NO:24 (HSD17B13 Transcript D).

Embodiment 175. The method of embodiment 172, wherein the one or more primers or probes further specifically hybridize to a region present in SEQ ID NO:28 (HSD17B13 Transcript G) that is not present in SEQ ID NO:24 (HSD17B13 Transcript D).

Embodiment 176. The method of embodiment 169 or 170, wherein the one or more primers or probes specifically hybridize to a region present in SEQ ID NO:25 (HSD17B13 Transcript E) that is not present in SEQ ID NO:21 (HSD17B13 Transcript A), optionally wherein the one or more primers or probes further specifically hybridize a region present in SEQ ID NO:25 (HSD17B13 Transcript E) that is not present in SEQ ID NO:29 (HSD17B13 Transcript H).

Embodiment 177. The method of embodiment 169 or 170, wherein the one or more primers or probes specifically hybridize to a region present in SEQ ID NO:26 (HSD17B13 Transcript F) that is not present in SEQ ID NO:21 (HSD17B13 Transcript A).

Embodiment 178. The method of embodiment 169 or 170, wherein the one or more primers or probes specifically hybridize to a region present in SEQ ID NO:23 (HSD17B13 Transcript C) that is not present in SEQ ID NO:21 (HSD17B13 Transcript A).

Embodiment 179. The method of embodiment 169, wherein the assay comprises RNA sequencing (RNA-Seq).

Embodiment 180. A method of detecting the presence of one or more of HSD17B13 Isoforms C, D, E, F, G, or H in a human subject, comprising: (a) obtaining a biological sample from the human subject comprising mRNA or cDNA; and (b) performing an assay on the

biological sample that determines the presence of one or more of HSD17B13 Isoforms C, D, E, F, G, or H in the biological sample.

Embodiment 181. A method of determining a human subject's susceptibility to developing a chronic liver disease, comprising: (a) obtaining a biological sample from the human subject comprising an HSD17B13 gene; (b) performing an assay on the biological sample that determines whether a thymine is inserted between positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1 when the HSD17B13 gene and SEQ ID NO:1 are optimally aligned; and (c) classifying the human subject as being at decreased risk for developing the chronic liver disease if a thymine is inserted between the positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1, or classifying the human subject as being at increased risk for developing the chronic liver disease if a thymine is not inserted between the positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1.

Embodiment 182. The method of embodiment 181, wherein the chronic liver disease is selected from the group consisting of nonalcoholic fatty liver disease (NAFLD), alcoholic liver fatty liver disease, cirrhosis, and hepatocellular carcinoma.

Embodiment 183. A method of diagnosing a human subject with fatty liver disease, comprising: (a) obtaining a biological sample from the human subject comprising an HSD17B13 gene; (b) performing an assay on the biological sample that determines whether a thymine is inserted between positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1 when the HSD17B13 gene and SEQ ID NO:1 are optimally aligned; and (c) classifying the human subject as being at decreased risk for progression to more clinically advanced stages of chronic liver disease if a thymine is inserted between the positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1, or classifying the human subject as being at increased risk for progression to more clinically advanced stages of chronic liver disease if a thymine is not inserted between the positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1.

Embodiment 184. The method of embodiment 183, wherein the assay comprises sequencing a portion of the HSD17B13 gene including positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 when the HSD17B13 gene and SEQ ID NO:1 are optimally aligned.

Embodiment 185. The method of any one of embodiments 181-184, wherein the assay comprises: (i) contacting the biological sample with a primer hybridizing to a segment of the HSD17B13 gene that is proximate to positions of the HSD17B13 gene corresponding to

positions 12665 and 12666 of SEQ ID NO:1 when the HSD17B13 gene and SEQ ID NO:1 are optimally aligned; (ii) extending the primer at least through the positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1; and (iii) determining whether a thymine is inserted between the positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1 in an extension product of the primer.

Embodiment 186. The method of any one of embodiments 181-184, wherein the assay comprises contacting the biological sample with a primer or probe that specifically hybridizes to the HSD17B13 rs72613567 variant and not the corresponding wild type HSD17B13 sequence under stringent conditions, and determining whether hybridization has occurred.

Embodiment 187. A method of determining a human subject's susceptibility to developing a chronic liver disease, comprising: (a) obtaining a biological sample from the human subject comprising mRNA or cDNA; (b) performing an assay on the biological sample that determines the presence of one or more of HSD17B13 Transcripts C, D, F, G, and H in the biological sample; and (c) classifying the human subject as being at decreased risk for developing the chronic liver disease if HSD17B13 Transcript C, D, F, G, or H is present in the biological sample, or classifying the human subject as being at increased risk for developing the chronic liver disease if HSD17B13 Transcript C, D, F, G, or H is not present in the biological sample.

Embodiment 188. The method of embodiment 187, wherein the assay in step (b) determines the presence of HSD17B13 Transcript D.

Embodiment 189. The method of embodiment 187, wherein the assay in step (b) determines the expression level of one or more of HSD17B13 Transcripts C, D, F, G, and H in the biological sample, wherein an increased expression level of HSD17B13 Transcript C, D, F, G, or H compared to a control sample from a control human subject homozygous for a wild type HSD17B13 allele indicates a decreased risk for developing the chronic liver disease, and wherein the same or a decreased expression level of HSD17B13 Transcript C, D, F, G, or H compared to the control sample indicates an increased risk for developing the chronic liver disease.

Embodiment 190. The method of embodiment 187, wherein the assay in step (b) determines the expression level of one or more of HSD17B13 Transcripts C, D, F, G, and H relative to HSD17B13 Transcript A, B, or E in the biological sample, wherein an increased ratio of HSD17B13 Transcript C, D, F, G, or H expression relative to HSD17B13 Transcript A, B, or E expression compared to the ratio in a control sample from a control human subject homozygous for a wild type HSD17B13 allele indicates a decreased risk for developing the

chronic liver disease, or wherein the same or a decreased ratio of HSD17B13 Transcript C, D, F, G, or H expression relative to HSD17B13 Transcript A, B, or E expression compared to the ratio in the control sample indicates an increased risk for developing the chronic liver disease.

Embodiment 191. The method of embodiment 190, wherein the assay in step (b) determines the expression level of HSD17B13 Transcript D relative to the expression level of HSD17B13 Transcript A in the biological sample.

Embodiment 192. The method of any one of embodiments 187-191, wherein the chronic liver disease is selected from the group consisting of nonalcoholic fatty liver disease (NAFLD), alcoholic liver fatty liver disease, cirrhosis, and hepatocellular carcinoma.

Embodiment 193. A method of diagnosing a human subject with fatty liver disease, comprising: (a) obtaining a biological sample from the human subject comprising mRNA or cDNA; (b) performing an assay on the biological sample that determines the presence of one or more of HSD17B13 Transcripts C, D, F, G, and H in the biological sample; and (c) classifying the human subject as being at decreased risk for progression to more clinically advanced stages of chronic liver disease if HSD17B13 Transcript C, D, F, G, or H is present in the biological sample, or classifying the human subject as being at increased risk for progression to more clinically advanced stages of chronic liver disease if HSD17B13 Transcript C, D, F, G, or H is not present in the biological sample.

Embodiment 194. The method of embodiment 193, wherein the assay in step (b) determines the presence of HSD17B13 Transcript D.

Embodiment 195. The method of embodiment 193, wherein the assay in step (b) determines the expression level of one or more of HSD17B13 Transcripts C, D, F, G, and H in the biological sample, wherein an increased expression level of HSD17B13 Transcript C, D, F, G, or H compared to a control sample from a control human subject homozygous for a wild type HSD17B13 allele indicates a decreased risk for progression to more clinically advanced stages of chronic liver disease, or wherein the same or a decreased expression level of HSD17B13 Transcript C, D, F, G, or H compared to the control sample indicates an increased risk for progression to more clinically advanced stages of chronic liver disease.

Embodiment 196. The method of embodiment 193, wherein the assay in step (b) determines the expression level of one or more of HSD17B13 Transcripts C, D, F, G, and H relative to HSD17B13 Transcript A, B, or E in the biological sample, wherein an increased ratio of HSD17B13 Transcript C, D, F, G, or H expression relative to HSD17B13 Transcript A, B, or E expression compared to the ratio in a control sample from a control human subject homozygous for a wild type HSD17B13 allele indicates a decreased risk for progression to more

clinically advanced stages of chronic liver disease, or wherein the same or a decreased ratio of HSD17B13 Transcript C, D, F, G, or H expression relative to HSD17B13 Transcript A, B, or E expression compared to the ratio in the control sample indicates an increased risk for progression to more clinically advanced stages of chronic liver disease.

Embodiment 197. The method of embodiment 196, wherein the assay in step (b) determines the expression level of HSD17B13 Transcript D relative to the expression level of HSD17B13 Transcript A in the biological sample.

Embodiment 198. The method of any one of embodiments 187-197, wherein the assay in step (b) comprises contacting the biological sample with one or more primers or probes that specifically hybridize to a segment present in SEQ ID NO:24 (HSD17B13 Transcript D), SEQ ID NO:28 (HSD17B13 Transcript G), and SEQ ID NO:29 (HSD17B13 Transcript H) that is not present in SEQ ID NO:21 (HSD17B13 Transcript A), and determining whether hybridization has occurred.

Embodiment 199. The method of any one of embodiments 187-197, wherein the assay in step (b) comprises contacting the biological sample with one or more primers or probes that specifically hybridize to a segment present within SEQ ID NO:26 (HSD17B13 Transcript F) that is not present in SEQ ID NO:21 (HSD17B13 Transcript A), and determining whether hybridization has occurred.

Embodiment 200. The method of any one of embodiments 187-197, wherein the assay in step (b) comprises contacting the biological sample with one or more primers or probes that specifically hybridize to a segment present within SEQ ID NO:23 (HSD17B13 Transcript C) that is not present in SEQ ID NO:21 (HSD17B13 Transcript A), and determining whether hybridization has occurred.

Embodiment 201. The method of any one of embodiments 187-200, wherein the assay in step (b) comprises reverse transcription polymerase chain reaction (RT-PCR) or quantitative RT-PCR (qRT-PCR).

Embodiment 202. The method of any one of embodiments 187-200, wherein the assay in step (b) comprises RNA sequencing (RNA-Seq) followed by determination of the presence and quantity of one or more of HSD17B13 Transcripts C, D, F, G, and H in the biological sample.

Embodiment 203. A method of determining a human subject's susceptibility to developing a chronic liver disease, comprising: (a) obtaining a biological sample from the human subject comprising proteins; (b) detecting whether one or more of HSD17B13 Isoforms C, D, F, G, and H is present in the biological sample; and (c) classifying the human subject as being at

decreased risk for developing the chronic liver disease if HSD17B13 Isoform C, D, F, G, or H is detected in the biological sample.

Embodiment 204. The method of embodiment 203, wherein the chronic liver disease is selected from the group consisting of nonalcoholic fatty liver disease (NAFLD), alcoholic liver fatty liver disease, cirrhosis, and hepatocellular carcinoma.

Embodiment 205. The method of any one of embodiments 203 or 204, wherein the detecting in step (b) determines the expression level of one or more of HSD17B13 Isoforms C, D, F, G, and H in the biological sample, wherein an increased expression level of HSD17B13 Isoform C, D, F, G, or H compared to a control sample from a control human subject homozygous for a wild type HSD17B13 allele indicates a decreased risk for developing the chronic liver disease.

Embodiment 206. A method of diagnosing a human subject with fatty liver disease, comprising: (a) obtaining a biological sample from the human subject comprising proteins; (b) detecting whether one or more of HSD17B13 Isoforms C, D, F, G, or H is present in the biological sample; and (c) classifying the human subject as being at decreased risk for progression to more clinically advanced stages of chronic liver disease if HSD17B13 Isoform C, D, F, G, or H is detected in the biological sample.

Embodiment 207. The method of embodiment 206, wherein the detecting in step (b) determines the expression level of one or more of HSD17B13 Isoforms C, D, F, G, and H in the biological sample, wherein an increased expression level of HSD17B13 Isoform C, D, F, G, or H compared to a control sample from a control human subject homozygous for a wild type HSD17B13 allele indicates a decreased risk for progression to more clinically advanced stages of chronic liver disease.

Embodiment 208. The method of any one of embodiments 181-207, further comprising: (d) administering a therapeutic that prevents or alleviates symptoms associated with progression from simple steatosis to one or more of steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma, or performing the method of any one of embodiments 233-237.

Embodiment 209. A method for decreasing expression of an HSD17B13 gene in a cell, comprising: contacting the genome of the cell with an antisense RNA, an siRNA, or an shRNA that hybridizes to a sequence within exon 7 of SEQ ID NO:21 (HSD17B13 Transcript A) and decreases expression of HSD17B13 Transcript A.

Embodiment 210. The method of embodiment 209, further comprising introducing an expression vector into the cell, wherein the expression vector comprises a recombinant HSD17B13 gene comprising a thymine inserted between nucleotides corresponding to positions

12665 and 12666 of SEQ ID NO:1 when the recombinant HSD17B13 gene is optimally aligned with SEQ ID NO:1.

Embodiment 211. The method of embodiment 210, wherein the recombinant HSD17B13 gene is a human gene.

Embodiment 212. The method of embodiment 210 or 211, wherein the recombinant HSD17B13 gene is an HSD17B13 minigene in which one or more nonessential segments of the gene have been deleted with respect to a corresponding wild type HSD17B13 gene.

Embodiment 213. The method of embodiment 212, wherein the deleted segments comprise one or more intronic sequences.

Embodiment 214. The method of embodiment 212 or 213, wherein the HSD17B13 minigene comprises an intron corresponding to intron 6 of SEQ ID NO:2 when optimally aligned with SEQ ID NO:2.

Embodiment 215. The method of embodiment 209, further comprising introducing an expression vector into the cell, wherein the expression vector comprises a nucleic acid encoding an HSD17B13 protein that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:42 (HSD17B13 Isoform D).

Embodiment 216. The method of embodiment 215, wherein the nucleic acid encoding the HSD17B13 protein is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:24 (HSD17B13 Transcript D) when optimally aligned with SEQ ID NO:24.

Embodiment 217. The method of embodiment 209, further comprising introducing an HSD17B13 protein or fragment thereof into the cell.

Embodiment 218. The method of embodiment 217, wherein the HSD17B13 protein or fragment thereof is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:42 (HSD17B13 Isoform D).

Embodiment 219. A method for modifying a cell, comprising introducing an expression vector into the cell, wherein the expression vector comprises a recombinant HSD17B13 gene comprising a thymine inserted between nucleotides corresponding to positions 12665 and 12666 of SEQ ID NO:1 when the recombinant HSD17B13 gene is optimally aligned with SEQ ID NO:1.

Embodiment 220. The method of embodiment 219, wherein the recombinant HSD17B13 gene is a human gene.

Embodiment 221. The method of embodiment 219 or 220, wherein the recombinant HSD17B13 gene is an HSD17B13 minigene in which one or more nonessential segments of the gene have been deleted with respect to a corresponding wild type HSD17B13 gene.

Embodiment 222. The method of embodiment 221, wherein the deleted segments comprise one or more intronic sequences.

Embodiment 223. The method of embodiment 221 or 222, wherein the HSD17B13 minigene comprises an intron corresponding to intron 6 of SEQ ID NO:2 when optimally aligned with SEQ ID NO:2.

Embodiment 224. A method for modifying a cell, comprising introducing an expression vector into the cell, wherein the expression vector comprises a nucleic acid encoding an HSD17B13 protein that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:42 (HSD17B13 Isoform D).

Embodiment 225. The method of embodiment 224, wherein the nucleic acid encoding the HSD17B13 protein is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:24 (HSD17B13 Transcript D) when optimally aligned with SEQ ID NO:24.

Embodiment 226. A method for modifying a cell, comprising introducing an HSD17B13 protein or fragment thereof into the cell.

Embodiment 227. The method of embodiment 226, wherein the HSD17B13 protein or fragment thereof is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:42 (HSD17B13 Isoform D).

Embodiment 228. The method of any one of embodiments 209-227, wherein the cell is a rodent cell, a mouse cell, or a rat cell.

Embodiment 229. The method of any one of embodiments 209-227, wherein the cell is a human cell.

Embodiment 230. The method of any one of embodiments 209-228, wherein the cell is a pluripotent cell.

Embodiment 231. The method of any one of embodiments 209-229, wherein the cell is a differentiated cell.

Embodiment 232. The method of embodiment 231, wherein the cell is a liver cell.

Embodiment 233. A method of treating a subject who is not a carrier of the HSD17B13 rs72613567 variant and has or is susceptible to developing a chronic liver disease comprising introducing into the subject: an antisense RNA, an siRNA, or an shRNA that hybridizes to a sequence within exon 7 or a sequence spanning the exon 6-exon 7 boundary of SEQ ID NO:21

(HSD17B13 Transcript A) and decreases expression of HSD17B13 Transcript A in a liver cell in the subject.

Embodiment 234. The method of embodiment 233, further comprising introducing an expression vector into the subject, wherein the expression vector comprises a recombinant HSD17B13 gene comprising a thymine inserted between nucleotides corresponding to positions 12665 and 12666 of SEQ ID NO:1 when the recombinant HSD17B13 gene is optimally aligned with SEQ ID NO:1, wherein the expression vector expresses the recombinant HSD17B13 gene in the liver cell in the subject.

Embodiment 235. The method of embodiment 233, further comprising introducing an expression vector into the subject, wherein the expression vector comprises a nucleic acid encoding an HSD17B13 protein that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:42 (HSD17B13 Isoform D), wherein the expression vector expresses the nucleic acid encoding the HSD17B13 protein in the liver cell in the subject.

Embodiment 236. The method of embodiment 235, wherein the nucleic acid encoding the HSD17B13 protein is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:24 (HSD17B13 Transcript D) when optimally aligned with SEQ ID NO:24.

Embodiment 237. The method of embodiment 233, further comprising introducing a messenger RNA into the subject, wherein the messenger RNA encodes an HSD17B13 protein that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:42 (HSD17B13 Isoform D), wherein the mRNA expresses the HSD17B13 protein in the liver cell in the subject.

Embodiment 238. The method of embodiment 237, wherein a complementary DNA reverse transcribed from the messenger RNA is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:24 (HSD17B13 Transcript D) when optimally aligned with SEQ ID NO:24.

Embodiment 239. The method of embodiment 233, further comprising introducing an HSD17B13 protein or fragment thereof into the subject.

Embodiment 240. The method of embodiment 239, wherein the HSD17B13 protein or fragment thereof is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:42 (HSD17B13 Isoform D).

Embodiment 241. A method of treating a subject who is not a carrier of the HSD17B13 rs72613567 variant and has or is susceptible to developing a chronic liver disease comprising

introducing an expression vector into the subject, wherein the expression vector comprises a recombinant HSD17B13 gene comprising a thymine inserted between nucleotides corresponding to positions 12665 and 12666 of SEQ ID NO:1 when the recombinant HSD17B13 gene is optimally aligned with SEQ ID NO:1, wherein the expression vector expresses the recombinant HSD17B13 gene in a liver cell in the subject.

Embodiment 242. The method of any one of embodiments 234 and 241, wherein the recombinant HSD17B13 gene is a human gene.

Embodiment 243. The method of any one of embodiments 234, 241, and 242, wherein recombinant HSD17B13 gene is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO:2 when optimally aligned with SEQ ID NO:2.

Embodiment 244. The method of any one of embodiments 234, 241, and 242, wherein the recombinant HSD17B13 gene is an HSD17B13 minigene in which one or more nonessential segments of the gene have been deleted with respect to a corresponding wild type HSD17B13 gene.

Embodiment 245. The method of embodiment 244, wherein the deleted segments comprise one or more intronic sequences.

Embodiment 246. The method of embodiment 244 or 245, wherein the HSD17B13 minigene comprises an intron corresponding to intron 6 of SEQ ID NO:2 when optimally aligned with SEQ ID NO:2.

Embodiment 247. A method of treating a subject who is not a carrier of the HSD17B13 rs72613567 variant and has or is susceptible to developing a chronic liver disease comprising introducing an expression vector into the subject, wherein the expression vector comprises a nucleic acid encoding an HSD17B13 protein that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:42 (HSD17B13 Isoform D), wherein the expression vector expresses the nucleic acid encoding the HSD17B13 protein in a liver cell in the subject.

Embodiment 248. The method of embodiment 247, wherein the nucleic acid encoding the HSD17B13 protein is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:24 (HSD17B13 Transcript D) when optimally aligned with SEQ ID NO:24.

Embodiment 249. A method of treating a subject who is not a carrier of the HSD17B13 rs72613567 variant and has or is susceptible to developing a chronic liver disease comprising introducing a messenger RNA into the subject, wherein the messenger RNA encodes an HSD17B13 protein that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at

least 99%, or 100% identical to SEQ ID NO:42 (HSD17B13 Isoform D), wherein the mRNA expresses the HSD17B13 protein in the liver cell in the subject.

Embodiment 250. The method of embodiment 249, wherein a complementary DNA reverse transcribed from the messenger RNA is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:24 (HSD17B13 Transcript D) when optimally aligned with SEQ ID NO:24.

Embodiment 251. A method of treating a subject who is not a carrier of the HSD17B13 rs72613567 variant and has or is susceptible to developing a chronic liver disease comprising introducing an HSD17B13 protein or fragment thereof into the liver of the subject.

Embodiment 252. The method of embodiment 251, wherein the HSD17B13 protein or fragment thereof is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:42 (HSD17B13 Isoform D).

Embodiment 253. The method of any one of embodiments 233-252, wherein the subject is a human.

Embodiment 254. The method of any one of embodiments 233-253, wherein the chronic liver disease is nonalcoholic fatty liver disease (NAFLD), alcoholic liver fatty liver disease, cirrhosis, or hepatocellular carcinoma.

Embodiment 255. The method of any one of embodiments 233-254, wherein the introducing into the subject comprises hydrodynamic delivery, virus-mediated delivery, lipid-nanoparticle-mediated delivery, or intravenous infusion.

All patent filings, websites, other publications, accession numbers and the like cited above or below are incorporated by reference in their entirety for all purposes to the same extent as if each individual item were specifically and individually indicated to be so incorporated by reference. If different versions of a sequence are associated with an accession number at different times, the version associated with the accession number at the effective filing date of this application is meant. The effective filing date means the earlier of the actual filing date or filing date of a priority application referring to the accession number if applicable. Likewise, if different versions of a publication, website or the like are published at different times, the version most recently published at the effective filing date of the application is meant unless otherwise indicated. Any feature, step, element, embodiment, or aspect of the invention can be used in combination with any other unless specifically indicated otherwise. Although the present embodiments have been described in some detail by way of illustration and example for purposes of clarity and understanding, it will be apparent that certain changes and modifications may be practiced within the scope of the appended claims.

The nucleotide and amino acid sequences recited herein are shown using standard letter abbreviations for nucleotide bases, and one-letter code for amino acids. The nucleotide sequences follow the standard convention of beginning at the 5´ end of the sequence and proceeding forward (i.e., from left to right in each line) to the 3´ end. Only one strand of each nucleotide sequence is shown, but the complementary strand is understood to be included by any reference to the displayed strand. The amino acid sequences follow the standard convention of beginning at the amino terminus of the sequence and proceeding forward (i.e., from left to right in each line) to the carboxy terminus.

The following examples are provided to describe the embodiments in greater detail. They are intended to illustrate, not to limit, the claimed embodiments.

Examples

The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how the compounds, compositions, articles, devices and/or methods claimed herein are made and evaluated, and are intended to be purely exemplary and are not intended to limit the scope of what the inventors regard as their invention. Efforts have been made to ensure accuracy with respect to numbers (*e.g.*, amounts, temperature, etc.), but some errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, temperature is in °C or is at ambient temperature, and pressure is at or near atmospheric. The diseases identified in the Examples are preferred.

Example 1: Variant 17Beta-Hydroxysteroid Dehydrogenase 13 Protects Against Chronic Liver Disease

To identify genetic factors contributing to chronic liver disease, we utilized exome sequence data and electronic health records from 46,544 participants in the DiscovEHR human genetics study. We identified genetic variants associated with established biomarkers of hepatic injury (serum alanine aminotransferase (ALT) and aspartate aminotransferase (AST)) to nominate candidates that might be associated with chronic liver disease. Candidate variants replicating in three additional cohorts (12,527 individuals) were subsequently evaluated for association with clinical diagnoses of chronic liver disease in DiscovEHR and two independent cohorts (total of 37,892 individuals). We also examined the association with histopathological severity of liver disease in an independent bariatric surgery cohort (n=2,391 human liver samples).

A splice variant (rs72613567:TA) in HSD17B13, encoding the hepatic lipid droplet

protein 17-beta hydroxysteroid dehydrogenase 13, was reproducibly associated with reduced ALT ($P=4.2x10^{-12}$) and AST ($P=6.2x10^{-10}$) levels. In DiscovEHR, this variant was associated with reduced risk of alcoholic and nonalcoholic liver disease (by 38%, 95% confidence interval (CI) 19%-52%; and by 16%, 95% CI 9%-22%, respectively, for each rs72613567:TA allele) and cirrhosis (by 44%, 95% CI 22-59%; and by 26%, 95% CI 12%-38% for alcoholic and nonalcoholic cirrhosis, respectively, for each rs72613567:TA allele) in an allele dosage-dependent manner; associations were confirmed in two independent cohorts. rs72613567:TA was associated with decreased severity of histological features of nonalcoholic steatohepatitis (NASH) (23% reduction, 95% CI 10%-34% for each rs72613567:TA allele among individuals with fatty liver disease). rs72613567:TA results in an unstable and truncated protein with reduced enzymatic activity against steroid substrates.

A loss-of-function variant in HSD17B13 was associated with reduced risk of alcoholic and nonalcoholic liver disease, and progression from steatosis to NASH.

Study Design and Participants

Human genetics studies were conducted as part of the DiscovEHR collaboration of the Regeneron Genetics Center and Geisinger Health System (GHS). The two DiscovEHR study populations (discovery cohort and bariatric surgery cohort) originated from the first 50,726 consented participants ≥ 18 years of age from the MyCode® Community Health Initiative of GHS. The GHS discovery cohort consisted of 46,544 European individuals recruited from outpatient primary care and specialty clinics between 2007 and 2016, excluding all those recruited to the bariatric surgery cohort. The GHS bariatric surgery cohort consisted of 2,644 European individuals who had been referred for bariatric surgery.

Replication studies of associations with liver transaminases included 1,357 European individuals from the Dallas Heart Study and 8,527 European individuals from the Penn Medicine Biobank. The Dallas Heart Study is a probability-based population cohort study of Dallas County residents aged 30 to 65 years (Victor et al., Am. J. Cardiol., 2004; 93, 1473-80). The Penn Medicine Biobank includes participants recruited from the University of Pennsylvania Health System and consented for biospecimen storage, access to EHR data, and permission to recontact.

Replication studies of the associations with chronic liver disease included 517 individuals from the Dallas Liver Study (DLS) and 447 individuals from the Dallas Pediatric Liver Study (DPLS). The DLS is a biobank of patients with liver disease of non-viral etiology. Recruitment began in January 2015 and is ongoing. Participants were recruited from liver clinics at UT Southwestern and Parkland Health and Hospital System, Dallas. Participants completed a

questionnaire on ethnic/racial background, medical history, lifestyle factors, and family history of liver disease and other diseases. Additional clinical information was extracted from medical records by a trained technician. We included all African American, European American, and Hispanic American patients with DNA available at the time of the present study (n=517) with controls from the Dallas Heart Study. The DPLS is a biobank of Hispanic children recruited from pediatric liver clinics at UT Southwestern and Parkland Health and Hospital System, Dallas, and from an obesity clinic at Children's Medical Center, Dallas. Clinical information was extracted from medical records by a trained technician. As more than 95% of the patients were Hispanic Americans, we only included Hispanic American patients and controls in the present study (n=205 patients and 234 controls).

Clinical Measurements and Chronic Liver Disease Definitions in the Discovery Cohort

Clinical laboratory measurements for ALT and AST were extracted from EHRs of participants from the GHS discovery cohort and bariatric surgery cohort. Median ALT and AST values were calculated for all participants with two or more measurements, and were $\log_{10}$-transformed to normalize the distribution prior to association analyses.

International Classification of Diseases, Ninth Revision (ICD-9) disease diagnosis codes were extracted from EHRs and collapsed into clinical disease categories for non-viral, nonalcoholic (ICD-9 571.40, 571.41, 571.49, 571.5, 571.8, 571.9) or alcoholic (ICD-9 571.0, 571.1, 571.2, 571.3) liver disease case definitions. Additional case definitions based on single diagnosis codes included: alcoholic cirrhosis (ICD-9 571.2), nonalcoholic cirrhosis (ICD-9 571.5), and HCC (ICD-9 155.0). For these case definitions, a common control group without liver disease ("no liver disease") was defined as participants with no case criteria or single-encounter or problem-list diagnosis code indicating any type of liver disease.

Liver Histopathologic Phenotype Definitions in the Bariatric Surgery Cohort

The GHS bariatric surgery cohort consisted of 2,644 individuals of European descent. Wedge biopsies of the liver were obtained intraoperatively during bariatric surgery from 2,391 of these individuals. The biopsies were consistently obtained 10 cm to the left of falciform ligament prior to any liver retraction or surgery on the stomach. The biopsy was divided into sections, with the primary section delivered to the clinical pathologists for liver histology (fixed in 10% neutral buffered formalin and stained with hematoxylin and eosin for routine histology and Masson's trichrome for assessment of fibrosis) and remaining sections stored within a research biobank (frozen in RNAlater and/or liquid nitrogen). Liver histology was conducted by an experienced pathologist and subsequently re-reviewed by a second experienced pathologist using the NASH Clinical Research Network scoring system (Kleiner et al., Hepatology, 2005, 41,

1313-21) as follows: steatosis grade 0 (<5% parenchymal involvement), 1 (5 to <33 %), 2 (34 to <66 %), and 3 (>67 %); lobular inflammation grade 0 (no foci), grade 1 (mild, <2 foci per 200X field), grade 2 (moderate, 2-4 foci per 200X field), grade 3 (severe, >4 foci per 200X field); fibrosis Stage 0 (none), Stage 1 (perisinusoidal or periportal fibrosis), Stage 2 (perisinusoidal and periportal fibrosis), Stage 3 (bridging fibrosis), and Stage 4 (cirrhosis). These histologic diagnoses were used to defined the following phenotypes: 1) Normal: no evidence of steatosis, NASH, or fibrosis; 2) Simple steatosis: Steatosis (regardless of grade) with no evidence of NASH or fibrosis; 3) NASH: Any presence of lobular inflammation or hepatocyte ballooning (regardless of grade), or any presence of fibrosis (regardless of stage); 4) Fibrosis: Any presence of fibrosis (regardless of stage).

Sample Preparation, Sequencing, and Genotyping

DNA sample preparation and whole exome sequencing for the participants in the DiscovEHR study, the Dallas Heart Study, and the Penn Medicine Biobank were performed at the Regeneron Genetics (Dewey et al., Science In Press, 2016). HSD17B13 rs72613567 was genotyped by Taqman assay (and verified by Sanger sequencing in 5 individuals of each genotype) in the Dallas Liver Study and Dallas Pediatric Liver Study.

In particular, exome capture was performed using NimbleGen probes according to the manufacturer's recommended protocol (Roche NimbleGen). The captured DNA was PCR amplified and quantified by qRT-PCR (Kapa Biosystems). The multiplexed samples were sequenced using 75 bp paired-end sequencing on an Illumina v4 HiSeq 2500 to a coverage depth sufficient to provide greater than 20x haploid read depth of over 85% of targeted bases in 96% of samples (approximately 80x mean haploid read depth of targeted bases). Raw sequence data from each Illumina Hiseq 2500 run were uploaded to the DNAnexus platform (Reid et al., BMC Bioinformatics, 2014, 15, 30) for sequence read alignment and variant identification. In brief, raw sequence data were converted from BCL files to sample-specific FASTQ-files, which were aligned to the human reference build GRCh37.p13 with BWA-mem (Li et al., Bioinformatics, 2009, 25, 1754-60). Single nucleotide variants (SNV) and insertion/deletion (indel) sequence variants were identified using the Genome Analysis Toolkit (McKenna et al., Genome Res., 2010, 20, 1297-303).

Exome-wide Association Analysis of Liver Enzymes and Chronic Liver Disease Phenotypes

We used linear mixed models to test 502,219 biallelic variants that had missing data rate of < 1%, Hardy-Weinberg equilibrium P-value > 1.0×10^{-6}, and minor allele frequency > 0.1% for association with transaminase levels. For variants with exome wide significant associations with transaminases ($p < 1 \times 10^{-7}$) in the GHS discovery cohort, we performed association analyses and

meta-analysis, in the European-ancestry replication studies described above. We used a Bonferroni significance threshold determined by the number of variants tested to define replicated associations. Meta-analysis of discovery and replication studies was also performed. All P-values reported in the text correspond to the allelic model.

We subsequently tested transaminase-associated single nucleotide variants for associations with chronic liver disease phenotypes. We used a Bonferroni significance threshold determined by the number of variants and broad chronic liver disease categories tested to determine significance of associations. We further tested replicated novel variants for association with histopathologically defined liver phenotypes from the GHS bariatric surgery cohort. We also performed a phenome-wide study of associations of replicated novel variants with 405 quantitative clinical measurements and 3,168 clinical diagnoses.

In particular, we tested 502,219 biallelic variants with missing data rate < 1%, Hardy-Weinberg equilibrium P-value > 1.0×10^{-6}, and minor allele frequency > 0.1% for association with transaminase levels. Log_{10}-transformed median ALT and AST were adjusted for age, age^2, sex, BMI, and the first four principal components of ancestry. To account for relatedness among study participants, we also fit a genetic relatedness matrix as a random-effects covariate. Both principal components and the genetic relatedness matrix were constructed from 39,858 non-MHC markers in approximate linkage equilibrium and with minor allele frequency > 0.1%. We used linear mixed models as implemented in the GCTA package (Yang et al., Am. J. Hum. Genet., 2011, 88, 76-82) to test for association between trait residuals and single nucleotide variants. All P-values reported in the text correspond to the allelic model.

We attempted to replicate associations in the GHS discovery cohort in three separate European-ancestry cohorts: the GHS bariatric surgery cohort, the Dallas Heart Study, and the Penn Medicine Biobank (described above). ALT and AST measures from the GHS bariatric surgery cohort and from Penn Medicine Biobank were log_{10}-transformed and adjusted for age, age^2, sex, BMI, and the first four principal components of ancestry. Genetic relatedness matrices were included as random-effects covariates, and analysis was performed using linear mixed models in GCTA. In the Dallas Heart study, log_{10}-transformed ALT and AST measures were adjusted for age, age^2, sex, BMI, and the first ten principal components of ancestry, and analysis was performed using linear regression implemented in PLINK. Summary statistics for the three replication cohorts were meta-analyzed using METAL (Willer et al., Bioinformatics, 2010, 26, 2190-1) (replication meta-analysis). Summary statistics for the discovery cohort and the three replication cohorts were meta-analyzed similarly (joint meta-analysis).

Association Analysis with Chronic Liver Disease Phenotypes

We analyzed thirteen significant and replicated single nucleotide variants from the liver enzyme ExWAS for associations with chronic liver disease phenotypes defined from the GHS discovery cohort, as described above. We used a Bonferroni significance threshold of $P<0.05/26$ ($P<1.92x10^{-3}$) to account for the thirteen variants and two broad chronic liver disease categories (alcoholic and nonalcoholic) tested. The HSD17B13 rs72613567 variant was further tested for association with histopathologically defined liver phenotypes from the GHS bariatric surgery cohort, as described above. Odds ratios were estimated with the use of Firth's penalized likelihood method of logistic regression after adjustment for age, age^2, sex, BMI, and the first four principal components of ancestry. Genotypic odds ratios were estimated for HSD17B13 rs72613567 using the same covariates.

Odds ratios for liver disease in the DLS were estimated by logistic regression, adjusted for age, age^2, sex, body mass index, and self-reported ethnicity. Participants from the Dallas Heart Study with available rs72613567 genotypes were used as normal controls (n=4,279). Odds ratios in the DPLS were estimated by logistic regression.

Phenome-Wide Association Study of HSD17B13 rs72613567

We performed a phenome-wide study of associations of HSD17B13 rs72613567 with 405 quantitative EHR-derived anthropometric, vital sign, laboratory, electrocardiographic, echocardiographic, and bone densitometry measurements, and also with 3,168 EHR-derived clinical diagnoses. Median laboratory values for individuals with serial outpatient measures were calculated following removal of likely spurious values that were > 3 standard deviations from the intra-individual median value; maximum and minimum values were also calculated. We then calculated trait residuals for all laboratory traits after adjustment for age, age^2, sex, and the first ten principal components of ancestry, and applied appropriate transformations prior to association analysis. ICD-9 based diagnosis codes were collapsed to hierarchical clinical disease groups and corresponding controls using a modified version of the groupings proposed by Denny et al (Denny et al., Nature Biotechnology, 2013, 31, 1102-10; and Denny et al., Bioinformatics, 2010, 26, 1205-10). ICD-9 based diagnoses required one or more of the following: a problem list entry of the diagnosis code or an encounter diagnosis code entered for two separate clinical encounters on separate calendar days.

Analyses of association with transformed quantitative clinical measurement residuals were performed using linear regression, and analyses of association with clinical diagnoses were performed using logistic regression adjusted for age, age^2, sex, and the first four principal components. Alleles were coded using both additive (0 for reference allele homozygotes, 1 for heterozygotes, and 2 for alternative allele homozygotes) and recessive (0 for reference allele

homozygotes and heterozygotes, 1 for alternative allele homozygotes) models.

Software

Genetic association analyses were performed using GCTA software, version 1.25.07 and PLINK, version 1.9.0. Quantile-quantile and Manhattan plots were generated using R software, version 3.2.1 (R Project for Statistical Computing). Regional association plots were generated using LocusZoom (Pruim et al., Bioinformatics, 2010, 26, 2336-7).

RNA Sequencing Studies

RNA quality and concentration was evaluated by running total RNA on an Agilent RNA Nano Bioanalyzer chip; all samples had an RNA integrity number (RIN) greater than 8. Polyadenlylated RNA transcripts were isolated using two rounds of enrichment with oligo(dT)25 beads (Thermo Fisher Scientific). Samples were purified and concentrated with RNAclean XP beads (Beckman Coulter) and heat-fragmented to approximately 140 base pairs. First-strand synthesis was completed with SuperScript III reverse transcriptase (Thermo Fisher Scientific) using random hexamers; dTTP was replaced with dUTP during second-strand synthesis. Samples were processed according to our standard DNA library preparation method referenced above for exomes with the addition of a uracil DNA-glycosylase step to generate strand-specific sequencing libraries.

Identification and Validation of Novel HSD17B13 Transcripts

Reads were mapped to the Human.B38 using ArrayStudio® software (OmicSoft®, Cary, NC) allowing two mismatches. Two approaches were employed to identify novel HSD17B13 transcripts. Novel exon junctions were discovered based on Gencode v24 using ArrayStudio. *De novo* transcript assembly was carried out using Trinity (v2.2.0) in default setting. Custom gene models were built to incorporate novel transcripts of HSD17B13, and transcript quantification was estimated by read alignment to the custom gene model. Protein sequence alignment of all identified HSD17B13 isoforms is shown in Figures 7A and 7B. RT-PCR was performed on total RNA from human liver samples was performed using the SuperScriptTM One-Step RT-PCR System with PlatinumTM Taq DNA Polymerase (Thermo Fisher). Each 50 µL RT-PCR reaction contained 1X Reaction Mix, 500 nM each forward and reverse primers (PST516: ATGAACATCATCCTAGAAATCCTTC (SEQ ID NO:48) and PST517: ATCATGCATACATCTCTGGCTGGAG (SEQ ID NO:49)), 1 µL of RT/Platinum Taq, and 75 ng RNA. Cycling conditions were: one cycle of 45°C for 30 minutes; one cycle of 94°C for 2 minutes; 40 cycles of 94°C for 20 seconds, 53°C for 30 seconds, and 72°C for 90 seconds; one cycle of 72°C for 5 minutes; then a 10°C hold. Products were purified using the QIAquick PCR Purification Kit (Qiagen) and submitted for direct Sanger sequencing using the

primer DE002 (ATCAGAACTTC AGGCCTTGG (SEQ ID NO:50)). To identify the B and C transcripts, the RT-PCR products were run out on a 2% agarose gel stained with SYBR GoldSYBR® Gold Nucleic Acid Gel Stain (ThermoFisher), and bands of the expected molecular weight were excised and purified using the QIAquick Gel Extraction Kit (Qiagen), then subjected to cloning with the TOPO® TA Cloning Kit (ThermoFisher). Sequencing of the TOPO clones was performed using M13F and M13R sequencing primers. Sequence analysis was performed using the Sequencher DNA analysis software (Gene Codes Corporation).

Full-length HSD17B13 transcripts were amplified directly from 50 ng of total RNA with the SuperScript III One-step RT-PCR System with Platinum Taq High Fidelity (ThermoFisher Scientific) using gene-specific primers in the first (GCAAAGCCATGAACATC ATCC (SEQ ID NO:51)) and last exons (TCTTGATGTAGTGGGAGTCGGATT (SEQ ID NO:52)) to generate an amplicon of about 2.2 kb (maximum predicted size transcript). Amplicons were verified on an Agilent Bioanalyzer. PacBio-compatible barcoded adapters were ligated to the amplicons and cleaned with PacBio PB beads (Pacific Biosciences). Libraries were pooled in equal amounts and sequenced on one SMRT cell for 180 minutes on the PacBio RSII platform. The data was demultiplexed using PacBio software smrtanalysis v2.3 tool labelzmw and then analyzed with ConsensusTools AmpliconAnalysis. Resulting amplicons were compared to HSD17B13 RefSeq genes to determine isoform and genotype status.

Subcellular Localization of HSD17B13 Isoforms

HepG2 cells were cultured in Eagle's Minimum Essential Medium supplemented with 10% fetal bovine serum. HSD17B13 transcripts A and D were sub-cloned into Myc-DDK backbone lentivirus constructs, and lentivirus were generated. HepG2 cells were infected with lentivirus carrying the HSD17B13 transcripts. Stable cell lines expressing each HSD17B13 transcript were selected with 1-3 mg/ml Geneticin G-418 sulfate in complete culture medium for two weeks. Following fixation, HSD17B13 isoforms were detected with mouse anti-Myc antibody. Lipid droplets were labeled with BODIPY FL dye (Sigma). Secondary antibodies for immunofluorescence were Alexa Fluor 488 donkey anti-rabbit IgG and Alexa Fluor 594 donkey anti-mouse IgG (Jackson ImmunoResearch).

Quantification of HSD171B3 Protein Expression in Human Liver Biopsy Tissue and Stable Cell Lines

Human liver and cell pellet samples were homogenized in ice-cold 1x RIPA lysis buffer (EMD Millipore) in the presence of protease and phosphatase inhibitor mixtures (ThermoFisher). Supernatant was collected and used for protein concentration using BCA protein assay (ThermoFisher). Human tissue and cell lysates were loaded and separated on SDS/PAGE gels

(Bio-Rad) and transferred to PVDF membranes (Bio-Rad). The membranes were blocked for 1 hour with 5% (wt/vol) milk in 1x TBS supplemented with 0.1% Tween20 (Bio-Rad). Membranes were incubated with antibody at 4°C overnight against HSD17B13 (1:200, Thermo-Fisher) and B-Actin (1:500, Cell Signaling Technology). Bound antibody was detected using HRP-conjugated anti-rabbit antibody (1:10,000, Jackson ImmunoResearch) and enhanced using chemi-luminescence reagent (ThermoFisher). Band intensities were quantified using Image J software.

Real-Time Semi-Quantitative PCR

RNA was extracted from cell using TRIzol® (Invitrogen, Carlsbad, CA). First-strand cDNA was synthesized using Superscript III RT (Invitrogen) and utilized for Semi-Quantitative PCR based on intron-spanning primers. A QuantStudio 6 Flex Real-Time PCR System was used to measure the expression level of transcripts. Primers of HSD17B13 and TBP were ordered from IDT (Integrated DNA Technologies). Relative gene expression was analyzed with the ΔΔCt method, providing a fold-change of expression normalized to the house-keeping gene TBP (ΔCt).

Lipid Droplet Isolation and Characterization by Western Blotting

Lipid droplets were prepared from HepG2 cells stably expressing HSD17B13 transcript A (IsoA) or transcript D (IsoD) as previously reported (Brasaemle DL, Wolins NE. Isolation of lipid droplets from cells by density gradient centrifugation, Current protocols in cell biology 2006;Chapter 3:Unit 3 15; and Ding et al., Nature Protocols, 2013, 8, 43-51). In brief, HepG2 cells stably expressing HSD17B13 IsoA, IsoD, or the parental line were incubated overnight with 1 mM oleic acid. The following lipid loading, cells were scraped and resuspended in hypotonic lysis buffer (20 mM Tris, pH 7.5, 1 mM EDTA) supplemented with 1X HaltTM protease/phosphatase inhibitors (Thermo) and lysed by cavitation at 50 bar for 8 minutes. Lysates were centrifuged at 1000g/4°C for 10 minutes, and the post-nuclear supernatant (PNS) was mixed with sucrose to a final volume of 2 mL and concentration of 20% in ultracentrifuge tubes. Then 1.5 mL of 5% sucrose and another 1.5 mL of hypotonic lysis buffer was layered on top of the lysate. Tubes were centrifuged at 182,000g/4°C for 40 minutes, and the lipid droplet (LD) layers were transferred to new tubes. The remaining volume in the tube was aspirated, and the pelleted (total membrane, TM) was resuspended in 0.5 mL hypotonic lysis buffer. The PNS, LD, and TM fractions were mixed with 1x radioimmunoprecipitation (RIPA) buffer (EMD) + NuPAGETM LDS Sample Buffer (Thermo) and β-mercaptoethanol and sonicated for 3 hours at 37°C. The TM lysate was diluted 2.5-fold to normalize to the PNS. Lysates were run on 4-20% SDS-PAGE gels (Biorad), transferred using the Trans-Blot (Biorad) onto low fluorescence

PVDF membranes, and blocked for 1 hour in Odyssey TBS Blocking Buffer. Membranes were incubated overnight with the following antibodies: α-HSD17B13 (Abgent, cat # AP5729a 1:500); LD marker: α-ADRP (Proteintech, 152-94-1-AP, 1:2500); LD marker: α-TIP47 (Proteintech, 10694 1:2000); lysosome marker: α-LAMP1 (Novus, NBP2-25183, 1:1000); cytosolic marker: α-GAPDH (Proteintech, 60004-1-Ig, 1:2000); endoplasmic reticulum marker: α-calreticulin (Abcam, ab92516, 1:1000); mitochondrial marker: α-COX IV (Abcam, ab33985, 1:500); cytoskeleton marker: α-actin (Sigma, A5441, 1:4000). The next day membranes were washed 4 times with Tris-buffered saline + 0.1% Tween, then incubated for 1 hour at room temperature with blocking buffer containing IRDye® α-rabbit (800CW) and α-mouse (680RD) secondary antibodies (Li-Cor) at 1:5,000 and 1:10,000 dilutions, respectively. Gels were washed again with TBST and imaged using the Odyssey.

Quantification of Intracellular Triglyceride Content

The triglyceride (TG) content from the stable cells was determined using a TG quantification kit (Abcam). In the assay, TG are converted to free fatty acids and glycerol. The glycerol is then oxidized to generate a product which is quantified (spectrophotometry at λ= 570 nm).

Substrate Screening of Steroid and Bioactive Lipid Libraries Against Purified Recombinant HSD17B13

Reactions were performed in a final volume of 40 µl of assay buffer (0.2 M Tris-HCl, pH 7.5) which contained 500 µM NAD^+, 5 µM bioactive lipid or 50 µM steroid (all in a final concentration of 5% DMSO), and 100 ng recombinant human HSD17B13. Reactions were incubated for 3 hours, at 23°C, after which an equal volume NADH-Glo Detection Reagent (Promega) was added. Following a 1 hour incubation at 23°C, the relative light units (RLUs) were measured on an Envision Plate Reader (Perkin Elmer). Raw RLU values were normalized as percent of control (50 µM estradiol) following subtraction of negative control (5% DMSO) using the following formula: Percent of control (POC) = 100 x (Sample (RLU) – Negative CTRLaverage) / (Positive CTRLaverage – Negative CTRLaverage).

In vitro and Cellular Characterization of HSD17B13 Enzymatic Activity

Recombinant human HSD17B13 protein was purified from *E. coli* (Genscript) transformed with plasmid DNA harboring HSD17B13 transcript A or transcript D. The HSD17B13 variants contained a 10xHis tag at the C terminus and were purified from soluble fraction using a $Ni2^+$ affinity purification. Enzymatic activity was determined through measurement of NADH production using the NAD(P)H-Glo Detection System (Promega). Reactions were performed for 3 hours at 25°C in 0.2 M Tris-HCl, pH 7.5, 0.5mM NAD^+, 75 µM

of substrate (Sigma) and 500 ng purified enzyme in a final volume of 100 µL. After incubation, 20 µL of the reaction was combined with 20 µL luciferase reagent (Promega), incubated at room temperature for 1 hour and read on an Envision Plate Reader (Perkin Elmer).

HEK293 cells overexpressing HSD17B13 transcript A, transcript D or green fluorescent protein (GFP, control) were used to investigate the activity of HSD17B13 against estradiol in a cell-based assay. Estradiol (1 µM) was fed to each cell type. After 48 hours, the media was collected and the concentration of estradiol and its converted product estrone were identified and quantified by LC-MS.

Association of Exonic Variants with Asparatate and Alanine Aminotransferases

We tested 502,219 biallelic single genetic variants for association with serum ALT or AST levels in 46,544 individuals of European descent from the DiscovEHR study ("GHS discovery cohort"; basic demographics in Table 1). A total of 35 variants in 19 genes were found to be associated with ALT or AST at $P<1.0x10^{-7}$ (Figures 1A and 1B, and Table 2). We performed replication studies in three cohorts of European-ancestry individuals: 1) bariatric surgery patients (n=2,644) from DiscovEHR ("GHS bariatric surgery cohort"); 2) 1,357 individuals from the Dallas Heart Study; and 3) 8,526 individuals from the Penn Medicine Biobank. In meta-analysis of the replication cohorts, thirteen variants in nine genes were significantly associated with serum levels of ALT or AST (Bonferroni significance threshold of $P<1.43x10^{-3}$ for 35 variants tested, Table 3). These included variants that were previously reported to be associated with elevated transaminase levels, such as PNPLA37, TM6SF211, SERPINA122, SAMM5023, and ERLIN124. SERPINA1 encodes alpha-1-antitrypsin, whose functional deficiency causes liver disease; the association with SAMM50 is mediated via linkage disequilibrium with variation in PNPLA3, and ERLIN1 has been implicated in liver fat deposition. We also identified variants that were not previously reported to be associated with liver disease. These included several variants in GPT and GOT1, the genes encoding ALT and AST, respectively, and SLC39A12, which encodes solute carrier family 39 member 12.

We also identified a reproducible association between a variant in HSD17B13, the gene encoding hydroxysteroid 17-beta dehydrogenase 13, an uncharacterized member of the 17-beta hydroxysteroid dehydrogenase family, and decreased levels of ALT (discovery $P=4.2x10^{-12}$, replication $P=1.7x10^{-4}$) and AST (discovery $P=6.2x10^{-10}$, replication $P=1.7x10^{-4}$, Table 3). The associated variant, rs72613567, is an insertion of an adenine adjacent to the donor splice site of exon six (TA allele), and had an allele frequency of 26.0% in the GHS discovery cohort. Previously, Chambers et al. identified a nearby locus at 4q22 (rs6834314) associated with ALT levels (Chambers et al., Nat. Genet., 2011, 43, 1131-1138, doi:10.1038/ng.970); rs72613567 has

not heretofore been reported to be associated with transaminase levels. HSD17B13 is 30 kb upstream of HSD17B11, another member of the same gene family. We did not observe exome-wide significant associations between coding or splice variants in HSD17B11 and transaminase levels in the discovery cohort (Figures 5A and 5B) or in the joint meta-analysis of the discovery cohort and three replication cohorts. Furthermore, linkage disequilibrium of rs72613567 with variants in HSD17B11 was modest across all ancestry groups (r^2<0.4 with all ascertained variants in HSD17B11 in all ancestry groups). Collectively, these findings suggest HSD17B13 as the gene in the genomic region that is most likely to be functionally related to transaminase levels.

Table 1. Demographics and clinical characteristics of sequenced European-ancestry individuals from the discovery and replication cohorts.

Characteristic	**Discovery Cohort (N = 46,544)**	**Bariatric Surgery Cohort (N = 2,644)**	**Dallas Heart Study (N = 1,357)**	**Penn Medicine Biobank (N = 8,526)**
Age (years) – median (IQR)	62.9 (49.6 - 73.8)	52.9 (44.1 - 61.2)	46.0 (38.0 - 54.0)	68.0 (60.0 - 76.0)
Female sex – number (%)	26,875 (57.7)	2,119 (80.1)	724 (53.4)	3,242 (38.0)
Body mass index – median (IQR)	29.9 (35.4 - 44.8)	47.4 (42.0 - 53.7)	28 (25-32)	30 (25-32)
Transaminase level (U/L) – median (IQR)				
Alanine aminotransferase (ALT)	22.0 (17.0 - 29.0)	23.0 (17.5 - 29.5)	20.0 (15.0 - 27.0)	22.0 (17.0 - 30.0)
Aspartate aminotransferase (AST)	23.0 (20.0 - 27.5)	23.0 (20.0 - 27.0)	21.0 (18.0 - 25.0)	24.0 (20.0 - 30.5)
Presence of liver disease (by ICD-9 code) – N (%)				
Alcoholic liver disease	197 (0.4)	7 (0.3)	-	-
Alcoholic cirrhosis	130 (0.3)	3 (0.1)	-	-
Nonalcoholic, non-viral liver disease	1,938 (4.2)	1,543 (58.4)	-	-
Nonalcoholic cirrhosis	382 (0.8)	24 (0.9)	-	-
Hepatocellular carcinoma	76 (0.2)	1 (0.04)	-	-
No liver disease	30,628 (65.8)	1 (0.04)	-	-

Table 2. Single nucleotide variants associated with serum transaminase levels at $P < 1.0x10^{-7}$ in the discovery cohort.

Trait	**CHR**	**BP**	**REF**	**ALT**	**rsID**	**Gene**	**Annotation**	**AA Substitution**	**Beta (SE)**
ALT	1	220970028	A	G	rs2642438	*MARC1*	missense	p.Thr165Ala	0.008 (0.001)

Trait	CHR	BP	REF	ALT	rsID	Gene	Annotation	AA Substitution	Beta (SE)
	4	88231392	T	TA	***rs72613567**	*HSD17B13*	splice donor		-0.009 (0.001)
	8	144997604	C	T	rs371119003	*PLEC*	missense	p.Ala2302Thr	-0.160 (0.026)
	8	145008502	G	A		*PLEC*	missense	p.Arg522Cys	-0.268 (0.032)
	8	145692918	G	A	rs35968570	*KIFC2*	missense	p.Glu174Lys	-0.033 (0.005)
	8	145730072	G	A	rs143408057	*GPT*	missense	p.Arg83His	-0.314 (0.036)
	8	145730161	C	T	rs201815297	*GPT*	missense	p.Ala87Val	-0.224 (0.014)
	8	145730221	G	A	rs112574791	*GPT*	missense	p.Arg107Lys	-0.033 (0.005)
	8	145731636	T	G	rs145155876	*GPT*	stop gained	p.Tyr326*	-0.235 (0.031)
	8	145732114	G	C	rs141505249	*GPT*	missense	p.Glu430Gln	-0.224 (0.013)
	8	145732151	G	A	rs143462595	*GPT*	missense	p.Arg442His	-0.077 (0.013)
	8	145732180	G	C	rs147998249	*GPT*	missense	p.Val452Leu	-0.225 (0.013)
	8	145732305	G	GC		*GPT*	frameshift	p.Glu475fs	-0.271 (0.031)
	8	145748532	A	G	rs567402720	*LRRC24*	missense	p.Leu290Ser	-0.185 (0.028)
	9	117122202	C	T	rs3748177	*AKNA*	synonymous	p.Glu755Glu	-0.007 (0.001)
	9	117124731	G	A	rs3748176	*AKNA*	missense	p.Pro624Leu	-0.007 (0.001)
	10	101595996	T	A	rs17222723	*ABCC2*	missense	p.Val1188Glu	-0.015 (0.003)
	10	101606861	G	T	rs1137968	*ABCC2*	synonymous	p.Val1430Val	-0.015 (0.003)
	10	101610533	C	T	rs8187707	*ABCC2*	synonymous	p.His1496His	-0.015 (0.003)

Trait	CHR	BP	REF	ALT	rsID	Gene	Annotation	AA Substitution	Beta (SE)
	10	101611294	G	A	rs8187710	*ABCC2*	missense	p.Cys1515Tyr	-0.015 (0.003)
	10	101912064	T	C	***rs2862954**	*ERLIN1*	missense	p.Ile291Val	-0.012 (0.001)
	10	101977883	C	T	rs2230804	*CHUK*	missense	p.Val268Ile	-0.009 (0.001)
	10	113917085	T	A	rs2254537	*GPAM*	synonymous	p.Pro681Pro	-0.008 (0.001)
	10	113940329	T	C	rs2792751	*GPAM*	missense	p.Ile43Val	-0.008 (0.001)
	14	94844947	C	T	***rs28929474**	*SERPINA1*	missense	p.Glu366Lys	0.042 (0.005)
	19	19379549	C	T	***rs58542926**	*TM6SF2*	missense	p.Glu167Lys	0.014 (0.002)
	22	44324727	C	G	***rs738409**	*PNPLA3*	missense	p.Ile148Met	0.023 (0.002)
	22	44324730	C	T	***rs738408**	*PNPLA3*	synonymous	p.Pro149Pro	0.023 (0.002)
	22	44342116	A	G	rs2294918	*PNPLA3*	missense	p.Lys434Glu	0.007 (0.001)
	22	44368122	A	G	***rs3761472**	*SAMM50*	missense	p.Asp110Gly	0.019 (0.002)
	22	44395451	T	C	***rs1007863**	*PARVB*	missense	p.Trp37Arg	0.011 (0.001)
AST	4	88231392	T	TA	***rs72613567**	*HSD17B13*	splice donor		-0.005 (0.001)
	10	18242311	A	G	rs10764176	*SLC39A12*	missense	p.Ser36Gly	-0.006 (0.001)
	10	101157378	CGTT	C		*GOT1*	inframe indel	p.Asn389del	-0.221 (0.024)
	10	101165533	G	C	rs374966349	*GOT1*	missense	p.Gln208Glu	0.271 (0.027)
	10	101912064	T	C	***rs2862954**	*ERLIN1*	missense	p.Ile291Val	-0.005 (0.001)
	11	22271870	A	T	rs7481951	*ANO5*	missense	p.Leu322Phe	0.004 (0.001)
	14	94844947	C	T	***rs28929474**	*SERPINA1*	missense	p.Glu366Lys	0.027 (0.003)
	19	19379549	C	T	***rs58542926**	*TM6SF2*	missense	p.Glu167Lys	0.008 (0.002)
	22	44324727	C	G	***rs738409**	*PNPLA3*	missense	p.Ile148Met	0.014 (0.001)
	22	44324730	C	T	***rs738408**	*PNPLA3*	synonymous	p.Pro149Pro	0.014 (0.001)

Trait	CHR	BP	REF	ALT	rsID	Gene	Annotation	AA Substitution	Beta (SE)
	22	44368122	A	G	***rs3761472**	*SAMM50*	missense	p.Asp110Gly	0.011 (0.001)
	22	44395451	T	C	***rs1007863**	*PARVB*	missense	p.Trp37Arg	0.006 (0.001)

Table 2 (cont.)

								N			Mean AST or ALT level (U/L)		
Trait	**CHR**	**BP**	**REF**	**ALT**	**P**	**AAF**	**N**	**REF/ REF**	**REF/ ALT**	**ALT/ ALT**	**REF/ REF**	**REF/ ALT**	**ALT/ ALT**
ALT	1	220970028	A	G	4.67E-08	0.7067	41,414	3,515	17,262	20,637	23.88	24.52	24.92
	4	88231392	T	TA	4.16E-12	0.2634	41,414	22,441	16,130	2,843	25.02	24.26	24.1
	8	144997604	C	T	1.30E-09	0.0005	41,413	41,373	40	0	24.67	18.1	NA
	8	145008502	G	A	3.26E-17	0.0003	41,414	41,387	27	0	24.67	13.8	NA
	8	145692918	G	A	1.40E-11	0.0139	41,414	40,271	1,133	10	24.67	12.07	NA
	8	145730072	G	A	3.28E-18	0.0003	41,414	41,393	21	0	24.67	12.07	NA
	8	145730161	C	T	6.28E-59	0.0018	41,414	41,270	144	0	24.7	14.68	NA
	8	145730221	G	A	4.25E-11	0.0136	41,414	40,293	1,111	10	24.71	23.09	18.35
	8	145731636	T	G	1.76E-14	0.0004	41,394	41,364	30	0	24.67	14.07	NA
	8	145732114	G	C	8.84E-64	0.0019	41,375	41,223	150	2	24.7	14.48	13.75
	8	145732151	G	A	1.18E-09	0.0021	41,406	41,232	174	0	24.68	20.87	NA
	8	145732180	G	C	8.19E-65	0.0019	41,413	41,254	159	0	24.7	14.74	NA
	8	145732305	G	G	1.00E-18	0.0004	41,414	41,385	29	0	24.67	14.24	NA
	8	145748532	A	G	3.42E-11	0.0004	41,393	41,358	35	0	24.67	17.71	NA
	9	117122202	C	T	9.51E-09	0.5232	41,414	9,414	20,645	11,355	25.12	24.72	24.18
	9	117124731	G	A	4.31E-09	0.5230	41,412	9,427	20,634	11,351	25.12	24.73	24.17
	10	101595996	T	A	2.97E-08	0.0608	41,414	36,543	4,704	167	24.77	23.97	22.12
	10	101606861	G	T	2.71E-08	0.0608	41,414	36,543	4,704	167	24.77	23.97	22.04
	10	101610533	C	T	2.77E-08	0.0608	41,414	36,542	4,706	166	24.77	23.97	22.03

								N			Mean AST or ALT level (U/L)		
Trait	CHR	BP	REF	ALT	P	AAF	N	REF/ REF	REF/ ALT	ALT/ ALT	REF/ REF	REF/ ALT	ALT/ ALT
	10	101611294	G	A	2.15E-08	0.0611	41,414	36,519	4,726	169	24.77	23.97	21.99
	10	101912064	T	C	2.43E-21	0.4755	41,414	11,318	20,819	9,277	25.32	24.71	23.77
	10	101977883	C	T	1.93E-13	0.5072	41,414	10,048	20,733	10,633	25.18	24.75	24.01
	10	113917085	T	A	4.61E-10	0.7073	41,414	3,627	16,984	20,803	25	24.97	24.36
	10	113940329	T	C	2.54E-10	0.7097	41,412	3,567	16,910	20,935	25	24.98	24.35
	14	94844947	C	T	9.28E-21	0.0171	41,414	40,006	1,399	9	24.58	26.91	43.89
	19	19379549	C	T	4.76E-09	0.0759	41,413	35,388	5,780	245	24.52	25.46	26.84
	22	44324727	C	G	1.34E-50	0.2351	41,414	24,257	14,837	2,320	24.06	24.99	28.91
	22	44324730	C	T	1.11E-50	0.2349	41,414	24,273	14,824	2,317	24.06	24.98	28.92
	22	44342116	A	G	8.26E-08	0.5986	41,412	6,691	19,833	14,888	24.15	24.47	25.15
	22	44368122	A	G	8.85E-30	0.1682	41,413	28,626	11,618	1,169	24.23	25.36	28.45
	22	44395451	T	C	7.98E-16	0.3963	41,414	15,036	19,920	6,458	24.15	24.6	26.09
AST	4	88231392	T	TA	6.24E-10	0.2638	40,753	22,068	15,870	2,815	24.47	24.1	23.96
	10	18242311	A	G	1.09E-10	0.2881	40,753	20,645	16,738	3,370	24.47	24.15	23.85
	10	101157378	GT	C	1.96E-20	0.0002	40,753	40,733	20	0	24.29	14.7	NA
	10	101165533	G	C	2.43E-24	0.0002	40,753	40,736	17	0	24.28	44.5	NA
	10	101912064	T	C	4.82E-09	0.4754	40,753	11,138	20,486	9,129	24.59	24.26	23.99
	11	22271870	A	T	9.61E-08	0.5833	40,722	7,123	19,686	13,913	24.03	24.22	24.53
	14	94844947	C	T	2.44E-20	0.0172	40,753	39,361	1,384	8	24.24	25.76	34.5
	19	19379549	C	T	6.54E-08	0.0760	40,752	34,811	5,698	243	24.21	24.74	25.43
	22	44324727	C	G	8.31E-46	0.2343	40,753	23,889	14,622	2,242	23.96	24.48	26.62
	22	44324730	C	T	8.93E-46	0.2341	40,753	23,905	14,609	2,239	23.96	24.47	26.63
	22	44368122	A	G	1.22E-22	0.1680	40,752	28,170	11,450	1,132	24.07	24.64	26.24
	22	44395451	T	C	1.31E-13	0.3961	40,753	14,761	19,678	6,314	24.02	24.23	25.1

* Indicates variants having exome-wide significant associations with both ALT and AST. Abbreviations: AAF, alternate allele frequency; Alt, alternate allele; ALT, alanine aminotransferase; AST, aspartate aminotransferase; Ref, reference allele; SE, standard error.

Table 3. Replication and joint meta-analysis of 35 exome-wide significant single nucleotide variants from the discovery cohort in three separate European-ancestry cohorts.

									GHS Discovery Cohort		
Trait	**Chr**	**BP**	**Ref**	**Alt**	**RSID**	**Gene**	**Ann**	**AA Substitution**	**Beta (SE)**	**P**	**N**
ALT	1	220970028	A	G	rs2642438	*MARC1*	mis	p.Thr165Ala	0.008 (0.001)	4.67E-08	41,414
	4	88231392	T	TA	rs72613567	*HSD17B13*	spl		-0.009 (0.001)	4.16E-12	41,414
	8	144997604	C	T	rs371119003	*PLEC*	mis	p.Ala2302Thr	-0.160 (0.026)	1.30E-09	41,413
	8	145008502	G	A		*PLEC*	mis	p.Arg522Cys	-0.268 (0.032)	3.26E-17	41,414
	8	145692918	G	A	rs35968570	*KIFC2*	mis	p.Glu174Lys	-0.033 (0.005)	1.40E-11	41,414
	8	145730072	G	A	rs143408057	*GPT*	mis	p.Arg83His	-0.314 (0.036)	3.28E-18	41,414
	8	145730161	C	T	rs201815297	*GPT*	mis	p.Ala87Val	-0.224 (0.014)	6.28E-59	41,414
	8	145730221	G	A	rs112574791	*GPT*	mis	p.Arg107Lys	-0.033 (0.005)	4.25E-11	41,414
	8	145731636	T	G	rs145155876	*GPT*	stop	p.Tyr326*	-0.235 (0.031)	1.76E-14	41,394
	8	145732114	G	C	rs141505249	*GPT*	mis	p.Glu430Gln	-0.224 (0.013)	8.84E-64	41,375
	8	145732151	G	A	rs143462595	*GPT*	mis	p.Arg442His	-0.077 (0.013)	1.18E-09	41,406
	8	145732180	G	C	rs147998249	*GPT*	mis	p.Val452Leu	-0.225 (0.013)	8.19E-65	41,413
	8	145732305	G	GC		*GPT*	fs	p.Glu475fs	-0.271 (0.031)	1.00E-18	41,414

									GHS Discovery Cohort		
Trait	**Chr**	**BP**	**Ref**	**Alt**	**RSID**	**Gene**	**Ann**	**AA Substitution**	**Beta (SE)**	**P**	**N**
	8	145748532	A	G	rs567402720	*LRRC24*	mis	p.Leu290Ser	-0.185 (0.028)	3.42E-11	41,393
	9	117122202	C	T	rs3748177	*AKNA*	syn	p.Glu755Glu	-0.007 (0.001)	9.51E-09	41,414
	9	117124731	G	A	rs3748176	*AKNA*	mis	p.Pro624Leu	-0.007 (0.001)	4.31E-09	41,412
	10	101595996	T	A	rs17222723	*ABCC2*	mis	p.Val1188Glu	-0.015 (0.003)	2.97E-08	41,414
	10	101606861	G	T	rs1137968	*ABCC2*	syn	p.Val1430Val	-0.015 (0.003)	2.71E-08	41,414
	10	101610533	C	T	rs8187707	*ABCC2*	syn	p.His1496His	-0.015 (0.003)	2.77E-08	41,414
	10	101611294	G	A	rs8187710	*ABCC2*	mis	p.Cys1515Tyr	-0.015 (0.003)	2.15E-08	41,414
	10	101912064	T	C	rs2862954	*ERLIN1*	mis	p.Ile291Val	-0.012 (0.001)	2.43E-21	40,834
	10	101977883	C	T	rs2230804	*CHUK*	mis	p.Val268Ile	-0.009 (0.001)	1.93E-13	41,414
	10	113917085	T	A	rs2254537	*GPAM*	syn	p.Pro681Pro	-0.008 (0.001)	4.61E-10	41,414
	10	113940329	T	C	rs2792751	*GPAM*	mis	p.Ile43Val	-0.008 (0.001)	2.54E-10	41,412
	14	94844947	C	T	rs28929474	*SERPINA1*	mis	p.Glu366Lys	0.042 (0.005)	9.28E-21	41,414
	19	19379549	C	T	rs58542926	*TM6SF2*	mis	p.Glu167Lys	0.014 (0.002)	4.76E-09	41,413
	22	44324727	C	G	rs738409	*PNPLA3*	mis	p.Ile148Met	0.023 (0.002)	1.34E-50	41,414
	22	44324730	C	T	rs738408	*PNPLA3*	syn	p.Pro149Pro	0.023 (0.002)	1.11E-50	41,414
	22	44342116	A	G	rs2294918	*PNPLA3*	mis	p.Lys434Glu	0.007 (0.001)	8.26E-08	41,412
	22	44368122	A	G	rs3761472	*SAMM50*	mis	p.Asp110Gly	0.019 (0.002)	8.85E-30	41,413

									GHS Discovery Cohort		
Trait	**Chr**	**BP**	**Ref**	**Alt**	**RSID**	**Gene**	**Ann**	**AA Substitution**	**Beta (SE)**	**P**	**N**
	22	44395451	T	C	rs1007863	*PARVB*	mis	p.Trp37Arg	0.011 (0.001)	7.98E-16	41,414
AST	4	88231392	T	TA	rs72613567	*HSD17B13*	spl		-0.005 (0.001)	6.24E-10	40,753
	10	18242311	A	G	rs10764176	*SLC39A12*	mis	p.Ser36Gly	-0.006 (0.001)	1.09E-10	40,753
	10	101157378	CGTT	C		*GOT1*	inf	p.Asn389del	-0.221 (0.024)	1.96E-20	40,753
	10	101165533	G	C	rs374966349	*GOT1*	mis	p.Gln208Glu	0.271 (0.027)	2.43E-24	40,753
	10	101912064	T	C	rs2862954	*ERLIN1*	mis	p.Ile291Val	-0.005 (0.001)	4.82E-09	40,753
	11	22271870	A	T	rs7481951	*ANO5*	mis	p.Leu322Phe	0.004 (0.001)	9.61E-08	40,722
	14	94844947	C	T	rs28929474	*SERPINA1*	mis	p.Glu366Lys	0.027 (0.003)	2.44E-20	40,753
	19	19379549	C	T	rs58542926	*TM6SF2*	mis	p.Glu167Lys	0.008 (0.002)	6.54E-08	40,192
	22	44324727	C	G	rs738409	*PNPLA3*	mis	p.Ile148Met	0.014 (0.001)	8.31E-46	40,753
	22	44324730	C	T	rs738408	*PNPLA3*	syn	p.Pro149Pro	0.014 (0.001)	8.93E-46	40,753
	22	44368122	A	G	rs3761472	*SAMM50*	mis	p.Asp110Gly	0.011 (0.001)	1.22E-22	40,752
	22	44395451	T	C	rs1007863	*PARVB*	Mis	p.Trp37Arg	0.006 (0.001)	1.31E-13	40,753

Table 3 (cont.)

			GHS Bariatric Surgery Cohort			Dallas Heart Study			U. Penn		
Trait	**Chr**	**BP**	**Beta (SE)**	**P**	**N**	**Beta (SE)**	**P**	**N**	**Beta (SE)**	**P**	**N**
ALT	1	220970028	0.005 (0.005)	3.10E-01	2475	0.011 (0.008)	1.76E-01	1357	0.007 (0.004)	1.02E-01	6158

			GHS Bariatric Surgery Cohort			Dallas Heart Study			U. Penn		
Trait	Chr	BP	Beta (SE)	P	N	Beta (SE)	P	N	Beta (SE)	P	N
	4	88231392	-0.010 (0.005)	5.57E-02	2475	-0.016 (0.008)	6.60E-02	1357	-0.013 (0.004)	1.33E-03	6158
	8	144997604	-0.492 (0.165)	2.84E-03	2475	NA (NA)	NA	NA	-0.051 (0.072)	4.79E-01	6158
	8	145008502	-0.161 (0.165)	3.29E-01	2475	NA (NA)	NA	NA	-0.247 (0.143)	8.48E-02	6158
	8	145692918	-0.009 (0.020)	6.48E-01	2475	0.032 (0.036)	3.76E-01	1356	-0.053 (0.018)	3.72E-03	6158
	8	145730072	-0.189 (0.165)	2.50E-01	2475	NA (NA)	NA	NA	-0.298(0.101)	3.26E-03	6158
	8	145730161	-0.341 (0.074)	3.64E-06	2475	NA (NA)	NA	NA	-0.143 (0.054)	8.50E-03	6158
	8	145730221	-0.009 (0.020)	6.45E-01	2475	0.028 (0.036)	4.37E-01	1357	-0.060 (0.018)	5.60E-04	6158
	8	145731636	-0.314 (0.165)	5.71E-02	2475	-0.317 (0.140)	2.35E-02	1356	-0.148 (0.143)	3.04E-01	6157
	8	145732114	-0.273 (0.048)	9.83E-09	2474	-0.240 (0.075)	1.36E-03	1357	-0.197 (0.041)	1.31E-06	6157
	8	145732151	-0.115 (0.058)	4.82E-02	2475	-0.106 (0.099)	2.86E-01	1356	-0.049 (0.041)	2.27E-01	6157
	8	145732180	-0.273 (0.050)	4.26E-08	2475	-0.191 (0.070)	6.58E-03	1357	-0.197 (0.041)	1.31E-06	6158
	8	145732305	-0.161 (0.165)	3.29E-01	2475	NA (NA)	NA	NA	-0.509 (0.203)	1.21E-02	6158
	8	145748532	-0.161 (0.165)	3.29E-01	2475	NA (NA)	NA	NA	-0.307 (0.143)	3.21E-02	6158
	9	117122202	-0.004 (0.005)	4.09E-01	2475	0.004 (0.008)	6.18E-01	1357	-0.007 (0.004)	5.29E-02	6158
	9	117124731	-0.004 (0.005)	3.90E-01	2475	0.003 (0.008)	7.33E-01	1356	-0.007 (0.004)	4.24E-02	6158
	10	101595996	-0.002 (0.010)	8.01E-01	2475	-0.007 (0.017)	6.88E-01	1357	-0.017 (0.007)	1.55E-02	6158
	10	101606861	-0.003 (0.010)	7.74E-01	2475	-0.008 (0.017)	6.28E-01	1357	-0.017 (0.007)	1.70E-02	6158

			GHS Bariatric Surgery Cohort			Dallas Heart Study			U. Penn		
Trait	**Chr**	**BP**	**Beta (SE)**	**P**	**N**	**Beta (SE)**	**P**	**N**	**Beta (SE)**	**P**	**N**
	10	101610533	-0.003 (0.010)	7.93E-01	2475	-0.008 (0.017)	6.28E-01	1357	-0.017 (0.007)	1.76E-02	6158
	10	101611294	-0.001 (0.010)	9.11E-01	2475	-0.010 (0.017)	5.40E-01	1357	-0.016 (0.007)	2.77E-02	6158
	10	101912064	-0.010 (0.005)	2.91E-02	2475	-0.006 (0.007)	4.02E-01	1356	-0.009 (0.004)	2.06E-02	6158
	10	101977883	-0.006 (0.005)	2.05E-01	2475	0.0001 (0.008)	9.94E-01	1357	-0.011 (0.004)	3.91E-03	6158
	10	113917085	-0.003 (0.005)	5.80E-01	2475	-0.013 (0.008)	1.15E-01	1357	-0.008 (0.004)	5.12E-02	6158
	10	113940329	-0.003 (0.005)	5.61E-01	2475	-0.013 (0.008)	1.33E-01	1357	-0.008 (0.004)	4.77E-02	6158
	14	94844947	0.035 (0.020)	7.97E-02	2475	0.034 (0.032)	2.92E-01	1357	0.054 (0.013)	1.63E-05	6158
	19	19379549	0.040 (0.010)	2.40E-05	2475	0.024 (0.014)	9.50E-02	1357	0.013 (0.008)	7.51E-02	6158
	22	44324727	0.019 (0.006)	5.54E-04	2475	0.006 (0.009)	5.43E-01	1357	0.016 (0.004)	2.05E-04	6158
	22	44324730	0.019 (0.006)	5.51E-04	2475	0.006 (0.009)	5.43E-01	1357	0.016 (0.004)	2.14E-04	6158
	22	44342116	0.001 (0.005)	7.77E-01	2475	0.005 (0.008)	5.18E-01	1357	0.005 (0.004)	2.16E-01	6158
	22	44368122	0.009 (0.006)	1.66E-01	2475	-0.001 (0.01)	9.37E-01	1357	0.018 (0.005)	4.02E-04	6158
	22	44395451	0.003 (0.005)	5.22E-01	2475	0.008 (0.008)	3.13E-01	1357	0.009 (0.004)	2.50E-02	6158
AST	4	88231392	-0.010 (0.003)	3.12E-03	2469	-0.012 (0.006)	5.32E-02	1357	-0.007 (0.004)	5.56E-02	6166
	10	18242311	-0.010 (0.003)	2.91E-03	2469	-0.003 (0.006)	5.80E-01	1357	-0.009 (0.004)	1.03E-02	6166
	10	101157378	-0.205 (0.062)	8.57E-04	2469	NA (NA)	NA	NA	-0.243 (0.088)	5.97E-03	6165
	10	101165533	NA (NA)	NA	NA	NA (NA)	NA	NA	0.339 (0.079)	1.85E-05	6166

			GHS Bariatric Surgery Cohort			Dallas Heart Study			U. Penn		
Trait	**Chr**	**BP**	**Beta (SE)**	**P**	**N**	**Beta (SE)**	**P**	**N**	**Beta (SE)**	**P**	**N**
	10	101912064	-0.004 (0.003)	1.54E-01	2469	-0.007 (0.006)	2.21E-01	1357	-0.004 (0.003)	1.94E-01	6166
	11	22271870	-0.001 (0.003)	7.85E-01	2466	0.006 (0.006)	2.85E-01	1357	-0.002 (0.003)	5.46E-01	6165
	14	94844947	0.023 (0.013)	7.79E-02	2469	0.044 (0.024)	6.98E-02	1357	0.055 (0.011)	4.01E-07	6166
	19	19379549	0.023 (0.006)	1.99E-04	2469	0.010 (0.011)	3.42E-01	1356	0.004 (0.007)	5.94E-01	6166
	22	44324727	0.014 (0.004)	1.27E-04	2469	0.004 (0.007)	5.44E-01	1357	0.015 (0.004)	4.87E-05	6166
	22	44324730	0.014 (0.004)	1.32E-04	2469	0.004 (0.007)	5.44E-01	1357	0.015 (0.004)	4.96E-05	6166
	22	44368122	0.008 (0.004)	6.03E-02	2469	-0.001 (0.008)	9.45E-01	1357	0.016 (0.004)	2.64E-04	6166
	22	44395451	0.003 (0.003)	4.12E-01	2469	0.006 (0.006)	2.95E-01	1357	0.009 (0.003)	6.17E-03	6166

Table 3 (cont.)

			**Replication Meta-Analysis (N=3)		***Joint Meta-Analysis (N = 4)	
Trait	**Chr**	**BP**	**Beta (SE)**	**P**	**Beta (SE)**	**P**
ALT	1	220970028	0.007 (0.003)	2.31E-02	0.008 (0.001)	3.38E-09
	4	88231392	-0.013 (0.003)	***3.85E-05**	-0.010 (0.001)	1.17E-15
	8	144997604	-0.121 (0.066)	6.56E-02	-0.155 (0.025)	2.68E-10
	8	145008502	-0.210 (0.108)	5.23E-02	-0.264 (0.031)	5.54E-18
	8	145692918	-0.025 (0.013)	4.69E-02	-0.032 (0.005)	2.25E-12
	8	145730072	-0.268 (0.086)	1.88E-03	-0.308 (0.033)	2.79E-20

			**Replication Meta-Analysis (N=3)		***Joint Meta-Analysis (N = 4)	
Trait	**Chr**	**BP**	**Beta (SE)**	**P**	**Beta (SE)**	**P**
	8	145730161	-0.213 (0.044)	***1.14E-06**	-0.223 (0.013)	4.49E-64
	8	145730221	-0.031 (0.013)	1.36E-02	-0.033 (0.005)	1.92E-12
	8	145731636	-0.256 (0.086)	2.79E-03	-0.237 (0.029)	1.94E-16
	8	145732114	-0.231 (0.029)	***7.24E-16**	-0.225 (0.012)	6.06E-78
	8	145732151	-0.074 (0.032)	1.88E-02	-0.076 (0.012)	7.03E-11
	8	145732180	-0.221 (0.029)	***1.41E-14**	-0.224 (0.012)	1.04E-77
	8	145732305	-0.299 (0.128)	1.93E-02	-0.273 (0.030)	6.44E-20
	8	145748532	-0.244 (0.108)	2.40E-02	-0.189 (0.027)	2.93E-12
	9	117122202	-0.005 (0.003)	8.42E-02	-0.007 (0.001)	3.08E-09
	9	117124731	-0.005 (0.003)	6.15E-02	-0.007 (0.001)	1.00E-09
	10	101595996	-0.012 (0.005)	3.43E-02	-0.014 (0.002)	3.44E-09
	10	101606861	-0.012 (0.005)	3.25E-02	-0.014 (0.002)	2.99E-09
	10	101610533	-0.012 (0.005)	3.43E-02	-0.014 (0.002)	3.23E-09
	10	101611294	-0.011 (0.005)	5.21E-02	-0.014 (0.002)	4.09E-09
	10	101912064	-0.009 (0.003)	***1.14E-03**	-0.011 (0.001)	1.76E-23
	10	101977883	-0.008 (0.003)	4.33E-03	-0.009 (0.001)	3.59E-15
	10	113917085	-0.007 (0.003)	2.07E-02	-0.008 (0.001)	3.28E-11

			**Replication Meta-Analysis (N=3)		***Joint Meta-Analysis (N = 4)	
Trait	**Chr**	**BP**	**Beta (SE)**	**P**	**Beta (SE)**	**P**
	10	113940329	-0.007 (0.003)	2.00E-02	-0.008 (0.001)	1.77E-11
	14	94844947	0.047 (0.010)	***2.82E-06**	0.043 (0.004)	1.59E-25
	19	19379549	0.024 (0.006)	***1.37E-05**	0.016 (0.002)	1.15E-12
	22	44324727	0.016 (0.003)	***7.45E-07**	0.021 (0.001)	3.55E-55
	22	44324730	0.016 (0.003)	***7.73E-07**	0.021 (0.001)	3.10E-55
	22	44342116	0.004 (0.003)	1.91E-01	0.006 (0.001)	6.24E-08
	22	44368122	0.012 (0.004)	***7.69E-04**	0.018 (0.002)	1.08E-31
	22	44395451	0.007 (0.003)	1.78E-02	0.010 (0.001)	1.16E-16
AST	4	88231392	-0.009 (0.002)	***8.38E-05**	-0.006 (0.001)	6.82E-13
	10	18242311	-0.009 (0.002)	***1.16E-04**	-0.006 (0.001)	1.10E-13
	10	101157378	-0.218 (0.051)	***1.66E-05**	-0.220 (0.022)	1.68E-24
	10	101165533	0.339 (0.079)	***1.85E-05**	0.278 (0.025)	3.25E-28
	10	101912064	-0.005 (0.002)	2.51E-02	-0.005 (0.001)	3.68E-10
	11	22271870	0.000 (0.002)	8.43E-01	0.004 (0.001)	1.13E-06
	14	94844947	0.042 (0.008)	***9.54E-08**	0.029 (0.003)	6.71E-26
	19	19379549	0.014 (0.004)	***1.20E-03**	0.009 (0.002)	5.92E-10
	22	44324727	0.013 (0.002)	***5.51E-08**	0.014 (0.001)	3.14E-52

			**Replication Meta-Analysis (N=3)		***Joint Meta-Analysis (N = 4)	
Trait	**Chr**	**BP**	**Beta (SE)**	**P**	**Beta (SE)**	**P**
	22	44324730	0.013 (0.002)	***5.81E-08**	0.014 (0.001)	3.55E-52
	22	44368122	0.010 (0.003)	***3.40E-04**	0.011 (0.001)	1.91E-25
	22	44395451	0.006 (0.002)	7.34E-03	0.006 (0.001)	3.62E-15

* Indicates P-values meeting the Bonferroni significance threshold of $P < 1.43 \times 10^{-3}$.

** Replication meta-analysis includes the three replication cohorts: GHS Bariatric Surgery Cohort, Dallas Heart Study, and Penn Medicine Biobank.

*** Joint meta-analysis includes the discovery cohort and the three replication cohorts: GHS Discovery Cohort, GHS Bariatric Surgery Cohort, Dallas Heart Study, and Penn Medicine Biobank.

Abbreviations: AAF, alternate allele frequency; Alt, alternate allele; ALT, alanine aminotransferase; AST, aspartate aminotransferase; Ref, reference allele; SE, standard error; ann, annotation; mis, missense; syn, synonymous; spl, splice donor; stop, stop gained; fs, frameshift; inf, inframe indel.

Association of Exonic Variants with Clinical Diagnoses of Chronic Liver Disease

Next, we analyzed the relationship between the thirteen transaminase-associated variants in the nine genes found in the discovery and replication cohorts and chronic liver disease, including alcoholic and nonalcoholic (non-viral) liver disease, as well as the most advanced forms of chronic liver disease: alcoholic cirrhosis, nonalcoholic cirrhosis, and hepatocellular carcinoma (HCC). Using a Bonferroni significance threshold of $P<1.92\times10^{-3}$ for the thirteen variants tested, we found significant associations between six variants in five genes (HSD17B13, SERPINA1, TM6SF2, PNPLA3, and SAMM50) and chronic liver disease phenotypes (Table 4). The SERPINA1, TM6SF2, PNPLA3, and SAMM50 associations confirm previously reported associations. In the discovery cohort, HSD17B13 rs72613567:TA was associated with lower odds of all EHR-derived categories of both alcoholic and nonalcoholic liver disease in an allele dosage-dependent manner (Figure 2A): all categories of alcoholic liver disease, heterozygous odds ratio (OR_{het}) (95% confidence interval) 0.58 (0.42-0.80), homozygous OR (OR_{hom}) 0.47 (0.23-0.97), allelic OR ($OR_{allelic}$) 0.62 (0.48-0.81), $P=1.8\times10^{-4}$; all

categories of nonalcoholic liver disease, OR_{het} 0.83 (0.75-0.92), OR_{hom} 0.70 (0.57-0.87), $OR_{allelic}$ 0.84 (0.78-0.91), P=1.3x10^{-5}. HSD17B13 rs72613567:TA was also associated with lower odds of alcoholic and nonalcoholic cirrhosis, with 42% and 73% lower odds of alcoholic cirrhosis for heterozygotes and homozygotes, respectively, (OR_{het} 0.58 (0.39-0.86), OR_{hom} 0.27 (0.09-0.85), $OR_{allelic}$ 0.56 (0.41-0.78), P=3.4x10^{-4}) and 26% and 49% lower odds of nonalcoholic cirrhosis for heterozygotes and homozygotes, respectively (OR_{het} 0.74 (0.60-0.93), OR_{hom} 0.51 (0.31-0.85), $OR_{allelic}$ 0.74 (0.62-0.88), P=4.5x10^{-4}). HSD17B13 rs72613567:TA was also nominally associated with lower odds of HCC.

We sought to confirm and extend these findings in the multi-ethnic Dallas Liver Study (DLS) and the Dallas Pediatric Liver Study (DPLS, Table 5). In the DLS, the TA allele was associated with lower odds of any liver disease in an allele-dosage dependent manner (OR_{het} 0.74 (0.57-0.97), OR_{hom} 0.41 (0.21-0.83), $OR_{allelic}$ 0.70 (0.5-0.88), P=1.8x10^{-3}, Figure 2B). Similar effects were observed across EHR-derived liver disease subtypes, including protective associations with advanced, cirrhotic forms of alcoholic ($OR_{allelic}$ 0.72 (0.53-0.99), P=4.4x10^{-2}) and nonalcoholic ($OR_{allelic}$ 0.65 (0.40-1.07), P=9.0x10^{-2}) liver disease. In subset analyses of individuals grouped by self-reported ethnicity, the association with liver disease was significant in Hispanic Americans (n=326 cases and 722 controls, $OR_{allelic}$ 0.51 (0.35-0.74), P=4.0x10^{-4}); similar numerical trends, which did not achieve statistical significance, were also noted in the African American (n=33 cases and 2,291 controls, $OR_{allelic}$ 0.74 (0.25-2.47), P=0.67) and European American (n=158 cases and 1,266 controls, $OR_{allelic}$ 0.87 (0.65-1.15), P=0.32) subsets of the DLS. In the DPLS, a separate study of Hispanic American pediatric liver disease patients and obese controls, the TA allele was also associated with lower odds of liver disease ($OR_{allelic}$ 0.61 (0.37-0.99), P=4.6x10^{-2}). Thus, HSD17B13 rs72613567:TA was associated with reduced odds of multiple forms of chronic liver disease, including cirrhosis, in adults and children in three independent populations.

Table 4. Association of twelve exome-wide significant and replicating single nucleotide variants with liver disease phenotypes in the discovery cohort.

CHR:BP:Ref:Alt	**Gene**	**rsID**	**Alcoholic liver disease**		**Alcoholic cirrhosis**	
			OR (95% CI)	**P-value**	**OR (95% CI)**	**P-value**
4:88231392:T:TA	*HSD17B13*	rs72613567	0.62 (0.48-0.81)	***1.82E-04**	0.56 (0.41-0.78)	***3.35E-04**
8:145730161:C:T	*GPT*	rs201815297	3.83 (1.05-13.94)	8.88E-02	6.33 (1.71-	2.88E-02

					23.43)	
8:145732114:G:C	*GPT*	rs141505249	0.77 (0.06-10.73)	8.43E-01	1.13 (0.08-15.39)	9.30E-01
8:145732180:G:C	*GPT*	rs147998249	0.73 (0.05-11.76)	8.17E-01	1.07 (0.07-17.16)	9.60E-01
10:18242311:A:G	*SLC39A12*	rs10764176	0.85 (0.68-1.07)	1.64E-01	0.92 (0.70-1.22)	5.80E-01
10:101157378:CGTT:C	*GOT1*		4.60 (0.25-86.41)	3.93E-01	7.11 (0.38-133.19)	3.00E-01
10:101165533:G:C	*GOT1*	rs374966349	2.20 (0.13-37.68)	6.24E-01	3.47 (0.20 - 59.04)	4.70E-01
14:94844947:C:T	*SERPINA1*	rs28929474	2.49 (1.49-4.17)	2.30E-03	3.35 (1.93-5.83)	***3.01E-04**
19:19379549:C:T	*TM6SF2*	rs58542926	1.47 (1.06-2.04)	2.76E-02	1.35 (0.89-2.04)	1.80E-01
22:44324727:C:G	*PNPLA3*	rs738409	1.76 (1.43-2.18)	***4.98E-07**	2.07 (1.60-2.67)	***1.08E-07**
22:44324730:C:T	*PNPLA3*	rs738408	1.77 (1.43-2.18)	***4.70E-07**	2.07 (1.61-2.67)	***1.03E-07**
22:44368122:A:G	*SAMM50*	rs3761472	1.90 (1.52-2.38)	***1.36E-07**	2.28 (1.75-2.98)	***1.83E-08**

* Indicates P-values meeting the Bonferroni significance threshold of $P < 2.08 \times 10^{-3}$.

Table 4 (cont.)

CHR:BP:Ref:Alt	**Gene**	**rsID**	**Nonalcoholic liver disease**		**Nonalcoholic cirrhosis**		**Hepatocellular carcinoma**	
			OR (95% CI)	**P-value**	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**
4:88231392:T:TA	*HSD17B13*	rs72613567	0.84 (0.78-0.91)	***1.31E-05**	0.74 (0.62-0.88)	***4.48E-04**	0.67 (0.45-1.00)	4.66E-02
8:145730161:C:T	*GPT*	rs201815297	0.23 (0.04-1.14)	1.86E-02	1.25 (0.24-6.38)	7.98E-01	3.66 (0.70-19.01)	2.01E-01

8:145732114:G:C	*GPT*	rs141505249	1.02 (0.49-2.11)	9.70E-01	0.36 (0.02-5.37)	3.82E-01	1.84 (0.15-23.25)	6.88E-01
8:145732180:G:C	*GPT*	rs147998249	1.03 (0.49-2.17)	9.30E-01	0.34 (0.02-5.59)	3.67E-01	1.74 (0.11-27.05)	7.21E-01
10:18242311:A:G	*SLC39A12*	rs10764176	0.92 (0.86 (0.99)	3.43E-02	1.03 (0.88-1.21)	7.15E-01	1.29 (0.93-1.79)	1.37E-01
10:101157378:CGTT:C	*GOT1*		2.37 (0.61-9.27)	2.50E-01	8.27 (1.44-47.49)	5.92E-02	9.81 (0.52-183.54)	2.43E-01
10:101165533:G:C	*GOT1*	rs374966349	1.63 (0.53-4.96)	4.20E-01	1.17 (0.07-20.09)	9.13E-01	5.37 (0.32-91.12)	3.55E-01
14:94844947:C:T	*SERPINA1*	rs28929474	1.50 (1.21-1.87)	***5.29E-04**	2.99 (2.11-4.24)	***9.08E-08**	1.86 (0.74-4.67)	2.40E-01
19:19379549:C:T	*TM6SF2*	rs58542926	1.36 (1.21-1.52)	***2.42E-07**	1.64 (1.31-2.05)	***6.04E-05**	1.93 (1.22-3.04)	1.08E-02
22:44324727:C:G	*PNPLA3*	rs738409	1.65 (1.54-1.78)	***1.31E-41**	2.05 (1.76-2.38)	***1.70E-19**	2.20 (1.60-3.02)	***5.59E-06**
22:44324730:C:T	*PNPLA3*	rs738408	1.65 (1.54-1.78)	***1.42E-41**	2.05 (1.77-2.38)	***1.45E-19**	2.20 (1.60-3.03)	***5.41E-06**
22:44368122:A:G	*SAMM50*	rs3761472	1.52 (1.41-1.65)	***7.33E-24**	1.86 (1.58-2.19)	***1.81E-12**	1.66 (1.16-2.39)	1.05E-02

Table 5. Demographics and clinical characteristics of genotyped multi-ethnic cases and controls from the Dallas Liver and Pediatric Liver Studies.

Characteristic	**Dallas Liver Study Cases (N = 517)**	**Dallas Liver Study Controls (N = 4,279)**	**Dallas Pediatric Liver Study Cases (N = 203)**	**Dallas Pediatric Liver Study Controls (N = 244)**
Age (years) – median (IQR)	55 (48 - 60)	44 (36 - 53)	12 (10 - 15)	12 (11 - 14)
Female sex – number (%)	277 (54)	2,494 (58)	65 (32)	126 (52)

Body mass index – median (IQR)	30 (27 - 35)	30 (26 - 35)	30 (27 - 34)	31 (28 - 35)
Self-reported ethnicity				
African American	33 (6)	2,291 (54)	-	-
European American	158 (31)	1,266 (30)	-	-
Hispanic American	326 (63)	722 (17)	203 (100)	244 (100)
Presence of liver disease (by ICD-9 code) – N (%)				
Alcoholic liver disease	223 (43)	-	-	-
Alcoholic cirrhosis	215 (42)	-	-	-
Nonalcoholic, non-viral liver disease	212 (20)	-	-	-
Nonalcoholic cirrhosis	100 (19)	-	-	-
Hepatocellular carcinoma	44 (9)	-	-	-
No liver disease	-	4,279 (100)	-	-244 (100)

Association of HSD17B13 rs72613567:TA with Liver Pathology

NAFLD describes a disease spectrum ranging from liver fat accumulation without evidence of significant inflammation (simple steatosis), to more clinically impactful NASH. To confirm the association between the HSD17B13 rs72613567:TA and EHR-derived liver disease diagnoses codes, and to further understand its association with histopathological progression of steatosis to NASH, we performed tests of association in the GHS bariatric surgery cohort. In this cohort of 2,391 of the whole exome sequenced individuals assessed by liver biopsy at the time of bariatric surgery, a total of 555 (23%) individuals had no evidence of steatosis, steatohepatitis, or fibrosis ("normal"), 830 (35%) had simple steatosis, and 1006 (42%) had NASH. When comparing prevalence of normal liver, simple steatosis, and NASH by genotype, it was observed that the prevalence of normal liver did not appear to differ by genotype (23%, 24%, and 23% for T/T, T/TA, and TA/TA carriers, respectively, P = 0.5 by Chi-squared test for trend in proportions), but that the prevalence of NASH decreased (45%, 40%, and 31% for T/T, T/TA, and TA/TA carriers, respectively, $P = 1.6x10^{-4}$) and that of simple steatosis increased (33%, 35%, and 47% for T/T, T/TA, and TA/TA carriers, respectively, $P = 1.1x10^{-3}$) with each TA allele (Figure 3A). Among individuals with steatosis, the TA allele was associated with statistically significantly lower odds of both NASH and fibrosis, as compared to simple steatosis ($OR_{allelic}$ 0.77 (0.66-0.90), $P=6.5x10^{-4}$ for NASH; $OR_{allelic}$ 0.74 (0.62-0.88), $P=4.15x10^{-4}$ for

fibrosis; Figure 3B), in an allele dosage-dependent manner. Altogether, these data suggest a role for HSD17B13 in mediating NAFLD progression from simple steatosis to more advanced stages of NASH and fibrosis.

Association of HSD17B13 rs72613567:TA with Clinical Quantitative Traits and Diagnoses

To more comprehensively examine the clinical consequences of the HSD17B13 splice variant, we performed a phenome-wide study of associations of HSD17B13 rs72613567:TA with 405 quantitative EHR-derived anthropometric, vital sign, laboratory, electrocardiographic, echocardiographic, and bone densitometry measurements, and also with 3,168 EHR-derived clinical diagnoses. Using Bonferroni significance thresholds of 1.23×10^{-4} and 1.58×10^{-5} for associations with quantitative clinical measurements and clinical diagnoses, respectively, we identified statistically significant associations of the HSD17B13 rs72613567:TA allele with higher platelet counts, in addition to the associations with hepatic transaminases (Table 6). There were no statistically significant associations with clinical diagnoses other than chronic liver disease (OR (95% CI) =0.88 (0.84-0.93); P = 9.14×10^{-6}; AAF = 0.263; N Cases total = 4031, T/T = 2331, T/TA = 1449, TA/TA = 251; N Controls Total = 35701, T/T = 19238, T/TA = 13984, TA/TA = 2479).

Table 6. Phenome-Wide Study of Associations of HSD17B13 rs72613567:TA with Quantitative Clinical Measurements

Phenotype	**Effect**	**SE**	**P**	**AAF**	**N**			
					Total	**T/T**	**T/TA**	**TA/TA**
Alanine Aminotransferase median:Adjusted(Residual Log)	-0.009	0.001	1.74E-12	0.264	44038	23868	17115	3055
Aspartate Aminotransferase median:Adjusted(Residual Log)	-0.006	0.001	2.75E-11	0.264	43370	23493	16851	3026
Alanine Aminotransferase max:Adjusted(Residual Log)	-0.013	0.002	1.39E-09	0.264	43905	23797	17065	3043
Aspartate Aminotransferase max:Adjusted(Residual Log)	-0.010	0.002	8.73E-09	0.264	42733	23145	16609	2979
Platelets median:Adjusted(Residual Log)	0.004	0.001	1.44E-08	0.264	46182	25020	17944	3218
Alanine Aminotransferase min:Adjusted(Residual Log)	-0.008	0.002	2.47E-07	0.264	44029	23864	17111	3054

Platelets min:Adjusted(Residual)	1.919	0.443	1.47E-05	0.264	46181	25020	17943	3218
Platelets max:Adjusted(Residual Log)	0.004	0.001	3.03E-05	0.264	46165	25014	17936	3215
Aspartate Aminotransferase min:Adjusted(Residual Log)	-0.004	0.001	5.00E-05	0.264	43327	23471	16831	3025
Gray shading indicates P-values meeting the Bonferroni significane threshold of $P < 1.23x10^{-4}$.								
Abbreviations: AAF, alternate allele frequency; SE, standard error.								

Effect of HSD17B13 rs72613567:TA on HSD17B13 mRNA and HSD17B13 Protein Expression

We next examined the effect of the HSD17B13 rs72613567:TA allele on expression of known and novel transcripts of the gene. We used RNA sequencing to assess HSD17B13 mRNA expression in histologically normal liver samples from 22 T/T homozygous, 30 T/TA heterozygous, and 17 TA/TA homozygous carriers of the HSD17B13 rs72613567 splice variant. In addition to the two known HSD17B13 transcripts, A and B, two novel transcripts were identified: transcript C, which lacked exon 6, and transcript D which contained an insertion of a guanine nucleotide at the 3' end of exon 6, which would be predicted to result in premature truncation of the protein. Four additional transcripts (E-H) were expressed at very low levels (Figures 6A, 6B, 6C, 6D, 6E, 6F, 6G, and 6H). The transcripts were validated by RT-PCR and Sanger sequencing. The D transcript was also validated using long read cDNA sequencing. Protein sequence alignment of all identified HSD17B13 isoforms (A-H) is shown in Figures 7A and 7B. The expression levels of these transcripts varied according to HSD17B13 rs72613567 genotype; levels of transcripts A and B decreased, while those of transcripts C and D increased in an allele dosage-dependent manner with each TA allele (Figure 4A). Transcript A, which encodes the full-length 300 amino acid protein, was the predominant transcript in T/T homozygotes, while transcript D, which encodes the prematurely truncated protein, was the predominant transcript in TA/TA homozygotes. In human liver biopsy tissue, the truncated isoform D protein was minimally present in heterozygotes and TA/TA homozygotes, and isoform A protein abundance was reduced in an allele dosage-dependent manner (Figure 4B-4C). Heterologous expression of isoforms A and D in HEK 293 cells indicated reduced abundance of isoform D relative to mRNA expression, suggesting instability of the D isoform when compared to isoform A (Figure 8). These data are consistent with HSD17B13 rs72613567 altering mRNA splicing, resulting in the synthesis of a truncated form of the protein with substantially reduced expression in human liver.

Expression of HSD17B13 in Human Liver Cells

HSD17B13 is expressed primarily in the liver (Liu et al., Acta Biochim. Pol. 2007, 54, 213-218), where it localizes to lipid droplets (Su et al., Proc. Natl. Acad. Sci. USA, 2014, 111, 11437-11442, doi:10.1073/pnas.1410741111), consistent with a role in the pathogenesis of fatty liver disease. We evaluated the expression of HSD17B13 and its localization in an immortalized human liver cell line stably transduced with lentivirus expressing HSD17B13 transcript A or D. HSD17B13 isoform A was mainly detected on membranes surrounding BODIPY-labeled lipid droplets (Figure 4D). Similar subcellular localization was observed for HSD17B13 isoform D at the lipid droplet surface (Figure 4D and Figure 9). No differences in intracellular triglyceride content were observed with oleic acid treatment of cell lines overexpressing GFP control or HSD17B13 isoforms A or D (Figure 10).

Effect of rs72613567:TA on HSD17B13 Activity in vitro and in Cellular Models

To understand the functional consequences of premature truncation of the HSD17B13 protein due to rs72613567:TA, we evaluated the enzymatic activity of isoforms A and D *in vitro* using recombinant protein and nicotinamide adenosine dinucleotide as cofactor. We tested 265 unique putative substrates, and identified steroid substrates and bioactive lipids (e.g. leukotriene B4) as enzymatic substrates of HS17B13. We focused subsequent characterization of HSD17B13 enzymatic activity on enzymatic conversion of estradiol (V_{max} and K_m values in Figure 11), which resulted in oxidation of a hydroxyl to a ketone group. HSD17B13 isoform D showed greatly reduced activity towards estradiol *in vitro* (Figure 4E) and in cell-based enzymatic conversion assays (Figure 4F) when compared to HSD17B13 isoform A.

By linking large-scale exome sequencing to EHR-derived clinical phenotypes, we identified a novel association between a splice variant in HSD17B13 and decreased serum transaminase levels, as well as reduced risk of nonalcoholic and alcoholic forms of liver disease. These associations were observed consistently in four independent cohorts, and across several different liver disease categories, including advanced cirrhotic forms of liver disease and HCC. The HSD17B13 rs72613567:TA allele was not associated with simple steatosis, but was associated with reduced risk of NASH and fibrosis, suggesting that this variant allele protects from progression to more clinically advanced stages of chronic liver disease. In a phenome-wide association study, HSD17B13 rs72613567:TA was not significantly associated with clinical diagnoses or measurements other than chronic liver disease and associated clinical measurements (hepatic transaminases and platelet counts), suggesting that the clinical effects of the variant allele may be specific to chronic liver disease.

Other hydroxysteroid 17-beta dehydrogenase family members are involved in sex steroid and fatty acid metabolism (Moeller, Mol. Cell. Endocrinol., 2009, 301, 7-19,

doi:10.1016/j.mce.2008.10.040), but little is known about the function of HSD17B13. HSD17B13 overexpression was shown previously to increase lipogenesis in mouse liver, and to increase the number and size of lipid droplets in cultured hepatocytes (Su et al., Proc. Natl. Acad. Sci. USA, 2014, 111, 11437-11442, doi:10.1073/pnas.1410741111). Two previous studies also showed that hepatic expression of HSD17B13 protein is increased in patients with fatty liver (Su et al., Proc. Natl. Acad. Sci. USA, 2014, 111, 11437-11442, doi:10.1073/pnas.1410741111; Kampf et al., FASEB J., 2014, 28, 2901-2914, doi:10.1096/fj.14-250555). Our data suggest that both HSD17B13 isoforms are expressed on the lipid droplet membrane, but do not appear to modulate intracellular neutral fat content, a finding that mirrors the lack of an association between the HSD17B13 rs72613567:TA and simple steatosis in humans. Although the physiological substrates of HSD17B13 are not known, enzymatic studies demonstrate that the HSD17B13 isoform encoded by the HSD17B13 rs72613567:TA allele is catalytically defective against estradiol. While at this time it is not clear if any of the substrates tested are critical for liver disease, it is intriguing that HSD17B13 has enzymatic activity against several bioactive lipid species (e.g. leukotriene B4) that have previously been implicated in lipid-mediated inflammation (Li et al., Nature Medicine, 2015, 21, 239-247, doi:10.1038/nm.3800).

This HSD17B13 variant may provide an avenue to new therapeutic strategies targeting chronic liver disease, similar to genetic variants that have guided the way to new therapeutics in other domains. Our data indicate that HSD17B13 modulates progression of liver disease from steatosis to later stages of NASH, fibrosis, and cirrhosis, which are associated with significant morbidity and mortality, and for which there are currently no effective treatments.

What is Claimed is:

1. A method of detecting a variant HSD17B13 gene in a human subject, comprising or consisting of performing an assay on a biological sample obtained from the human subject, wherein the assay determines whether a thymine is inserted between positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene, or whether a thymine is present at a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene, wherein the presence of said thymine is indicative for a variant HSD17B13 gene, wherein the presence of said thymine is indicative for a variant HSD17B13 gene encoding HSD17B13 Isoforms D-F, F', G, or H.

2. The method of claim 1, wherein the assay comprises or consists of:

a) sequencing a portion of the HSD17B13 gene including positions corresponding to positions 12665 and 12666 of SEQ ID NO:1, or including a position corresponding to position 12666 of SEQ ID NO:2; or

b) i) contacting the biological sample with a primer hybridizing to a region of the HSD17B13 gene that is within 50 nucleotides of a position of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1, or within 50 nucleotides of a position of the HSD17B13 gene corresponding to position 12666 of SEQ ID NO:2;

b) ii) extending the primer at least through the position of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1, or corresponding to position 12666 of SEQ ID NO:2; and

b) iii) determining whether a thymine is inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene, or whether a thymine is present at a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene, in an extension product of the primer.

3. A method of determining a human subject's susceptibility or risk of developing a liver disease, comprising or consisting of:

a) performing an assay on a biological sample obtained from the human subject, wherein the assay determines whether a thymine is inserted between positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene, or whether a thymine is present at a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene; and

b) classifying the human subject as being at decreased risk for developing the liver disease if a thymine is inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene or if a thymine is present at a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene, or classifying the human subject as being at increased risk for developing the liver disease if a thymine is not inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene or if a thymine is not present at a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene.

4. The method of claim 3, wherein the liver disease is selected from the group consisting of fatty liver disease, nonalcoholic fatty liver disease (NAFLD), alcoholic liver fatty liver disease, cirrhosis, viral hepatitis, hepatocellular carcinoma, simple steatosis, steatohepatitis, fibrosis, and non-alcoholic steatohepatitis (NASH).

5. The method of any one of claims 3 to 4, wherein the assay comprises or consists of:

a)

i) contacting the biological sample with a primer hybridizing to a region of the HSD17B13 gene that is within 50 nucleotides of positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene, or within 50 nucleotides of positions corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene;

ii) extending the primer at least through the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene, or corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene; and

iii) determining whether a thymine is inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene, or whether a thymine is present at a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene, in an extension product of the primer; or

b) contacting the biological sample with a primer or probe that specifically hybridizes to the variant HSD17B13 gene having a thymine at a position corresponding to position 12666 of SEQ ID NO:2, and does not hybridize to the corresponding wild type HSD17B13 gene under stringent conditions, and determining whether hybridization has occurred.

6. A method of determining a human subject's susceptibility or risk for developing a liver disease, comprising or consisting of:

a) performing an assay on a biological sample obtained from the human subject, wherein the assay determines the presence of an HSD17B13 Transcript D in the biological sample; and

b) classifying the human subject as being at decreased risk for developing the liver disease if an HSD17B13 Transcript D is present in the biological sample, or classifying the human subject as being at increased risk for developing the liver disease if an HSD17B13 Transcript D is not present in the biological sample:

wherein the HSD17B13 Transcript D comprises or consists of a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to SEQ ID NO:6, 15, 24, or 33; or

wherein the HSD17B13 Transcript D is RNA and comprises or consists of SEQ ID NO:6, or a cDNA thereof comprising or consisting of SEQ ID NO:24, or wherein the HSD17B13 Transcript D is mRNA and comprises or consists of SEQ ID NO:15, or a cDNA thereof comprising or consisting of SEQ ID NO:33.

7. A method of determining a human subject's susceptibility or risk for developing a liver disease, comprising or consisting of:

a) detecting whether HSD17B13 Isoform D is present in a biological sample obtained from the human subject; and

b) classifying the human subject as being at decreased risk for developing the liver disease if HSD17B13 Isoform D is detected in the biological sample, or classifying the human subject as being at decreased risk for developing the liver disease if HSD17B13 Isoform D is not detected in the biological sample.

8. A method of determining a human subject's risk for progression to more clinically advanced stages of fatty liver disease, comprising or consisting of:

a) performing an assay on a biological sample obtained from the human subject, wherein the assay determines whether a thymine is inserted between positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene, or whether a thymine is present at a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene; and

b) classifying the human subject as being at decreased risk for progression to more clinically advanced stages of fatty liver disease if a thymine is inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene or if a thymine is present at a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene, or classifying the human subject as being at increased risk for progression to more clinically advanced stages of fatty liver disease if a thymine is not inserted between the positions corresponding to positions 12665 and 12666 of SEQ ID NO:1 of the wild type HSD17B13 gene, or if a thymine is not present at a position corresponding to position 12666 of SEQ ID NO:2 of the variant HSD17B13 gene.

9. The method of claim 8, wherein the assay comprises or consists of:

i) contacting the biological sample with a primer hybridizing to a region of the HSD17B13 gene that is within 50 nucleotides of positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1, or corresponding to position 12666 of SEQ ID NO:2;

ii) extending the primer at least through the positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1, or corresponding to position 12666 of SEQ ID NO:2; and

iii) determining whether a thymine is inserted between the positions of the HSD17B13 gene corresponding to positions 12665 and 12666 of SEQ ID NO:1, or whether a thymine is present at a position of the variant HSD17B13 gene corresponding to position 12666 of SEQ ID NO:2, in an extension product of the primer; or

wherein the assay comprises or consists of contacting the biological sample with a primer or probe that specifically hybridizes to the variant HSD17B13 gene having a thymine at a position corresponding to position 12666 of SEQ ID NO:2, and not to the corresponding wild type HSD17B13 gene under stringent conditions, and determining whether hybridization has occurred.

10. A method of determining a human subject's risk for progression to more clinically advanced stages of fatty liver disease, comprising or consisting of:

a) performing an assay on a biological sample obtained from the human subject, wherein the assay determines the presence of an HSD17B13 Transcript D in the biological sample; and

b) classifying the human subject as being at decreased risk for progression to more clinically advanced stages of fatty liver disease if an HSD17B13 Transcript D is present in the biological sample, or classifying the human subject as being at increased risk for progression to more clinically advanced stages of fatty liver disease if an HSD17B13 Transcript D is not present in the biological sample.

11. The method of claim 10, wherein:

the HSD17B13 Transcript D comprises or consists of a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to SEQ ID NO:6, 15, 24, or 33; or

the HSD17B13 Transcript D is RNA and comprises or consists of SEQ ID NO:6, or a cDNA thereof comprising or consisting of SEQ ID NO:24, or wherein the HSD17B13 Transcript D is mRNA and comprises or consists of SEQ ID NO:15, or a cDNA thereof comprising or consisting of SEQ ID NO:33.

12. The method of any one of claims 10 to 11, wherein the assay:

determines the expression level of HSD17B13 Transcript D in the biological sample, wherein an increased expression level of HSD17B13 Transcript D compared to a control sample from a control human subject homozygous for a wild type HSD17B13 allele indicates a decreased risk for progression to more clinically advanced stages of fatty liver disease, and wherein the same or a decreased expression level of HSD17B13 Transcript D compared to the control sample indicates an increased risk for progression to more clinically advanced stages of fatty liver disease; or

comprises or consists of contacting the biological sample with one or more primers or probes that specifically hybridize to a nucleic acid sequence of HSD17B13 Transcript D, or to the complement thereof, and determining whether hybridization has occurred.

13. The method of any one of claims 10 to 12, further comprising specifically detecting Transcript D by using a nucleic acid molecule comprising from about 5 nucleotides up to about 50 nucleotides comprising or consisting of:

i) a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a nucleotide sequence of a SEQ ID NO:6, 15, 24, or 33, or to the complement thereof;

ii) a nucleic acid molecule which specifically hybridizes to exon 2 of Transcript D; and/or

iii) a nucleic acid molecule which specifically hybridizes to the region which bridges exons 3 and 4 of Transcript D.

14. The method of claim 13, wherein the one or more primers or probes specifically hybridize to SEQ ID NO:6, SEQ ID NO:15, SEQ ID NO:24, and/or SEQ ID NO:33.

15. A method of determining a human subject's risk for progression to more clinically advanced stages of fatty liver disease, comprising or consisting of:

a) detecting whether HSD17B13 Isoform D is present in a biological sample obtained from the human subject; and

b) classifying the human subject as being at decreased risk for progression to more clinically advanced stages of liver disease if HSD17B13 Isoform D is detected in the biological sample.

16. The method of claim 15, wherein the HSD17B13 Isoform D comprises or consists of an amino acid sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to SEQ ID NO:42.

17. A nucleic acid molecule comprising or consisting of at least 15 contiguous nucleotides of an HSD17B13 gene, wherein the contiguous nucleotides are at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to a corresponding sequence in SEQ ID NO:2, and having a thymine at a position corresponding to position 12666 of SEQ ID NO:2 when used for detecting a variant HSD17B13 gene in a human subject, wherein the nucleic acid molecule is linked to a heterologous nucleic acid or comprises a heterologous label.

18. A nucleic acid molecule comprising or consisting of a nucleotide sequence that encodes a polypeptide having an amino acid sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to the amino acid sequence of HSD17B13 Isoform D (SEQ ID NO:42) when used for detecting a variant HSD17B13 gene in a human subject, wherein the nucleic acid molecule is linked to a heterologous nucleic acid or comprises a heterologous label.

19. The nucleic acid molecule of claim 18, comprising or consisting of a nucleotide sequence that:

encodes a polypeptide having the amino acid sequence of HSD17B13 Isoform D (SEQ ID NO:42); or

is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a nucleotide sequence of an HSD17B13 Transcript D (SEQ ID NO:6, 15, 24, or 33).

20. The nucleic acid molecule of any one of claims 18 to 19, wherein the nucleic acid molecule is RNA and comprises or consists of SEQ ID NO:6, or a cDNA thereof comprising or consisting of SEQ ID NO:24, or wherein the nucleic acid molecule is mRNA and comprises or consists of SEQ ID NO:15, or a cDNA thereof comprising or consisting of SEQ ID NO:33.

21. A nucleic acid molecule comprising or consisting of from about 5 nucleotides up to about 50 nucleotides that specifically hybridizes to an HSD17B13 gene at a region that includes a position corresponding to position 12666 in SEQ ID NO:2, or the complement thereof, and wherein the nucleic acid molecule specifically hybridizes to an HSD17B13 gene having a thymine at a position corresponding to position 12666 of SEQ ID NO:2, or to the complement thereof when used for detecting a variant HSD17B13 gene in a human subject, wherein the nucleic acid molecule is linked to a heterologous nucleic acid or comprises a heterologous label.

22. A nucleic acid molecule when used for detecting a variant HSD17B13 gene in a human subject, wherein the nucleic acid molecule comprises or consists of from about 5 nucleotides up to about 50 nucleotides that:

a) specifically hybridizes to variant HSD17B13 Transcript D, and wherein the nucleic acid molecule specifically hybridizes to:

i) a nucleotide sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a nucleotide sequence of a SEQ ID NO:6, 15, 24, or 33, or

ii) to the complement of the nucleotide sequence of i) and

wherein the nucleic acid molecule is linked to a heterologous nucleic acid or comprises a heterologous label; or

b) comprises or consists of:

i) the nucleic acid molecule of a);

ii) a nucleic acid molecule which specifically hybridizes to exon 2 of Transcript D; and/or

iii) a nucleic acid molecule which specifically hybridizes to the region which bridges exons 3 and 4 of Transcript D,

wherein the nucleic acid molecule is linked to a heterologous nucleic acid or comprises a heterologous label.

23. The nucleic acid molecule of claim 22, wherein the nucleic acid molecule specifically hybridizes to an RNA molecule comprising or consisting of SEQ ID NO:6, or a cDNA thereof comprising or consisting of SEQ ID NO:24, or wherein the nucleic acid molecule specifically hybridizes to an mRNA comprising or consisting of SEQ ID NO:15, or a cDNA thereof comprising or consisting of SEQ ID NO:33, or the complement thereof.

24. The nucleic acid molecule of any one of claims 21 to 23, when used for detecting a variant HSD17B13 transcript, for determining a human subject's susceptibility or risk of developing a liver disease, or for determining a human subject's risk for progression to more clinically advanced stages of fatty liver disease.

25. A vector comprising the nucleic acid molecule of any one of claims 17 to 24.

26. A cell comprising the nucleic acid molecule of any one of claims 17 to 24, or the vector of claim 25.

27. A polypeptide comprising or consisting of an amino acid sequence that is at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100%

identical to the amino acid sequence of HSD17B13 Isoform D (SEQ ID NO:42), wherein the polypeptide is linked to a heterologous molecule.

28. A composition comprising the nucleic acid molecule of any one of claims 17 to 24, the vector of claim 25, the cell of claim 26, or the polypeptide of claim 27.