

(10) International Publication Number
WO 2016/201438 A1(43) International Publication Date
15 December 2016 (15.12.2016)(51) International Patent Classification:
A61K 31/7088 (2006.01) *C40B 30/04* (2006.01)
C12Q 1/68 (2006.01)(21) International Application Number:
PCT/US2016/037242(22) International Filing Date:
13 June 2016 (13.06.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/174,009 11 June 2015 (11.06.2015) US

(71) Applicant: THE CHILDREN'S HOSPITAL OF PHILADELPHIA [US/US]; 34th Street and Civic Center Boulevard, Philadelphia, PA 19104 (US).

(72) Inventors: JOHNSON, Matthew, E.; 1200 Lincoln Avenue, Apt. 13, Prospect Park, PA 19076 (US). XIA, Qianhua; 52-5 Revere Road, Drexel Hill, PA 19026 (US). JOHNSTON, Brian, T.; 400 Walnut Street, Apt 61, Philadelphia, PA 19106 (US). CHESI, Alessandra; 52 S. 3rd Street, Philadelphia, PA 19106 (US). GRANT, Struan, F. A.; 205 Darthmouth Avenue, Philadelphia, PA 19081 (US).

(74) Agents: RIGAUT, Kathleen, D. et al.; Dann, Dorfman, Herrell & Skillman, 1601 Market Street, Suite 2400, Philadelphia, PA 19103-2307 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

(54) Title: NEW TARGETS MODULATED BY A CASUAL VARIANT OF TYPE 2 DIABETES (T2D) EMBEDDED WITHIN THE TCF7L2 GENE AND METHODS OF USE THEREOF FOR IDENTIFYING AGENTS HAVING EFFICACY FOR THE TREATMENT OF TYPE 2 DIABETES AND OTHER METABOLIC DISORDERS

FIG. 1A

Upstream

1 CGTCTCTTT GGCTGAAGGI TTTCGCTGC AAGCTTGCAT TGTATGACG TGACACCTGA
61 CGTATCTGTA AATTTCTICA AATTTCTAAG TGTATTACAA CCCCCTGTGC AAAAGATGAT
121 TAATTAATTG CCTTGACAGI AAAACAAAAA AAAAAAAGG GTGTGGGGG TATATGGTAT
181 CCTGATTTA CTATAGAAGA TGCAGAGAGT GAAGGGAGAT GAGGTGGGGA GGAGGGGCC

Downstream

1 CTTGGGTICA AATGATCTC CIGGCTCAGC CTCTCTTAAG CAGTGCCTTT CCCCATTCTC
61 ATGGGACITT CCAATCCATG AGATACCTTG CTGCAGGGAA GCCCTGTCTG TCCAGGCCCTC
121 TGTAAATAGAC GACTTCACAT GTTCCTGTGT TGTGTCTTGC CTTCCTGTGTG GCTAAGTTTC
181 CATGACCTGG TGGCTTGGAA GCCCATGCC TGATTGTGTG GAGAGGCAGG GAGGCACCTT

Guide sequence

1. ACAGATACGTCAGGTGTCACIGG	6. AGTCGTCTATTACACAGGCTGG
2. GGATTACAATGCAAGGTTGCAGG	7. GTGTAATAGACGACTCCACATGG
3. TTAATCATCTTTGACACGEGG	8. TGTGAAGTCGTCTATACACAGG
4. AAGAAATTTACAGATACGTCAGG	9. GTGGCTAAGTTTCCATGACCTGG
5. TGTCACTGGATTCAATGCAAGG	10. TCCATGAGATACTTTGCTGCAGG

CRISPR/Cas9 PAM sites with the highest scores

(57) Abstract: Compositions and methods are provided which modulate the enhancer/silencer function embedded within TCF7L2 for treating metabolic disease including T2D are disclosed.



WO 2016/201438 A1

New Targets Modulated by a Causal Variant of Type 2 Diabetes (T2D) embedded within the TCF7L2 gene and Methods of Use Thereof for Identifying Agents having Efficacy for the Treatment of Type 2 Diabetes and Other Metabolic Disorders

By

Matthew E. Johnson

Qianghua Xia

Brian T. Johnston

Alessandra Chesi

Struan F. A. Grant

This application claims priority to US Provisional Application No. 62/174,009 filed June 11, 2015, the entire disclosure being incorporated herein as though set forth in full.

Field of the Invention

This invention relates to the fields of transcriptional regulation of metabolic disease and drug screening. More specifically, the invention provides new gene targets within and outside the same topologically associating domain (TAD) as *TCF7L2* and methods of use thereof for identifying new agents useful for modulating fatty acid metabolism for the treatment of metabolic diseases, particularly, T2D.

Background of the Invention

Several publications and patent documents are cited throughout the specification in order to describe the state of the art to which this invention pertains. Each of these citations is incorporated herein by reference as though set forth in full.

Common genetic variation within the gene encoding transcription factor 7-like 2 (*TCF7L2*, formerly known as *TCF4*) is strongly associated with type 2 diabetes (T2D). This observation was reported as early as 2006, when we first published the association [3]. Very recently, trans-ethnic meta-analysis of T2D genome wide association studies (GWAS) [4] have also implicated the *TCF7L2* locus. This gene encodes a transcription factor that is considered a 'master regulator' of the canonical Wnt signaling pathway and plays a role in multiple development processes. Indeed, we previously revealed via ChIP-seq that the genes bound by

TCF7L2 are significantly enriched for both endocrine-related pathways and GWAS-implicated loci for various cardio-metabolic traits

Many studies have presumed that *TCF7L2* is the main T2D causal gene at this GWAS-implicated locus and thus have been carried out efforts to understand the mechanism by which
5 TCF7L2 plays a regulatory role in T2D pathogenesis; however, in this filing we show that there are multiple genes influenced by this associated variant, of which *TCF7L2* is one.

The tissues affected by *TCF7L2* genetic variation in T2D have yet to be identified. Studies have focused on multiple candidate tissues, including pancreatic islets[5], liver[6], adipose[7] and intestinal cells[8]. The latter is particularly compelling given that *TCF7L2* and its
10 binding partner beta-catenin mediate cell-type-specific regulation of the proglucagon (*GCG*) gene, which in turn cleaves to the insulinotropic hormone GLP-1 in the intestinal tract[8]. Furthermore, missense mutations in *TCF7L2* have been known for years to cause colorectal cancer[9], while *TCF7L2*^{-/-} mouse studies suggest that crypt stem cells of the small intestine are genetically programmed by *TCF7L2* [10]. But a fundamental issue remains, where it
15 is still unclear if *TCF7L2* is in fact even the actual principal culprit gene at this locus.

As a consequence of fine mapping efforts in Africans and African Americans, along with Bayesian refinement efforts[11-13], there is now strong consensus in the T2D genetics field that the T allele of rs7903146 within *TCF7L2* is the causal variant at this locus. However, there is still uncertainty as to whether the region harboring this variant influences the promoter of
20 *TCF7L2* itself or a promoter further away.

Summary of the Invention

To address the drawbacks of current approaches for elucidating the mechanism underlying activity of the T2D-associated variant embedded within the *TCF7L2* gene we utilized
25 the CRISPR/Cas9 system. This approach entails direct and precise gene editing for assessing the role of the immediate genomic region harboring rs7903146 in the regulation of expression of genes in a homogenous manner. These studies coupled with 3C-qPCR and 4C sequencing revealed that rs7903146 is a coordinating center for a number of genes heretofore unassociated with the T2D pathogenic phenotype.

30 In accordance with the present invention, an isolated binding complex comprising a SNP containing transcription factor 7-like 2 (*TCF7L2*) encoding nucleic acid, wherein said SNP is

rs7903146, and gene or gene promoter region selected from *ACSL5*, *RBM20*, *PDCD4*, *MIR4680*, *BBIP1*, *SHOC2*, *RPL13AP6*, *ADRA2A*, *GPAM*, *TECTB*, *MIR6715B*, *GUCY2GP*, *ZDHHC6*, *VTI1A*, *MIR4295*, *LOC103344931*, *TCF7L2*, *SNORA87A*, *HABP2*, *NRAP*, *CASP7* is provided.

In a preferred embodiment, preferably the complex comprises at least the SNP containing

5 *TCF7L2* nucleic acid and *ACSL5*.

In another aspect of the invention, a method for identifying agents which disrupt the enhancer/silencer binding complexes described herein, thereby modulating the function of *TCF7L2*, *ACSL5* and the other genes listed above is disclosed. An exemplary method comprises incubating said complex in the presence and absence of an effective amount of said agent, said
10 complex comprising at least one detectably labeled protein or nucleic acid (step a); measuring disruption of said binding complex in the presence of said agent relative to that observed in the absence of said agent (step b), agents which disrupt said complex being identified as modulators of this enhancer/silencer function. The method can be performed in vitro or in vivo within a cell. Functions of the genes listed above, including *TCF7L2* and *ACSL5*, to be modulated,
15 include, without limitation, Wnt signaling, chromatin remodeling, activation of target gene expression and DNA damage detection and repair.

Finally, the invention also provides a method for identifying agents which modulate *ACSL5* activity. An exemplary method entails incubating said cells comprising a *TCF7L2* topologically associating domain and *ACSL5* in the presence and absence of an effective amount
20 of an agent and measuring one or more parameter of *ACSL5* activity in the presence of said agent relative to that observed in the absence of said agent, agents which alter said parameter in treated versus untreated cells being identified as modulators of *ACSL5* activity. In a preferred embodiment, the one or more parameters of *ACSL5* activity are selected from modulation of fatty acid synthesis, fatty acid oxidation, insulin production, glucagon production, glycolysis and
25 glucose uptake.

In another aspect of the invention, a method for identifying an agent which modulates function of one or more genes selected from *RBM20*, *PDCD4*, *MIR4680*, *BBIP1*, *SHOC2*, *RPL13AP6*, *ADRA2A*, *GPAM*, *TECTB*, *MIR6715B*, *GUCY2GP*, *ZDHHC6*, *VTI1A*, *LOC103344931*, *MIR4295*, *TCF7L2*, *SNORA87A*, *HABP2*, *NRAP*, and *CASP7* is disclosed. An
30 exemplary method comprises incubating cells comprising an enhancer element within a topologically associated domain containing *TCF7L2* and at least one of said genes in the

presence and absence of an effective amount of said agent; and measuring one or more parameter associated with fatty acid or glucose metabolism in the presence of said agent relative to that observed in the absence of said agent, agents which alter said parameter in treated versus untreated cells being identified as modulators of fatty acid or glucose metabolism.

- 5 Preferred parameters to be assessed include, without limitation, modulation of one or more of fatty acid synthesis, fatty acid oxidation, glycolysis, insulin production, glucagon production, and glucose uptake.

Brief Description of the Drawings

10 **Figures 1A – 1G. Targeting of HCT116 cells and NCM460 cells for the genomic region harboring rs7903146.** Fig. 1A and Fig. 1B. The Cas9 sgRNA-targeting sites in HCT116 cells. The guide RNA sequences and the PAM sites are labeled in red; the guide sequences #5, #7, #12, #17 and #18 were chosen for the study. Fig. 1C. Genotyping of targeted HCT116 cells. Wildtype and 1.4kb del are shown. Expected fragment size: WT = 1.8kb, 1.4kb del = 0.4 kb. Fig. 15 1D. The 1.4kb homozygous deletion in HCT116 cells validated by Sanger sequencing. Fig. 1E. Sanger sequencing result for 1.4kb deletion in HCT116 cells aligned to the genome. Fig. 1F. Genotyping of targeted HCT116 cells. Wildtype 104bp del and 66bp del are shown. Expected fragment size: WT = 501bp, 104bp del = 90bp, 66bp del=128bp Fig. 1G. Genotyping of targeted NCM460 cells. Wildtype and 1.4kb del are shown. Expected fragment size: WT = 1.8kb, 1.4kb 20 del = 0.4 kb.

Figure 2. Corresponding topologically associating domain (TAD) for TCF7L2. The image was generated by the 3D Genome Browser using data from the GM12878 cell line. The TAD harbors the following genes and microRNA: *RBM20*, *PDCD4*, *MIR4680*, *BBIP1*, *SHOC2*, 25 *RPL13AP6*, *ADRA2A*, *GPAM*, *TECTB*, *MIR6715B*, *GUCY2GP*, *ACSL5*, *ZDHHC6*, *VTI1A*, *MIR4295*, *LOC103344931*, *TCF7L2*, *SNORA87A*, *HABP2*, *NRAP*, *CASP7*..

Figures 3A -3B. Expression analysis of *ACSL5* and *TCF7L2* following gene editing of the genomic region harboring rs7903146. Relative levels of individual gene expression in the 30 control and targeted cells, determined by quantitative PCR with GAPDH normalization. Values

are the mean of three experiments. Fig. 3A: Gene expression in the HCT116 control and targeted cells; Fig. 3B: Gene expression in the NCM460 control and targeted cells.

Figures 4A -4C. Western blot for *ACSL5* and *TCF7L2* following the CRISPR/Cas9 deletion of the region harboring *TCF7L2* SNP rs7903146. Fig. 4A. Expression of *ACSL5* and *TCF7L2* proteins in the control and targeted cells determined by Western blot. β actin was used for the loading control. The experiment was repeated three times and a representative blot is displayed. Fig. 4B and Fig. 4C. Quantification of the three independent experiments described in A represented as a bar graphs. Western blots were scanned and intensities were determined using ImageJ. Bars represent quantified Western blot signals normalized to β actin and the HCT116 control cells.

Figure 5. Circos plot of significant contacts. The yellow circle represents the *TCF7L2* TAD region (expanded to fully contain *CASP7*; hg19 coordinates chr10:112,370,010-115,490,668) with coordinates increasing clockwise starting (and ending) at 'noon'. For each gene, the transcripts (from the UCSC gene models) are depicted (without the exon/intron details) in blue when on the forward strand, in orange when on the reverse strand (*TCF7L2* has 30 UCSC transcripts, we have hidden some for ease of view). Links between the bait regions and significant 4C contacts within 5kb of any TSS are displayed in red for HTC116 and in green for HCT116 1.4kb deletion. The contacts shown are only those that were consistently detected also in the HTC116 capture C, NCM460 4C and capture C assays.

Figures 6A -6B. Western blot for *ACSL5* following *TCF7L2* siRNA knockdown in HCT116 cells. Fig. 6A. Expression of *ACSL5* and *TCF7L2* proteins in the control and knockdown cells determined by Western blot. β actin was used for the loading control. The experiment was repeated three times and a representative blot is displayed. Fig. 6B. Quantification of the three independent experiments described in A represented as a bar graph. Western blots were scanned and intensities were determined using ImageJ. Bars represent quantified Western blot signals normalized to β actin and the HCT116 control cells. Error bars indicate standard deviation. Statistical analyses were performed by two-tailed Student's t-test to

compare expression level between the control and the targeted cells and significance is denoted by asterisks where $*P < 0.001$.

Figures 7A -7C. PARP-1 inhibition: Olaparib and siRNA. Fig. 7A. Relative levels of *ACSL5* expression in the control and treated HCT116 and NCM460 cells determined by quantitative PCR with GAPDH normalization. Fig. 7B. *ACSL5* protein levels were determined by Western blot in the control and treated cells with PARP-1 inhibitor Olaparib or PARP-1 siRNA determined by Western blot. β actin was used for the loading control. The experiment was repeated three times and a representative blot is displayed. Fig. 7C. Quantification of the three independent experiments described in B represented as a bar graph. Western blots were scanned and intensities were determined using ImageJ. Bars represent quantified western blot signals normalized to β actin and the HCT116 and NCM460 control cells.

Figures 8A -8C. Allele specific effects on *ACSL5* gene expression. Fig. 8A. Relative levels of *ACSL5* gene expression in the control and targeted cells, determined by quantitative PCR with *GAPDH* normalization. Values are the mean of three experiments. Fig. 8B. Expression of *ACSL5* protein in the control and targeted cells determined by Western blot. β actin was used for the loading control. The experiment was repeated three times and a representative blot is displayed. Fig. 8C. Quantification of the three independent experiments described in B represented as a bar graph. Western blots were scanned and intensities were determined using ImageJ. Bars represent quantified Western blot signals normalized to β actin and the HCT116 control cells.

Detailed Description of the Invention

The T allele of single nucleotide polymorphism (SNP), rs7903146, embedded within the gene encoding transcription factor 7-like 2 (*TCF7L2*) is considered the strongest associated variant with type 2 diabetes (T2D) reported to date; furthermore, this variant is widely presumed to be the actual causal lesion at this locus. To aid in the identification of the actual gene(s) under the influence of this variant, we first generated a CRISPR/Cas9 mediated 1.4kb deletion of the genomic region harboring rs7903146 in the HCT116 cell line followed by global gene expression analysis. We observed 99 genes with significant differential expression (FDR cut-off=10%) and an effect size of at least two-fold. We then carried out a combination of 4C and

Capture C in HCT116 and NCM460 cell lines in order to ascertain which of these perturbed genes' promoters made consistent physical contact with the genomic region harboring the variant. This revealed just one gene, Acyl-CoA synthetase 5 (*ACSL5*), which resides in the same topologically associating domain as *TCF7L2*. Acyl-CoA synthetases (ACSL) activate long-chain fatty acids for both synthesis of cellular lipids, and degradation via beta-oxidation. ACSL5 may also activate fatty acids from exogenous sources for the synthesis of triacylglycerol destined for intracellular storage. This protein utilizes a wide range of saturated fatty acids with a preference for C16-C18 unsaturated fatty acids. We then generated additional, smaller deletions (66 and 104bp) comprising rs7903146 and observed consistently reduced *ACSL5* mRNA levels across all three deletions up to 30-fold, with commensurate loss of ACSL5 protein. Notably, these deletions abolished significantly detectable chromatin contacts with the *ACSL5* promoter. We went on to confirm that contacts between rs7903146 and the *ACSL5* promoter regions were conserved in human colon tissue. *ACSL5* encodes 'acyl-CoA synthetase long chain family, member 5', an enzyme with known roles in fatty acid metabolism. This 'variant to gene mapping' effort implicates the genomic location harboring rs7903146 as a regulatory region for *ACSL5*.

In summary, our data point to rs7903146 being a 'coordinating center' for a number of novel genes playing a role in the pathogenesis of type 2 diabetes. By altering expression of these genes, better therapeutic outcomes are predicted.

Definitions

The following definitions are provided to facilitate an understanding of the present invention:

For purposes of the present invention, "a" or "an" entity refers to one or more of that entity; for example, "a cDNA" refers to one or more cDNA or at least one cDNA. As such, the terms "a" or "an," "one or more" and "at least one" can be used interchangeably herein. It is also noted that the terms "comprising," "including," and "having" can be used interchangeably. Furthermore, a compound "selected from the group consisting of" refers to one or more of the compounds in the list that follows, including mixtures (i.e. combinations) of two or more of the compounds. The transitional terms "comprising", "consisting essentially of" and "consisting

of”, when used in the appended claims, in original and amended form, define the claim scope with respect to what unrecited additional claim elements or steps, if any, are excluded from the scope of the claim(s). The term “comprising” is intended to be inclusive or open-ended and does not exclude any additional, unrecited element, method, step or material. The term “consisting of”
5 excludes any element, step or material other than those specified in the claim, and in the latter instance, impurities ordinarily associated with the specified material(s). The term “consisting essentially of” limits the scope of a claim to the specified elements, steps or materials and those that do not materially affect the basic and novel characteristic(s) of the claimed subject matter.

According to the present invention, an isolated, or biologically pure molecule is a
10 compound that has been removed from its natural milieu. As such, “isolated” and “biologically pure” do not necessarily reflect the extent to which the compound has been purified. An isolated compound of the present invention can be obtained from its natural source, can be produced using laboratory synthetic techniques or can be produced by any such chemical synthetic route.

“T2D-associated SNP or specific marker” is a SNP or marker which is associated with an
15 increased or decreased risk of developing T2D not found normal patients who do not have this disease. Such markers may include but are not limited to nucleic acids, proteins encoded thereby, or other small molecules. Relevant information for the marker of the invention can be found in the dbSNP entry on the world wide web at ncbi.nlm.nih.gov/SNP/snp_ref.cgi?type=rs&rs=rs7903146.

20 The phrase “Type 2 diabetes (T2D)” formerly noninsulin-dependent diabetes mellitus (NIDDM) or adult-onset diabetes) makes up about 90% of cases of diabetes. T2D is a metabolic disorder due to high blood glucose caused by insulin resistance and relative insulin deficiency. Without adequate insulin, glucose builds up in the bloodstream instead of going into the cells. The body is unable to use this glucose for energy despite high levels in the bloodstream, leading
25 to increased hunger. In addition, the high levels of glucose in the blood causes the patient to urinate more, which in turn causes excessive thirst.

A “single nucleotide polymorphism (SNP)” refers to a change in which a single base in the DNA differs from the usual base at that position. These single base changes are called SNPs or “snips.” Millions of SNP's have been cataloged in the human genome. Some SNPs such that
30 which causes sickle cell are responsible for disease. Other SNPs are normal variations in the genome.

The term "genetic alteration" as used herein refers to a change from the wild-type or reference sequence of one or more nucleic acid molecules. Genetic alterations include without limitation, base pair substitutions, additions and deletions of at least one nucleotide from a nucleic acid molecule of known sequence.

5 The "ACSL5 gene" encodes an isozyme of the long-chain fatty-acid-coenzyme A ligase family. Although differing in substrate specificity, subcellular localization, and tissue distribution, all isozymes of this family convert free long-chain fatty acids into fatty acyl-CoA esters, and thereby play a key role in lipid biosynthesis and fatty acid degradation. This isozyme is highly expressed in uterus and spleen, and in trace amounts in normal brain, but has markedly
10 increased levels in malignant gliomas. This gene functions in mediating fatty acid-induced glioma cell growth. Three transcript variants encoding two different isoforms have been found for this gene. Diseases associated with ACSL5 include chronic intestinal vascular insufficiency and myelodysplastic syndrome. Related pathways having ACSL5 involvement include metabolism and regulation of lipid metabolism by peroxisome proliferator-activated receptor
15 alpha (PPARalpha). GO annotations related to this gene include long-chain fatty acid-CoA ligase activity. An important paralog of this gene is ACSBG2.

 The term "solid matrix" as used herein refers to any format, such as beads, microparticles, a microarray, the surface of a microtitration well or a test tube, a dipstick or a filter. The material of the matrix may be polystyrene, cellulose, latex, nitrocellulose, nylon, polyacrylamide, dextran
20 or agarose. Nucleic acids of the invention may be affixed or chemically immobilized onto a solid matrix.

 "Target nucleic acid" as used herein refers to a previously defined region of a nucleic acid present in a complex nucleic acid mixture wherein the defined wild-type region contains at least one known nucleotide variation which may or may not be associated with T2D. The nucleic acid
25 molecule may be isolated from a natural source by cDNA cloning or subtractive hybridization or synthesized manually. The nucleic acid molecule may be synthesized manually by the triester synthetic method or by using an automated DNA synthesizer.

 With regard to nucleic acids used in the invention, the term "isolated nucleic acid" is sometimes employed. This term, when applied to DNA, refers to a DNA molecule that is
30 separated from sequences with which it is immediately contiguous (in the 5' and 3' directions) in the naturally occurring genome of the organism from which it was derived. For example, the

"isolated nucleic acid" may comprise a DNA molecule inserted into a vector, such as a plasmid or virus vector, or integrated into the genomic DNA of a prokaryote or eukaryote. An "isolated nucleic acid molecule" may also comprise a cDNA molecule. An isolated nucleic acid molecule inserted into a vector is also sometimes referred to herein as a recombinant nucleic acid

5 molecule.

With respect to RNA molecules, the term "isolated nucleic acid" primarily refers to an RNA molecule encoded by an isolated DNA molecule as defined above. Alternatively, the term may refer to an RNA molecule that has been sufficiently separated from RNA molecules with which it would be associated in its natural state (i.e., in cells or tissues), such that it exists in a

10 "substantially pure" form.

By the use of the term "enriched" in reference to nucleic acid it is meant that the specific DNA or RNA sequence constitutes a significantly higher fraction (2-5 fold) of the total DNA or RNA present in the cells or solution of interest than in normal cells or in the cells from which the sequence was taken. This could be caused by a person by preferential reduction in the amount of

15 other DNA or RNA present, or by a preferential increase in the amount of the specific DNA or RNA sequence, or by a combination of the two. However, it should be noted that "enriched" does not imply that there are no other DNA or RNA sequences present, just that the relative amount of the sequence of interest has been significantly increased.

It is also advantageous for some purposes that a nucleotide sequence be in purified form.

20 The term "purified" in reference to nucleic acid does not require absolute purity (such as a homogeneous preparation); instead, it represents an indication that the sequence is relatively purer than in the natural environment (compared to the natural level, this level should be at least 2-5 fold greater, e.g., in terms of mg/ml). Individual clones isolated from a cDNA library may be purified to electrophoretic homogeneity. The claimed DNA molecules obtained from these

25 clones can be obtained directly from total DNA or from total RNA. The cDNA clones are not naturally occurring, but rather are preferably obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The construction of a cDNA library from mRNA involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection of the cells carrying the cDNA

30 library. Thus, the process which includes the construction of a cDNA library from mRNA and isolation of distinct cDNA clones yields an approximately 10^6 -fold purification of the native

message. Thus, purification of at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated. Thus the term "substantially pure" refers to a preparation comprising at least 50-60% by weight the compound of interest (e.g., nucleic acid, oligonucleotide, etc.). More preferably, the preparation comprises
5 at least 75% by weight, and most preferably 90-99% by weight, the compound of interest. Purity is measured by methods appropriate for the compound of interest.

The term "complementary" describes two nucleotides that can form multiple favorable interactions with one another. For example, adenine is complementary to thymine as they can form two hydrogen bonds. Similarly, guanine and cytosine are complementary since they can
10 form three hydrogen bonds. Thus if a nucleic acid sequence contains the following sequence of bases, thymine, adenine, guanine and cytosine, a "complement" of this nucleic acid molecule would be a molecule containing adenine in the place of thymine, thymine in the place of adenine, cytosine in the place of guanine, and guanine in the place of cytosine. Because the complement can contain a nucleic acid sequence that forms optimal interactions with the parent nucleic acid
15 molecule, such a complement can bind with high affinity to its parent molecule.

With respect to single stranded nucleic acids, particularly oligonucleotides, the term "specifically hybridizing" refers to the association between two single-stranded nucleotide molecules of sufficiently complementary sequence to permit such hybridization under pre-determined conditions generally used in the art (sometimes termed "substantially
20 complementary"). In particular, the term refers to hybridization of an oligonucleotide with a substantially complementary sequence contained within a single-stranded DNA or RNA molecule of the invention, to the substantial exclusion of hybridization of the oligonucleotide with single-stranded nucleic acids of non-complementary sequence. For example, specific hybridization can refer to a sequence which hybridizes to the T2D specific marker gene or
25 nucleic acid, but does not hybridize to other nucleotides. Appropriate conditions enabling specific hybridization of single stranded nucleic acid molecules of varying complementarity are well known in the art.

For instance, one common formula for calculating the stringency conditions required to achieve hybridization between nucleic acid molecules of a specified sequence homology is set
30 forth below (Sambrook et al., Molecular Cloning, Cold Spring Harbor Laboratory (1989):
$$T_m = 81.5^{\circ}\text{C} + 16.6 \text{ Log } [\text{Na}^+] + 0.41(\% \text{ G+C}) - 0.63 (\% \text{ formamide}) - 600/\#\text{bp in duplex}$$

As an illustration of the above formula, using $[Na^+] = [0.368]$ and 50% formamide, with GC content of 42% and an average probe size of 200 bases, the T_m is 57°C. The T_m of a DNA duplex decreases by 1 - 1.5°C with every 1% decrease in homology. Thus, targets with greater than about 75% sequence identity would be observed using a hybridization temperature of 42°C.

5 The stringency of the hybridization and wash depend primarily on the salt concentration and temperature of the solutions. In general, to maximize the rate of annealing of the probe with its target, the hybridization is usually carried out at salt and temperature conditions that are 20-25°C below the calculated T_m of the hybrid. Wash conditions should be as stringent as possible for the degree of identity of the probe for the target. In general, wash conditions are selected to
10 be approximately 12-20°C below the T_m of the hybrid. In regards to the nucleic acids of the current invention, a moderate stringency hybridization is defined as hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 µg/ml denatured salmon sperm DNA at 42°C, and washed in 2X SSC and 0.5% SDS at 55°C for 15 minutes. A high stringency hybridization is defined as hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 µg/ml
15 denatured salmon sperm DNA at 42°C, and washed in 1X SSC and 0.5% SDS at 65°C for 15 minutes. A very high stringency hybridization is defined as hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 µg/ml denatured salmon sperm DNA at 42°C, and washed in 0.1X SSC and 0.5% SDS at 65°C for 15 minutes.

 The term "oligonucleotide," as used herein is defined as a nucleic acid molecule
20 comprised of two or more ribo- or deoxyribonucleotides, preferably more than three. The exact size of the oligonucleotide will depend on various factors and on the particular application and use of the oligonucleotide. Oligonucleotides, which include probes and primers, can be any length from 3 nucleotides to the full length of the nucleic acid molecule, and explicitly include every possible number of contiguous nucleic acids from 3 through the full length of the
25 polynucleotide. Preferably, oligonucleotides are at least about 10 nucleotides in length, more preferably at least 15 nucleotides in length, more preferably at least about 20 nucleotides in length.

 The term "probe" as used herein refers to an oligonucleotide, polynucleotide or nucleic acid, either RNA or DNA, whether occurring naturally as in a purified restriction enzyme digest
30 or produced synthetically, which is capable of annealing with or specifically hybridizing to a nucleic acid with sequences complementary to the probe. A probe may be either single-stranded

or double-stranded. The exact length of the probe will depend upon many factors, including temperature, source of probe and use of the method. For example, for diagnostic applications, depending on the complexity of the target sequence, the oligonucleotide probe typically contains 15-25 or more nucleotides, although it may contain fewer nucleotides. The probes herein are
5 selected to be complementary to different strands of a particular target nucleic acid sequence. This means that the probes must be sufficiently complementary so as to be able to "specifically hybridize" or anneal with their respective target strands under a set of pre-determined conditions. Therefore, the probe sequence need not reflect the exact complementary sequence of the target. For example, a non-complementary nucleotide fragment may be attached to the 5' or 3' end of the
10 probe, with the remainder of the probe sequence being complementary to the target strand. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the sequence of the target nucleic acid to anneal therewith specifically.

The term "primer" as used herein refers to an oligonucleotide, either RNA or DNA, either
15 single-stranded or double-stranded, either derived from a biological system, generated by restriction enzyme digestion, or produced synthetically which, when placed in the proper environment, is able to functionally act as an initiator of template-dependent nucleic acid synthesis. When presented with an appropriate nucleic acid template, suitable nucleoside triphosphate precursors of nucleic acids, a polymerase enzyme, suitable cofactors and conditions
20 such as a suitable temperature and pH, the primer may be extended at its 3' terminus by the addition of nucleotides by the action of a polymerase or similar activity to yield a primer extension product. The primer may vary in length depending on the particular conditions and requirement of the application. For example, in diagnostic applications, the oligonucleotide primer is typically 15-25 or more nucleotides in length. The primer must be of sufficient
25 complementarity to the desired template to prime the synthesis of the desired extension product, that is, to be able anneal with the desired template strand in a manner sufficient to provide the 3' hydroxyl moiety of the primer in appropriate juxtaposition for use in the initiation of synthesis by a polymerase or similar enzyme. It is not required that the primer sequence represent an exact complement of the desired template. For example, a non-complementary nucleotide sequence
30 may be attached to the 5' end of an otherwise complementary primer. Alternatively, non-complementary bases may be interspersed within the oligonucleotide primer sequence,

provided that the primer sequence has sufficient complementarity with the sequence of the desired template strand to functionally provide a template-primer complex for the synthesis of the extension product.

Polymerase chain reaction (PCR) has been described in US Patents 4,683,195, 4,800,195,
5 and 4,965,188, the entire disclosures of which are incorporated by reference herein.

The term "vector" relates to a single or double stranded circular nucleic acid molecule that can be infected, transfected or transformed into cells and replicate independently or within the host cell genome. A circular double stranded nucleic acid molecule can be cut and thereby linearized upon treatment with restriction enzymes. An assortment of vectors, restriction
10 enzymes, and the knowledge of the nucleotide sequences that are targeted by restriction enzymes are readily available to those skilled in the art, and include any replicon, such as a plasmid, cosmid, bacmid, phage or virus, to which another genetic sequence or element (either DNA or RNA) may be attached so as to bring about the replication of the attached sequence or element.

The term "promoter element" describes a nucleotide sequence that is incorporated into a
15 vector that, once inside an appropriate cell, can facilitate transcription factor and/or polymerase binding and subsequent transcription of portions of the vector DNA into mRNA. In one embodiment, the promoter element of the present invention precedes the 5' end of the T2D specific marker nucleic acid molecule such that the latter is transcribed into mRNA. Host cell machinery then translates mRNA into a polypeptide.

As used herein, the terms "regulatory element" and "regulatory polynucleotide" refer to
20 polynucleotide molecules having regulatory activity (i.e., one that has the ability to affect the transcription of an operably linked transcribable polynucleotide molecule). The terms refer to a polynucleotide molecule containing one or more elements such as core promoter regions, cis-elements, leaders or UTRs, enhancers, introns, and transcription termination regions, all of which
25 have regulatory activity and may play a role in the overall expression of nucleic acid molecules in living cells. The "regulatory elements" determine if, when, and at what level a particular polynucleotide is transcribed. The regulatory elements may interact with regulatory proteins or other proteins or be involved in nucleotide interactions, for example, to provide proper folding of a regulatory polynucleotide.

As used herein, the term "cis-element" refers to a cis-acting transcriptional regulatory
30 element that confers an aspect of the overall control of the expression of an operably linked

transcribable polynucleotide. A cis-element may function to bind transcription factors, which are trans-acting protein factors that regulate transcription. Some cis-elements bind more than one transcription factor, and transcription factors may interact with different affinities with more than one cis-element. Cis-elements can confer or modulate expression, and can be identified by a
5 number of techniques, including deletion analysis (i.e., deleting one or more nucleotides from the 5' end or internal to a promoter), DNA binding protein analysis using DNase I footprinting, methylation interference, electrophoresis mobility-shift assays, in vivo genomic footprinting by ligation-mediated PCR, and other conventional assays; or by DNA sequence similarity analysis with known cis-element motifs by conventional DNA sequence comparison methods. The fine
10 structure of a cis-element can be further studied by mutagenesis (or substitution) of one or more nucleotides or by other conventional methods. Cis-elements can be obtained by chemical synthesis or by isolation from regulatory polynucleotides that include such elements, and they can be synthesized with additional flanking nucleotides that contain useful restriction enzyme sites to facilitate subsequence manipulation.

15 As used herein, the term "enhancer" refers to a transcriptional regulatory element, typically 100-200 base pairs in length, which strongly activates transcription, for example, through the binding of one or more transcription factors. Enhancers can be identified and studied by methods such as those described above for cis-elements. Enhancer sequences can be obtained by chemical synthesis or by isolation from regulatory elements that include such elements, and
20 they can be synthesized with additional flanking nucleotides that contain useful restriction enzyme sites to facilitate subsequence manipulation.

As used herein, the term "intron" refers to a polynucleotide molecule that may be isolated or identified from the intervening sequence of a genomic copy of a transcribed polynucleotide which is spliced out during mRNA processing prior to translation. Introns may themselves
25 contain sub-elements such as cis-elements or enhancer domains that affect the transcription of operably linked polynucleotide molecules. Some introns are capable of increasing gene expression through a mechanism known as intron mediated enhancement (IME). IME, as distinguished from the effects of enhancers, is based on introns residing in the transcribed region of a polynucleotide. In general, IME is mediated by the first intron of a gene, which can reside in
30 either the 5'-UTR sequence of a gene or between the first and second protein coding (CDS) exons of a gene. Without being limited by theory, because IME may be particularly important in

highly expressed, constitutive genes, it may also play a role in the expression of genes expressed in a tissue-specific manner. Those skilled in the art will recognize that a nucleic acid vector can contain nucleic acid elements other than the promoter element and the T2D specific marker gene nucleic acid molecule. These other nucleic acid elements include, but are not limited to, origins
5 of replication, ribosomal binding sites, nucleic acid sequences encoding drug resistance enzymes or amino acid metabolic enzymes, and nucleic acid sequences encoding secretion signals, localization signals, or signals useful for polypeptide purification.

A "replicon" is any genetic element, for example, a plasmid, cosmid, bacmid, plastid, phage or virus, that is capable of replication largely under its own control. A replicon may be
10 either RNA or DNA and may be single or double stranded.

An "expression operon" refers to a nucleic acid segment that may possess transcriptional and translational control sequences, such as promoters, enhancers, translational start signals (e.g., ATG or AUG codons), polyadenylation signals, terminators, and the like, and which facilitate the expression of a polypeptide coding sequence in a host cell or organism.

As used herein, the terms "reporter," "reporter system", "reporter gene," or "reporter gene product" shall mean an operative genetic system in which a nucleic acid comprises a gene that encodes a product that when expressed produces a reporter signal that is a readily measurable, e.g., by biological assay, immunoassay, radio immunoassay, or by colorimetric, fluorogenic, chemiluminescent or other methods. The nucleic acid may be either RNA or DNA, linear or
20 circular, single or double stranded, antisense or sense polarity, and is operatively linked to the necessary control elements for the expression of the reporter gene product. The required control elements will vary according to the nature of the reporter system and whether the reporter gene is in the form of DNA or RNA, but may include, but not be limited to, such elements as promoters, enhancers, translational control sequences, poly A addition signals, transcriptional termination
25 signals and the like.

The introduced nucleic acid may or may not be integrated (covalently linked) into nucleic acid of the recipient cell or organism. In bacterial, yeast, plant and mammalian cells, for example, the introduced nucleic acid may be maintained as an episomal element or independent replicon such as a plasmid. Alternatively, the introduced nucleic acid may become integrated
30 into the nucleic acid of the recipient cell or organism and be stably maintained in that cell or organism and further passed on or inherited to progeny cells or organisms of the recipient cell or

organism. Finally, the introduced nucleic acid may exist in the recipient cell or host organism only transiently.

The term "selectable marker gene" refers to a gene that when expressed confers a selectable phenotype, such as antibiotic resistance, on a transformed cell.

5 The term "operably linked" means that the regulatory sequences necessary for expression of the coding sequence are placed in the DNA molecule in the appropriate positions relative to the coding sequence so as to effect expression of the coding sequence. This same definition is sometimes applied to the arrangement of transcription units and other transcription control elements (e.g. enhancers) in an expression vector.

10 The terms "recombinant organism," or "transgenic organism" refer to organisms which have a new combination of genes or nucleic acid molecules. A new combination of genes or nucleic acid molecules can be introduced into an organism using a wide array of nucleic acid manipulation techniques available to those skilled in the art. The term "organism" relates to any living being comprised of a least one cell. An organism can be as simple as one eukaryotic cell
15 or as complex as a mammal. Therefore, the phrase "a recombinant organism" encompasses a recombinant cell, as well as eukaryotic and prokaryotic organism.

 The term "isolated protein" or "isolated and purified protein" is sometimes used herein. This term refers primarily to a protein produced by expression of an isolated nucleic acid molecule of the invention. Alternatively, this term may refer to a protein that has been
20 sufficiently separated from other proteins with which it would naturally be associated, so as to exist in "substantially pure" form. "Isolated" is not meant to exclude artificial or synthetic mixtures with other compounds or materials, or the presence of impurities that do not interfere with the fundamental activity, and that may be present, for example, due to incomplete purification, addition of stabilizers, or compounding into, for example, immunogenic
25 preparations or pharmaceutically acceptable preparations.

 A "specific binding pair" comprises a specific binding member (sbm) and a binding partner (bp) which have a particular specificity for each other and which in normal conditions bind to each other in preference to other molecules. Examples of specific binding pairs are antigens and antibodies, ligands and receptors and complementary nucleotide sequences. The
30 skilled person is aware of many other examples. Further, the term "specific binding pair" is also applicable where either or both of the specific binding member and the binding partner comprise

a part of a large molecule. In embodiments in which the specific binding pair comprises nucleic acid sequences, they will be of a length to hybridize to each other under conditions of the assay, preferably greater than 10 nucleotides long, more preferably greater than 15 or 20 nucleotides long.

5 “Sample” or “patient sample” or “biological sample” generally refers to a sample which may be tested for a particular molecule, preferably a T2D specific marker molecule. Samples may include but are not limited to cells, body fluids, including blood, serum, plasma, urine, saliva, tears, pleural fluid and the like.

10 **Methods of using Gene Targets which interact with the**
Variant embedded within the *TCF7L2* gene to Screen Agents Useful for the Treatment of
T2D

15 Methods for identifying agents that modulate the enhancer activity of the SNP embedded within the *TCF7L2* gene containing nucleic acids its interacting promoters will result in the generation of efficacious therapeutic agents for the treatment of a variety of disorders associated with this condition.

20 These DNA-DNA binding complexes provide suitable targets for the rational design of therapeutic agents which modulate the activity of the gene targets identified herein, thereby interfering with the T2D phenotype. Small nucleic acid molecules or peptides corresponding to these regions may be used to advantage in the design of therapeutic agents which effectively modulate the activity of the encoded proteins.

25 Molecular modeling should facilitate the identification of specific organic molecules with capacity to bind to the active site of the proteins which bind the SNP containing *TCF7L2* nucleic acids based on conformation or key amino acid residues required for function. A combinatorial chemistry approach will be used to identify molecules with greatest activity and then iterations of these molecules will be developed for further cycles of screening.

30 The agents employed in drug screening assays may either be free in solution, affixed to a solid support or within a cell. One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant polynucleotides expressing the polypeptide or fragment, preferably in competitive binding assays. Such cells, either in viable or

fixed form, can be used for standard binding assays. One may determine, for example, formation of complexes between the polypeptide or fragment and the agent being tested, or examine the degree to which the formation of a complex between the polypeptide or fragment and a known substrate is interfered with by the agent being tested.

5 Agents which can be tested for disruption of enhancer binding, without limitation, those currently being tested in clinical trials for other disorders, such as cancer. Exemplary agents include, Iniparib (BSI 201), Olaparib (AZD-2281), Rucaparib (AG014699, PF-01367338), Veliparib (ABT-888), CEP 9722, MK 4827, Inhibitor of PARP1 and PARP2. BMN-673 and 3-aminobenzamide, a prototypical PARP inhibitor. Other agents which target ACSL
10 activity such as Triacsin c, a specific inhibitor of acyl-CoA synthetase (ACS) can also be used to modulate the mediated effects of the genes listed above on the T2D phenotype. ACS catalyzes the formation of acyl-CoA from fatty acid, a reaction that is involved in both the degradation of fatty acid and the synthesis of cellular lipids. Additional agents for this purpose include, rosiglitazone, Wy14643, 15d-PGJ2 and AICAR. (37)

15 Another technique for drug screening provides high throughput screening for compounds having suitable binding affinity for the encoded polypeptides and is described in detail in Geysen, PCT published application WO 84/03564, published on September 13, 1984. Briefly stated, large numbers of different, small peptide test compounds, such as those described above, are synthesized on a solid substrate, such as plastic pins or some other surface. The peptide test
20 compounds are reacted with the target polypeptide and washed. Bound polypeptide is then detected by methods well known in the art.

A further technique for drug screening involves the use of host eukaryotic cell lines or cells (such as described above) which have the *TCF7L2* containing SNP allele. The host cell lines or cells are grown in the presence of drug compound. The rate of cellular metabolism of the
25 host cells or expression levels of the target genes disclosed herein is measured to determine if the compound is capable of regulating cellular metabolism, gene expression, or other cellular parameters associated with the diabetic phenotype. A variety of cell lines are commercially available for use in such screening assays. Methods for introducing DNA molecules are also well known to those of ordinary skill in the art. Such methods are set forth in Ausubel et al. eds.,
30 Current Protocols in Molecular Biology, John Wiley & Sons, NY, N.Y. 1995, the disclosure of which is incorporated by reference herein.

Cells and cell lines suitable for studying the effects of disrupting the interaction between TCF7L2 containing SNP nucleic acids and different promoter elements on glucose metabolism and methods of use thereof for drug discovery are provided. Such cells and cell lines will be transfected with the SNP encoding nucleic acids described herein or variants thereof and the effects on glucagon secretion, insulin secretion and/or beta cell apoptosis can be determined. Such cells and cell lines will also be contacted with the siRNA molecules provided herein to assess the effects thereof on glucagon secretion, insulin secretion and/or beta cell apoptosis. The siRNA molecules will be tested alone and in combination of 2, 3, 4, and 5 siRNAs to identify the most efficacious combinations. Cells suitable for these purposes include, without limitation, INS cells (ATCC CRL 11605), PC12 cells (ATCC CRL 1721), MIN6 cells, alpha-TC6 cells, H116 cells, NCM460 cells, and INS-1 832/13 cells (Fernandez et al., J. of Proteome Res. (2007). 7:400-411). Pancreatic islet cells can be isolated and cultured as described in Joseph, J. et al., (J. Biol. Chem. (2004) 279:51049). Diao et al. (J. Biol. Chem. (2005) 280:33487-33496), provide methodology for assessing the effects of the SNP encoding nucleic acids and/or the siRNAs provided herein on glucagon secretion and insulin secretion. Park, J. et al. (J. of Bioch. and Mol. Biol. (2007) 40:1058-68) provide methodology for assessing the effect of these nucleic acid molecules on glucosamine induced beta cell apoptosis in pancreatic islet cells.

A wide variety of expression vectors are available that can be modified to express the sequences of this invention. The specific vectors exemplified herein are merely illustrative, and are not intended to limit the scope of the invention. Expression methods are described by Sambrook et al. Molecular Cloning: A Laboratory Manual or Current Protocols in Molecular Biology 16.3-17.44 (1989). Expression methods in *Saccharomyces* are also described in Current Protocols in Molecular Biology (1989).

Suitable vectors for use in practicing the invention include prokaryotic vectors such as the pNH vectors (Stratagene Inc., 11099 N. Torrey Pines Rd., La Jolla, Calif. 92037), pET vectors (Novogen Inc., 565 Science Dr., Madison, Wis. 53711) and the pGEX vectors (Pharmacia LKB Biotechnology Inc., Piscataway, N.J. 08854). Examples of eukaryotic vectors useful in practicing the present invention include the vectors pRc/CMV, pRc/RSV, and pREP (Invitrogen, 11588 Sorrento Valley Rd., San Diego, Calif. 92121); pcDNA3.1/V5&His (Invitrogen); baculovirus vectors such as pVL1392, pVL1393, or pAC360 (Invitrogen); and yeast vectors such as YRP17, YIP5, and YEP24 (New England Biolabs, Beverly, Mass.), as well

as pRS403 and pRS413 Stratagene Inc.); *Pichia* vectors such as pHIL-D1 (Phillips Petroleum Co., Bartlesville, Okla. 74004); retroviral vectors such as PLNCX and pLPCX (Clontech); and adenoviral and adeno-associated viral vectors.

Promoters for use in expression vectors of this invention include promoters that are operable in prokaryotic or eukaryotic cells. Promoters that are operable in prokaryotic cells include lactose (lac) control elements, bacteriophage lambda (pL) control elements, arabinose control elements, tryptophan (trp) control elements, bacteriophage T7 control elements, and hybrids thereof. Promoters that are operable in eukaryotic cells include Epstein Barr virus promoters, adenovirus promoters, SV40 promoters, Rous Sarcoma Virus promoters, cytomegalovirus (CMV) promoters, baculovirus promoters such as AcMNPV polyhedrin promoter, *Pichia* promoters such as the alcohol oxidase promoter, and *Saccharomyces* promoters such as the gal4 inducible promoter and the PGK constitutive promoter, as well as neuronal-specific platelet-derived growth factor promoter (PDGF), and the Thy-1 promoter.

In addition, a vector of this invention may contain any one of a number of various markers facilitating the selection of a transformed host cell. Such markers include genes associated with temperature sensitivity, drug resistance, or enzymes associated with phenotypic characteristics of the host organisms.

Host cells expressing the T2D-associated SNP and the promoters that interact with it or functional fragments thereof provide a system in which to screen potential compounds or agents for the ability to modulate the development of T2D. Thus, in one embodiment, the nucleic acid molecules of the invention may be used to create recombinant cell lines for use in assays to identify agents which modulate aspects of ACSL5 expression as well as agents which impact the diabetic phenotype. Also provided herein are methods to screen for compounds capable of modulating the function of proteins which bind the nucleic acids described below. Assays for measurement of fatty acid oxidation rates in animal tissues and cell lines are described by Huynh et al. *Methods Enzymol.* (2014) 542:491-405 and by Bennet et al. in *Methods in Cell Biology*, 80:179-197 (2007). Fatty acid oxidation may also be studied using Seahorse Bioscience XF technology from Agilent. TeSlaa et al. provide methods for analyzing glycolysis in cells in *Methods Enzymol.* (2014) 542:91-114. Promega provides the Glucose Uptake-Glo® Assay which is a plate based assay for measuring glucose uptake in cells. Each of the aforementioned

references are hereby incorporated by reference in their entirety for the assays and cells they disclose.

The goal of rational drug design is to produce structural analogs of biologically active polypeptides of interest or of small molecules with which they interact (e.g., agonists, antagonists, inhibitors) in order to fashion drugs which are, for example, more active or stable forms of the polypeptide, or which, e.g., enhance or interfere with the function of a polypeptide in vivo. See, e.g., Hodgson, (1991) *Bio/Technology* 9:19-21. In one approach, discussed above, the three-dimensional structure of a protein of interest or, for example, of the protein-substrate complex, is solved by x-ray crystallography, by nuclear magnetic resonance, by computer modeling or most typically, by a combination of approaches. Less often, useful information regarding the structure of a polypeptide may be gained by modeling based on the structure of homologous proteins. An example of rational drug design is the development of HIV protease inhibitors (Erickson et al., (1990) *Science* 249:527-533). In addition, peptides may be analyzed by an alanine scan (Wells, (1991) *Meth. Enzym.* 202:390-411). In this technique, an amino acid residue is replaced by Ala, and its effect on the peptide's activity is determined. Each of the amino acid residues of the peptide is analyzed in this manner to determine the important regions of the peptide.

It is also possible to isolate a target-specific antibody, selected by a functional assay, and then to solve its crystal structure. In principle, this approach yields a pharmacophore upon which subsequent drug design can be based.

One can bypass protein crystallography altogether by generating anti-idiotypic antibodies (anti-ids) to a functional, pharmacologically active antibody. As a mirror image of a mirror image, the binding site of the anti-ids would be expected to be an analog of the original molecule. The anti-id could then be used to identify and isolate peptides from banks of chemically or biologically produced banks of peptides. Selected peptides would then act as the pharmacophore.

Pharmaceuticals and Small Molecule Therapies

The elucidation of the role played by the T2D associated SNP containing nucleic acids described herein in target gene regulation facilitates the development of pharmaceutical compositions useful for treatment and diagnosis of T2D. These compositions may comprise, in

addition to one of the above substances, a pharmaceutically acceptable excipient, carrier, buffer, stabilizer or other materials well known to those skilled in the art. Such materials should be non-toxic and should not interfere with the efficacy of the active ingredient. Whether it is a polypeptide, antibody, peptide, nucleic acid molecule, small molecule or other pharmaceutically useful compound according to the present invention that is to be given to an individual, administration is preferably in a “prophylactically effective amount” or a “therapeutically effective amount” (as the case may be, although prophylaxis may be considered therapy), this being sufficient to show benefit to the individual.

As it is presently understood, RNA interference involves a multi-step process. Double stranded RNAs are cleaved by the endonuclease Dicer to generate nucleotide fragments (siRNA). The siRNA duplex is resolved into 2 single stranded RNAs, one strand being incorporated into a protein-containing complex where it functions as guide RNA to direct cleavage of the target RNA (Schwarz et al, Mol. Cell. 10:537 548 (2002), Zamore et al, Cell 101:25 33 (2000)), thus silencing a specific genetic message (see also Zeng et al, Proc. Natl. Acad. Sci. 100:9779 (2003)).

The invention includes a method of treating T2D in a mammal. An exemplary method entails administering to the mammal a pharmaceutically effective amount of an siRNA molecule directed to gene target identified herein. Such molecules are commercially available from Dharmacon. The siRNA inhibits the expression of the aforementioned gene. Preferably, the mammal is a human. The term “patient” as used herein refers to a human. The invention also entails administration of small molecules or peptides which block binding between the DNA sequences described herein.

Specific formulations directed at inhibiting the expression of these target genes as well as delivery methods are provided as a novel therapy to treat T2D. The siRNA or small molecule for example can be delivered to a patient in vivo either systemically or locally with carriers, as discussed below. The compositions of the invention may be used alone or in combination with other agents or genes encoding proteins to augment the efficacy of the compositions.

A “membrane permeant peptide sequence” refers to a peptide sequence which is able to facilitate penetration and entry of the siRNA inhibitor across the cell membrane. Exemplary peptides include with out limitation, the signal sequence from Kaposi fibroblast growth factor exemplified herein, the HIV tat peptide (Vives et al., J Biol. Chem., 272:16010-16017, 1997),

Nontoxic membrane translocation peptide from protamine (Park et al., FASEB J. 19(11):1555-7, 2005), CHARIOT® delivery reagent (Active Motif; US Patent 6,841,535) and the antimicrobial peptide Buforin 2.

In one embodiment of the invention siRNAs or small molecules are delivered for
5 therapeutic benefit. There are several ways to administer such compounds of the invention in vivo to treat T2D including, but not limited to, naked siRNA delivery, siRNA conjugation and delivery, liposome carrier-mediated delivery, polymer carrier delivery, nanoparticle compositions, plasmid-based methods, and the use of viruses.

siRNA or small molecule compositions of the invention can comprise a delivery vehicle,
10 including liposomes, for administration to a subject, carriers and diluents and their salts, and/or can be present in pharmaceutically acceptable formulations. This can be necessary to allow the siRNA to cross the cell membrane and escape degradation. Methods for the delivery of nucleic acid molecules are described in Akhtar et al., 1992, Trends Cell Bio., 2, 139; Delivery Strategies for Antisense Oligonucleotide Therapeutics, ed. Akhtar, 1995, Maurer et al., 1999, Mol. Membr.
15 Biol., 16, 129-140; Hofland and Huang, 1999, Handb. Exp. Pharmacol., 137, 165-192; and Lee et al., 2000, ACS Symp. Ser., 752, 184-192; Beigelman et al., U.S. Pat. No. 6,395,713 and Sullivan et al., PCT WO 94/02595 further describe the general methods for delivery of nucleic acid molecules. These protocols can be utilized for the delivery of virtually any nucleic acid molecule.

20 The frequency of administration of the agent to a patient will also vary depending on several factors including, but not limited to, the type and severity of the T2D to be treated, the route of administration, the age and overall health of the individual, the nature of the agent and the like. It is contemplated that the frequency of administration of the agent to the patient may vary from about once every few months to about once a month, to about once a week, to about
25 once per day, to about several times daily.

Pharmaceutical compositions that are useful in the methods of the invention may be administered systemically in parenteral, oral solid and liquid formulations, ophthalmic, suppository, aerosol, topical or other similar formulations. In addition to the appropriate agent or other efficacious small molecule, these pharmaceutical compositions may contain
30 pharmaceutically-acceptable carriers and other ingredients known to enhance and facilitate drug administration. Thus such compositions may optionally contain other components, such as

adjuvants, e.g., aqueous suspensions of aluminum and magnesium hydroxides, and/or other pharmaceutically acceptable carriers, such as saline. Other possible formulations, such as nanoparticles, liposomes, resealed erythrocytes, and immunologically based systems may also be used to administer the appropriate siRNA or small molecules to a patient according to the methods of the invention. The use of nanoparticles to deliver siRNAs, as well as cell membrane permeable peptide carriers that can be used are described in Crombez et al., *Biochemical Society Transactions* v35:p44 (2007).

A pharmaceutically acceptable carrier can contain physiologically acceptable compounds that act, for example, to stabilize the agent, e.g., a siRNA or small molecule, or increase the absorption of the agent. Such physiologically acceptable compounds include, for example, carbohydrates, such as glucose, sucrose or dextrans, antioxidants, such as ascorbic acid or glutathione, chelating agents, low molecular weight proteins or other stabilizers or excipients. One skilled in the art would know that the choice of a pharmaceutically acceptable carrier, including a physiologically acceptable compound, depends, for example, on the route of administration of the agent.

Nucleic acid molecules can be administered to cells by incorporation into other vehicles, such as biodegradable polymers, hydrogels, cyclodextrins. (see for example Gonzalez et al., 1999, *Bioconjugate Chem.*, 10, 1068-1074; Wang et al., International PCT publication Nos. WO 03/47518 and WO 03/46185), poly(lactic-co-glycolic)acid (PLGA) and PLCA microspheres (see for example U.S. Pat. No. 6,447,796 and US Patent Application Publication No. US 2002130430), biodegradable nanocapsules, and bioadhesive microspheres, or by proteinaceous vectors (O'Hare and Normand, International PCT Publication No. WO 00/53722)

In order to treat an individual having T2D, to alleviate a sign or symptom of the disease, the agent should be administered in an effective dose. The total treatment dose can be administered to a subject as a single dose or can be administered using a fractionated treatment protocol, in which multiple doses are administered over a more prolonged period of time, for example, over the period of a day to allow administration of a daily dosage or over a longer period of time to administer a dose over a desired period of time. One skilled in the art would know that the amount of agent required to obtain an effective dose in a subject depends on many factors, including the age, weight and general health of the subject, as well as the route of administration and the number of treatments to be administered. In view of these factors, the

skilled artisan would adjust the particular dose so as to obtain an effective dose for treating an individual having T2D.

The effective dose of agent will depend on the mode of administration, and the weight of the individual being treated. The dosages described herein are generally those for an average adult but can be adjusted for the treatment of children. The dose will generally range from about 0.001 mg to about 1000 mg.

In an individual suffering from T2D, in particular a more severe form of the disease, administration of the agent can be particularly useful when administered in combination, for example, with a conventional agent for treating such a disease. The skilled artisan would administer agent alone or in combination and would monitor the effectiveness of such treatment using routine methods such as pancreatic beta cell function determination, radiologic, immunologic or, where indicated, histopathologic methods. Other conventional agents for the treatment of diabetes include insulin administration, glucagon administration or agents that alter levels of either of these two molecules. Glucophage®, Avandia®, Actos®, Januvia® and Glucovance® are examples of such agents.

Administration of the pharmaceutical preparation is preferably in an “effective amount” this being sufficient to show benefit to the individual. This amount prevents, alleviates, abates, or otherwise reduces the severity of T2D symptoms in a patient.

The pharmaceutical preparation is formulated in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form, as used herein, refers to a physically discrete unit of the pharmaceutical preparation appropriate for the patient undergoing treatment. Each dosage should contain a quantity of active ingredient calculated to produce the desired effect in association with the selected pharmaceutical carrier. Procedures for determining the appropriate dosage unit are well known to those skilled in the art.

Dosage units may be proportionately increased or decreased based on the weight of the patient. Appropriate concentrations for alleviation of a particular pathological condition may be determined by dosage concentration curve calculations, as known in the art.

Kits and Articles of Manufacture

Any of the aforementioned products can be incorporated into a kit which may contain an T2D-associated SNP specific marker TCF7L2 (or any of the genes listed above) polynucleotide

or one or more such markers immobilized on a Gene Chip, an oligonucleotide, a polypeptide which binds the SNP containing nucleic acid described herein, a peptide designed to disrupt such binding, an siRNA, a small molecule, an antibody, a label, marker, or reporter, a pharmaceutically acceptable carrier, a physiologically acceptable carrier, instructions for use, a container, a vessel for administration, an assay substrate, or any combination thereof.

The materials and methods set forth below are provided to facilitate the practice of the present invention.

Cell Culture

The NCM460 cells were purchased from INCELL Corporation, LLC and maintained in M3 medium according the manufacture's protocol. The human colorectal cancer cell line, HCT116 (ATCC), was maintained at 37°C with 5% CO₂ incubation in Dulbecco's modified Eagle's medium (DMEM), supplemented with 10% fetal bovine serum (HyClone), 100 U/ml penicillin and 100 mg/ml streptomycin.

Targeting the genomic region encompassing rs7903146 in HCT116 cells – 1.4kb deletion

A bicistronic vector px330, expressing both a human codon-optimized SpCas9 and a chimeric guide RNA, was purchased from Addgene (Plasmid 42230)[1]. The plasmid was linearized with *BbsI*, dephosphorylated and then gel purified. We designed the 10 guide RNA according to the different PAM sites to generate the constructs. These oligonucleotides for targeting the *TCF7L2* rs7903146 region were annealed and ligated to the linearized vector. The guide sequences are listed in **Figure 1A**. The primers used for molecular cloning are listed in **Table 1**.

The targeting sites yielding high scores (i.e. good cutting efficiency with low chance of off-target effects), as predicted by the CRISPR design tool developed at the Broad institute, are highlighted in **Figure 1A**, along with the corresponding guide sequences. We tested all of 10 guide sequences and #5 and #7 showed the best cutting efficiency. The guide RNA #5 in the upstream of rs7903146 and the guide RNA #7 in the downstream were predicted to generate a 1.4kb deletion (**Figure 1A**).

To determine whether the CRISPR/Cas9 system could generate targeted cleavage in the designed region across the SNP, we transfected plasmids expressing both the mammalian-codon optimized Cas9 and sgRNA into HCT116 cells and then isolated the single clones followed by

PCR-based genotyping. Using one pair of flanking primers (P1 and P2) the expected size for the PCR product was 1.8kb for the uncut version. The expected size for the PCR products was 0.4kb for the 1.4kb cleaved version with primers P1 and P2. As shown in **Figure 1B**, we successfully identified the homozygous deletion, namely '1.4kb del' (1.4kb deletion). The sequence of the 0.4kb fragment was validated by Sanger sequencing and then aligned with the wild-type 1.8kb fragment (**Figures 1C and 1D**).

Table I

10 Primers: 4C and Capture C probe experiments

Sequence Name	Sequence
4C 1.4kb del upstream	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTTCAGG
TCF7L2 SNP.2F	CTTCACCAGCAC
4C 1.4kb del upstream	ACACTCTTTCCTACACGACGCTCTTCCGATCTTCCACTGA
TCF7L2 SNP.2R	GGGAAGTGAAAG
4C TCF7L2 SNP.1F	ACACTCTTTCCTACACGACGCTCTTCCGATCTTCCTTTTAA AATGGTGACAAATTC
4C TCF7L2 SNP.1R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTTCTCTTT GGTCCGGCAGA
TCF7L2_rs7903146	GATCTCTGCCGGACCAAAGAGAAGATTCTTTTAAATGG
Capture C probe 1	TGACAAATTCATGGGCTTTCTCTG CCTCAAAACCTAGCACAGCTGTTATTTACTGAACAATTAGA GAGCTAAGCACTTTTTAGATACTATATAATTTAAT
TCF7L2_rs7903146	CTCTTCGCTATAAACATTTTAGCTTTTGTGTTTGCTGACTG
Capture C probe 2	GCAACAATACATAGTGAAAGTTCTAATAATTTGTAATGCTT TTGCATGTCTTTGTATTTTCTTGGTTATCACATCACATCAA ATTAAGATACTGATC
4C SNP TCF7L2.2F	ACACTCTTTCCTACACGACGCTCTTCCGATCTTCCTTCAACCTCCACTGAGG
4C SNP TCF7L2.2R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTCATTGGCCTTCAGGCTTCAC

Primers: CRISPR/Cas 9-px330 cloning

Sequence Name	Sequence
crispr1F	CACCGAACAGCTGTGCTAGGTTTG
crispr1R	AAACCAAAACCTAGCACAGCTGTTT

Crispr2F	CACCGCAGTAAATAACAGCTGTGCT
Crispr2R	AAACAGCACAGCTGTTATTTACTGC
Crispr3F	CACCGAAAATAAGGGTGCCTCATA
Crispr3R	AAACTATGAGGCACCCTTAGTTTTT
Crispr4F	CACCGTTTCTCGTCTGAAAATAA
Crispr4R	AAACTTAGTTTTTCAGACGAGAAAC
Crispr5F	CACCGGTTTCTCGTCTGAAAATA
Crispr5R	AAACTAGTTTTTCAGACGAGAAACC
Crispr6F	CACCGACTTGCCTTCCCTGTAAC TG
Crispr6R	AAACCAGTTACAGGGAAGGCAAGTC
Crispr7F	CACCGTATAATTTAATTGCCGTATG
Crispr7R	AAACCATACGGCAATTAAATTATAC
Crispr8F	CACCGAGACGAGAAACCACAGTTAC
Crispr8R	AAACGTAACTGTGGTTTCTCGTCTC
Crispr9F	CACCGACGAGAAACCACAGTTACA
Crispr9R	AAACTGTAACTGTGGTTTCTCGTC
Crispr10F	CACCGAGAAACCACAGTTACAGGGA
Crispr10R	AAACTCCCTGTAACTGTGGTTTCTC
Crispr11F	CACCGACAGATACGTCAGGTGTCAC
crispr11R	AAACGTGACACCTGACGTATCTGTC
Crispr12F	CACCGGATTACAATGCAAGGTTGC
Crispr12R	AAACGCAACCTTGCATTGTAATCC
Crispr13F	CACCGTTAATCATCTTTTGCACACG
Crispr13R	AAACCGTGTGCAAAAGATGATTAAC
Crispr14F	CACCGAAGAAATTTACAGATACGTC
Crispr14R	AAACGACGTATCTGTAAATTTCTTC
Crispr15F	CACCGTGTCACTGGATTACAATGCA
Crispr15R	AAACTGCATTGTAATCCAGTGACAC
Crispr16F	CACCGAGTCGTCTATTACACAGGCC
Crispr16R	AAACGGCCTGTGTAATAGACGACTC
Crispr17F	CACCGTGTAATAGACGACTTCACA
Crispr17R	AAACTGTGAAGTCGTCTATTACAC
Crispr18F	CACCGTGTGAAGTCGTCTATTACAC
Crispr18R	AAACGTGTAATAGACGACTTCACAC
Crispr19F	CACCGTGGCTAAGTTTCCATGACC
Crispr19R	AAACGGTCATGGAACTTAGCCAC
Crispr20F	CACCGTCCATGAGATACTTTGCTGC

Crispr20R AAACGCAGCAAAGTATCTCATGGAC

Primers: Genotyping

Sequence Name	Sequence
P1	CCCGGTACCCCTGGTCTCATCAGCGATATTC
P2	CCCCTCGAGGGCACCAAGCTAAGACAAGATA
P3	TCTCTGCCGGACCAAAGAGAAGATTCC
P4	TGTGTCCAGGGCCCCTCTAACCTTTTC

Primers: Expression analysis of the key genes

Sequence Name	Sequence
TCF7L2 for	TCCATTTTCAGTCCGGCAG
TCF7L2 rev	CATCCTTGAGGGCTTGTCTAC
ACSL5 for	AGGCATGATAGTTTCTGGGAC
ACSL5 rev	ACCTGACATCCCATTGCTG

Subsequent additional deletions

In order to generate 104bp and 66bp deletions, we generated sgRNA expressing vectors in a similar fashion as described above. These additional oligonucleotides for targeting the *TCF7L2* rs7903146 region were annealed and ligated to the linearized vector. Leveraging the 10 guide sequences (**Figure 1E**), the guide RNA#12 in the upstream of rs7903146 and the guide RNA #17 or #18 in the downstream generated 66bp and 104bp deletions, respectively. Using one pairs of flanking primers (P3 and P4), the expected size for the PCR product was 194bp for the uncut version. The expected size for the PCR products with primers P3 and P4 was 128bp or 90bp for the 66bp or 104bp cleaved versions. As shown in **Figure 1F**, we successfully identified two homozygous deletions, namely '66bp del' (66bp deletion), '104bp del' (104bp deletion). We also successfully identified a homozygous 1.4kb deletion in the NCM460 cells (**Figure 1G**). HCT116 cells are heterozygous for the SNP rs7903146, harboring both C and T allele. We also identified heterozygous deletion clones, namely C/del and T/del using a similar approach in HCT116. The genotyping results confirmed the presence of the C allele (clone hap-C) or the T allele (clone hap-T) exclusively or the absence of both.

Array-based Genome wide expression.

Total RNA was isolated from the cell pellets using Trizol Reagent (Invitrogen), according to the manufacturer's protocol, and the RNA concentrations were determined by spectrophotometer (NanoDrop ND-1000). The GeneChip® Human Transcriptome Array 2.0 was leveraged at the Genomics Analysis Core at the University of Pennsylvania. All experiments were carried out in triplicate.

The Affymetrix Expression Console software was used to summarize probe set expression with the Robust Multichip Analysis (RMA) algorithm for normalization. The resulting values were used as input to the PaGE algorithm for Differential Expression analysis[2]. PaGE is a non-parametric permutation-based approach which corrects for multiple testing based on False Discovery Rate (FDR). We used an FDR cutoff of 10% to call significantly up- and down-regulated genes in HCT116 1.4kb deletion as compared to wild type. We then filtered the resulting lists of significant genes by a fold change greater than or equal to two.

4C and Capture C protocols

For both 4C (circularized chromosome conformation capture) and Capture C methods, the initial 3C libraries were generated for the HCT116, HCT116 1.4 kb deletion and NCM460 cells following a previously published protocol[15]. For each library, 10 million cells were harvested and fixed. Human colon tissue, flash frozen in liquid nitrogen, was obtained from the Cooperative Human Tissue Network (CHTN). 0.41g of human colon was ground in a liquid nitrogen-cooled mortar, with the resulting cells being fixed using a previously published protocol[15]. The DNA was digested using *DpnII*, then re-ligated together using T4 DNA ligase and finally isolated by phenol/chloroform extraction[15].

For the rest of the 4C method we followed another previously published protocol[16]. We used *NlaIII* as the second restriction enzyme to digest the DNA followed by T4 ligation. PCR primers were designed with Illumina adaptors to flank the region harboring rs7903146 near *DpnII* and *NlaIII* cut sites (**Table 1**). PCR and Illumina libraries were generated as previously described[16] with the exception of using a PCR annealing temperature of 55°C.

Following the previously published Capture C protocol[15] we utilized the above described 3C libraries for the capture procedure. The 3C libraries were sonicated to an average size of 200bp, ligated to Illumina adaptors, and amplified for 6 cycles with index primers to generate Illumina

libraries. 140bp biotinylated DNA oligo probes were designed to each *DpnII* cut site flanking rs7903146 (**Table 1**) and were used in the two sequential pull downs of the Capture libraries. All 4C and Capture C libraries were subsequently sequenced on the Illumina MiSeq platform.

5 RT-PCR and quantitative PCR

Reverse transcription reactions were performed using the High-Capacity RNA-to-cDNA™ Kit (Applied Biosystems). The resulting cDNA was diluted 1:10 and used for quantitative PCR. The primer sequences used for *ACSL5* and *TCF7L2* are listed in **Table 1**. The reactions were performed in a total volume of 15µl, consisting of 2µl cDNA and 2x SYBR®
10 Green Universal Master Mix (all quantitative PCR reagents were obtained from Applied Biosystems). Reactions were run in a 384-well plate in an ABI 7900 quantitative PCR machine and analyzed with Sequence Detection Systems 7900HT version 2.4 software (Applied Biosystems). Briefly, samples were heated to 50°C for 2 minutes and then 95°C for 10 minutes to allow for DNA polymerase activation followed by 40 cycles of 95° C for 15 seconds and 60°
15 C for 1 minute to allow for denaturing, annealing and extension. Average fold changes were calculated by differences in threshold cycles (Ct) between pairs of samples, with each sample run in triplicate; furthermore, the mean and standard deviation for the three values were calculated for comparison. Expression levels were normalized to *GAPDH*.

20 Western Blot Analysis

Cell pellets were lysed with RIPA buffer. Western blotting was performed according to standard procedures (<http://www.abcam.com/protocols/general-western-blot-protocol>). The primary antibodies for ACSL5 were purchased from Abcam (ab104892) while the antibodies for TCF7L2 (#2569) was purchased from Cell Signaling.

25

siRNA knockdown

To knock down *TCF7L2*, the cells were seeded in to 6-well plates (Corning) at a density of 1,000,000 cells/well. The cells were transfected via the addition of 100nM siRNA, (siGENOME Human TCF7L2 (6934) siRNA - SMARTpool, Dharmacon). PARP-1 siRNA was
30 purchased from Santa Cruz (sc-29437). After 48 hours of transfection, the cells were collected

for protein extraction. All siRNA experiments used an ON-TARGET plus Non-targeting siRNA #1 (Dharmacon) as a negative control.

PARP-1 inhibition

5 PARP-1 inhibitor Olaparib (AZD2281, KU0059436) was purchased from Selleckchem (Catalog No.S1060). The HCT116 and NCM460 cells were seeded in to 6-well plates (Corning) at a density of 1,000,000 cells/well and treated with 50uM Olaparib in 0.1% DMSO for 48 hours.

Sequence analyses

10 **4C:** primer sequences were trimmed from sequencing reads in paired-end mode using 'cutadapt'[17]. The hicup_truncater script from the HiCUP pipeline (<http://www.bioinformatics.babraham.ac.uk/projects/hicup>) was used to remove potential hybrid segments from the trimmed reads. Then reads were aligned to the hg19 assembly in paired-end mode with bowtie2 using a maximum fragment length of 2000-3000bp. Only concordant
15 alignments were retained. We used fourSig[18], which uses a non-parametric permutation-based approach that accounts for multiple-testing using False Discovery Rate (FDR), to identify statistically significant enrichment of reads in 5-fragment windows.

Capture C: Sequencing reads were processed using the HiCUP pipeline

20 (<http://www.bioinformatics.babraham.ac.uk/projects/hicup>), which consists of several steps: truncation of reads containing the *DpnII* restriction site, mapping to the human genome (hg19) with bowtie2, filtering out of reads resulting from experimental artifacts such as self- and re-ligations, and duplicate removal. Processed reads were assigned to either the bait or the target group, and reads in each *DpnII* fragment were counted using custom scripts. We used
25 fourSig[18] to identify statistically significant enrichment of reads in 5-fragment windows.

For both 4C and Capture C, due to the dependency of the number of reads on distance from the bait region, the analysis was performed separately in 4 regions (in: bait $\pm$ 50 kb; mid left: from bait - 200 kb to bait - 50 kb; mid right: from bait + 50kb to bait + 200 kb; out: everything else). Significant regions (FDR threshold =0.01) were then combined. Significant
30 regions were assigned to a promoter if within 5kb of a transcriptional start site (TSS).

The following examples are provided to illustrate certain embodiments of the invention. They are not intended to limit the invention in any way.

Example I

Targeting the genomic region encompassing rs7903146 in HCT116 cells

We generated a CRISPR/Cas9-mediated 1.4kb deletion in the HCT116 cell line (Figure 1). We then carried out array-based analysis to ascertain gene expression perturbation genome wide. We observed 99 genes with significant differential expression (with an FDR cutoff of 10%) and a fold-change ≥ 2 , of which 7 were down-regulated and the remaining up-regulated in the deletion setting (Table 2).

We then carried out a combination of 4C and Capture C followed by sequencing in two human colon cell lines (NCM460 and HCT116) in order to ascertain which of these perturbed genes' promoters made physical contact with the genomic region harboring the variant. We employed both techniques in parallel in order to ensure that there was consistency in the results irrespective of the chromatin capture method used. The oligos used for each experiment are shown in Table 1, and the metrics of sequencing and processing in Table 3. We focused our investigation on significant contacts to promoter regions (within 5 kb of the transcription start site or TSS) for the 99 genes identified in the microarray analysis, as described above, to ascertain if their expression was being directly physically impacted by the SNP region, as opposed to just as a consequence of indirect downstream perturbation. Among the differentially expressed genes in our array analysis, the only promoter interactions consistently detected in all experiments were between the SNP region and the promoter regions of ACSL5, residing in the same topologically associating domain (TAD) as TCF7L2, i.e. a chromatin compartment within which most enhancer-promoter contacts occur and where the causal gene most likely resides (Table 4, Figure 2).). We also detected consistent interactions with the promoters of five additional genes other than *TCF7L2* (*GUCY2GP*, *HABP2*, *NRAP*, *CASP7* and the miRNA *MIR4295*), all residing in the same TAD, plus one gene outside of the TAD boundaries (*NHLRC2*) (Table 4). These genes were not differentially expressed when comparing wild type versus edited cells (Table 2 and data not shown). As for *TCF7L2*, we detected consistent interactions with the promoter region of three short alternative transcripts (Table 4), however this gene was not among the 99 identified in the microarray analyses with the criteria described

above (see below for additional expression assays). We then elected to generate a series of deletions in the immediate region surrounding the SNP.

Along with the 1.4kb deletion described above, we generated two additional smaller deletions in HCT116 - 104bp and 66bp (Figure 1). Figure 3A shows the real time expression analysis of ACSL5 and how it was impacted by the various deletions. We observed consistent and highly significant reduction of ACSL5 gene expression; in contrast, when analyzing TCF7L2 itself, we observed a reduction in its expression but not as significant as for ACSL5. (Figure 3A).

In NCM460 cells, CRISPR/Cas9 proved relatively challenging; however, we were able to generate a clone with a 1.4kb deletion. The resulting cells grew extremely slowly but we were able to obtain sufficient RNA to test expression of ACSL5 and TCF7L2 (Figure 3B). We observed the same marked decrease in ACSL5 gene expression, while we observed a slight, albeit non-significant, increase in the gene expression of TCF7L2, contrary to the direction we observed in HCT116 cells.

Given these observations, we also assessed the expression of ACSL5 and TCF7L2 at the protein level in HCT116. Similar to the real time mRNA observations, we saw a dramatic decrease in ACSL5 protein expression, with almost total ablation with the biggest deletion (Figure 4A and 4B). With respect to TCF7L2, we also observed a reduction in its expression but not as marked as for ACSL5 (Figure 4A and 4C).

To investigate whether deletion of the genomic element harboring rs7903146 had an impact on the long-range contacts observed, we performed 4C using the HCT116 cell line harboring the 1.4kb deletion. Interestingly, the peaks corresponding to contacts with the ACSL5 promoters in the wild type setting were all abolished by the deletion, further suggesting that this region is critical to establishing this long-range interaction (overview in Figure 5, which also includes the other genes in the TAD with consistent promoter contacts to the SNP region; all contacts for CASP7, GUCY2GP and HABP2, plus some contacts for NRAP and TCF7L2, were also abolished by the CRISPR deletion, while the contact for the MIR4295 miRNA was not abolished; data not shown).

Table 2

GENE SYMBOL	PROBESET	log2(WT1)	log2(WT2)	log2(WT3)	AVG(log2(WT))	log2(DEL1)	log2(DEL2)	log2(DEL3)	AVG(log2(DEL))	Fold Change	q-value
GPR110	TC06001776.hg.1	8.592319	8.430242	8.494238	8.50599667	5.120279	5.157203	5.180912	5.152798	-10.2163055	0.084
ACS15**	TC10000813.hg.1	9.761443	9.63432	9.770786	9.722183	6.984656	6.985536	6.87293	6.947707333	-6.842272997	0.084
SQLE	TC08000729.hg.1	9.404976	9.411249	9.5443	9.453508333	6.845356	6.905558	6.923849	6.891587667	-5.904932903	0.084
RNA5SP156	TC04001038.hg.1	9.087831	9.163441	9.2151	9.15457333	6.786371	6.868299	6.975186	6.876618667	-4.852871525	0.1
PLAU*	TC10000475.hg.1	9.703062	9.770308	9.67142	9.71493	7.416939	7.5303	7.473956	7.473731667	-4.727896107	0.1
LRP1B	TC02002388.hg.1	6.984005	6.997366	7.16598	7.049117	4.817976	4.851053	4.887233	4.852087333	-4.585343032	0.1
CASP4	TC11002246.hg.1	7.78933	7.872336	7.784963	7.815543	5.731423	5.664412	5.781729	5.725854667	-4.256561082	0.1
EDA2R	TC0X001101.hg.1	5.815779	5.821071	5.812994	5.816614667	8.980083	8.984767	8.997868	8.987572667	9.006446488	0.05
KIAA0040	TC01005949.hg.1	8.257838	8.332206	8.318351	8.302798333	10.50424	10.4772	10.52071	10.49905	4.582870968	0.063
RPS4Y1	TC0Y000351.hg.1	6.291416	6.296387	6.324342	6.304048333	8.346901	8.287704	8.371258	8.335287667	4.087558373	0.063
SPARC	TC05003313.hg.1	7.146082	7.091359	6.634519	6.95732	10.70762	10.61122	10.83032	10.71638667	13.53916317	0.065
IFI27	TC14000584.hg.1	7.340064	7.33849	6.98732	7.221958	11.03049	11.11421	10.45394	10.86621333	12.50345879	0.065
IFI27	TC14001821.hg.1	7.452641	7.515418	7.29217	7.420076333	11.05321	11.24383	10.55625	10.95109667	11.5596061	0.065
IFI6	TC01002412.hg.1	9.697481	9.444965	9.075848	9.406098	12.75968	12.71234	12.14111	12.53771	8.764136778	0.065
TXLNGY	TC0Y000354.hg.1	5.265135	5.189979	5.547793	5.334302333	8.358295	8.250803	8.483814	8.364304	8.168106442	0.065
UTY	TC0Y000317.hg.1	5.720562	5.823657	6.443296	5.995838333	9.031212	8.553227	8.654224	8.746221	6.728955903	0.065
DDX3Y	TC0Y000053.hg.1	6.265801	6.108759	6.48471	6.286423333	8.835964	8.68859	8.91328	8.812611333	5.760475882	0.065
TTY15	TC0Y000051.hg.1	5.252861	5.09968	5.138774	5.163771667	7.828112	7.637826	7.427339	7.631092333	5.530157873	0.065
ANXA10	TC04000836.hg.1	5.632541	5.519178	5.678651	5.610123333	7.987438	7.870935	8.177071	8.011814667	5.284222935	0.065
BST2	TC19001275.hg.1	6.72316	6.717154	6.6285	6.689604667	9.270431	9.283128	8.644706	9.06088333	5.192695639	0.065
USP9Y	TC0Y000052.hg.1	5.27498	5.232227	5.484575	5.330594	7.765163	7.649754	7.682521	7.699146	5.164225506	0.065
SIC2A3	TC12002734.hg.1	9.08069	9.317242	9.373422	9.257118	11.45537	11.22819	11.66006	11.44787333	4.565444509	0.065
TXLNGY	TC0Y000353.hg.1	5.76019	5.724464	5.888038	5.790897333	8.190057	7.808205	7.850028	7.94943	4.464605391	0.065
BAMBI*	TC10000199.hg.1	7.365927	7.388284	7.455973	7.403394667	9.605789	9.30644	9.639117	9.517115333	4.328060502	0.065
LOC151760	TC03001641.hg.1	5.92014	5.975412	5.99036	5.961970667	8.245943	7.97744	7.970974	8.064785667	4.295467034	0.065
SIC2A3	TC12001170.hg.1	8.298094	8.518245	8.584589	8.466976	10.40323	10.20048	10.63352	10.41241	3.85153624	0.065
UTY	TC0Y000163.hg.1	5.515998	5.463736	5.582058	5.520597333	7.485515	7.386959	7.510232	7.460902	3.837866868	0.065
KRTAP2-3	TC17001485.hg.1	9.234018	9.468987	9.288866	9.330623667	11.31498	11.35853	11.01241	11.22864	3.727003912	0.065
KRTAP3-1	TC17001478.hg.1	5.634204	5.383188	5.234241	5.417211	7.33933	7.27337	7.226798	7.279832667	3.636679197	0.065
LOC101928687*	TC10001299.hg.1	6.794942	6.810825	7.006038	6.870601667	8.682969	8.634595	8.675056	8.664206667	3.466800938	0.065
HEPH	TC0X001747.hg.1	4.420486	4.367289	4.650907	4.479560667	6.36164	6.441517	5.995032	6.266063	3.449775151	0.065
RNF43	TC17001732.hg.1	8.157213	8.260593	8.18239	8.20065333	9.902535	9.975638	9.927682	9.935285	3.329301824	0.065
KDM5D	TC0Y000185.hg.1	6.035546	6.030663	6.045216	6.037141667	7.723364	7.777399	7.519302	7.673355	3.108488698	0.065
FUT8	TC14001735.hg.1	9.185688	9.1399	9.211544	9.179044	10.79283	10.78763	10.8652	10.81522	3.108408259	0.065
FUT8	TC14000397.hg.1	7.954855	7.921098	8.028675	7.968209333	9.595393	9.533814	9.63257	9.593925667	3.085953522	0.065
FAM49A	TC02001600.hg.1	5.802835	5.844225	5.799907	5.815655667	7.425217	7.426537	7.414908	7.422220667	3.045259137	0.065
RP11-675f6.3	TC08001150.hg.1	5.220485	5.184109	5.206293	5.203629	6.782826	6.788001	6.80935	6.793392333	3.00999679	0.065
MIR221	TC0X000985.hg.1	9.55582	9.684757	9.661674	9.634083667	11.262	11.28643	11.08225	11.21022667	2.981716307	0.065
CSF2RA	TC0Y000006.hg.1	7.595946	7.542838	7.53432	7.557701333	9.193482	9.167864	9.029435	9.130260333	2.974318203	0.065
KIAA0040	TC01003546.hg.1	8.299541	8.350165	8.344116	8.331274	9.795712	9.905576	9.991023	9.897437	2.961161138	0.065
ITGB8	TC07000125.hg.1	6.112734	6.123369	6.131175	6.122426	7.61638	7.488762	7.959381	7.688174333	2.960310148	0.065
PDIM5	TC04000501.hg.1	7.338852	7.280774	7.329552	7.316392667	8.903872	8.840239	8.895254	8.879788333	2.955486574	0.065

SPARC	TC05001955.hg.1	7.627716	7.685898	7.534703	7.616105667	9.117394	9.041168	9.254514	9.137692	2.871065677	0.065
SAMHD1	TC200008211.hg.1	8.202301	7.945538	8.141726	8.096521667	9.701351	9.579802	9.549488	9.610213667	2.855398297	0.065
ITGBL1	TC13001260.hg.1	4.316439	4.209159	4.324227	4.283275	5.773517	5.701839	5.914433	5.796596333	2.854664764	0.065
ENPP1	TC06000985.hg.1	6.973175	6.857592	6.942306	6.924357667	8.484623	8.337476	8.392363	8.404820667	2.790382699	0.065
SORT1	TC01002954.hg.1	7.731912	7.725256	7.86454	7.773902667	9.23362	9.176573	9.2722	9.227464333	2.738833695	0.065
BVES	TC060001976.hg.1	7.787455	7.596719	7.872069	7.752081	9.226841	9.049046	9.302456	9.192781	2.714525431	0.065
HUA-DRB1	TC6_apd_hap1000101.hg.1	8.051876	8.124769	8.059359	8.078668	9.515738	9.469229	9.504034	9.496333667	2.671528979	0.065
CSF2RA	TC0X000006.hg.1	7.255936	7.152455	7.209499	7.205963333	8.702178	8.605497	8.558788	8.62154333	2.668799641	0.065
MIR222	TC0X000986.hg.1	8.33212	8.38916	8.389478	8.370252667	9.689595	9.703409	9.956873	9.783292333	2.662976444	0.065
SEMA3A	TC07001568.hg.1	8.946352	8.749907	9.017911	8.904723333	10.35141	10.27683	10.29602	10.30808667	2.645175297	0.065
SERPINA3	TC14002309.hg.1	6.906319	6.904289	6.779182	6.863263333	8.143644	8.298797	8.245246	8.229229	2.577487921	0.065
GCH1	TC14002019.hg.1	8.11276	7.912842	7.934271	7.986624333	9.43968	9.268446	9.328276	9.345467333	2.564794079	0.065
HUA-DRB1	TC6_cox_hap2000193.hg.1	8.153399	8.188405	8.121857	8.154533667	9.534574	9.478212	9.519663	9.510816333	2.560210914	0.065
HUA-DRB1	TC6_qbl_hap6000182.hg.1	8.043717	8.091655	8.0353	8.056890667	9.432281	9.380839	9.42274	9.411953333	2.558082276	0.065
DZIP3	TC03000548.hg.1	7.163749	6.953986	7.265183	7.127639333	8.528519	8.38878	8.348589	8.421962667	2.452619334	0.065
HUA-DRB1	TC6_ssto_hap7000164.hg.1	8.757843	8.837186	8.659019	8.751349333	10.04433	9.9298	10.13273	10.03562	2.43558894	0.065
DKK1*	TC10000350.hg.1	9.852479	9.622636	9.879656	9.784923667	11.07257	10.98758	11.07257	11.04424	2.393822752	0.065
EIF1AY	TC0Y000072.hg.1	4.785937	4.706617	4.81019	4.767581333	6.14515	5.937682	5.995842	6.026224667	2.392706323	0.065
HUA-DRB1	TC6_dbb_hap3000180.hg.1	8.057549	8.180483	8.035787	8.091273	9.523892	9.215746	9.301241	9.346959667	2.387807719	0.065
TMEM2	TC090001204.hg.1	7.398883	7.410414	7.568018	7.491505	8.710566	8.678666	8.715393	8.701541667	2.365978013	0.065
SEMA3C	TC070001559.hg.1	7.044285	7.022013	7.226575	7.097624333	8.310979	8.245647	8.453609	8.336745	2.360546114	0.065
CASD1	TC07000561.hg.1	6.824064	6.761492	6.875274	6.820276667	8.097409	8.029128	8.018241	8.048259333	2.342392219	0.065
SIPA1L2	TC01003974.hg.1	7.767325	7.783907	7.816182	7.789138	9.066492	8.993592	8.972109	9.010731	2.332040753	0.065
CDK6	TC070003061.hg.1	8.721735	8.67684	8.882369	8.760314667	9.886213	9.847587	10.05454	9.929446667	2.24876359	0.065
SIC7A2	TC080001816.hg.1	5.869527	6.021367	6.15488	6.015258	7.153896	7.156886	7.160332	7.157038	2.206530975	0.065
RP56KA5	TC14001434.hg.1	7.893776	7.832983	7.889707	7.872155333	9.074492	8.988648	8.969036	9.010725333	2.201626896	0.065
SIC2A1	TC01002578.hg.1	10.30562	10.50402	10.44369	10.41777667	11.5287	11.47651	11.55701	11.52074	2.147954351	0.065
PARM1	TC04002072.hg.1	6.230497	6.048861	6.197627	6.158995	7.358914	7.17873	7.215385	7.251009667	2.131715143	0.065
KALP	TC0Y000055.hg.1	6.025469	6.159663	6.068521	6.084551	7.194567	7.194262	7.127708	7.172179	2.125243282	0.065
PRKD3	TC02001742.hg.1	7.10182	7.044928	7.304277	7.150341667	8.25629	8.185747	8.257496	8.233177667	2.118195867	0.065
SIC20A1	TC02000711.hg.1	9.09886	9.133432	9.156731	9.129674333	10.30754	10.23629	10.09077	10.21153333	2.116761901	0.065
RBM47	TC04002523.hg.1	8.181325	8.060573	8.006923	8.082940333	9.164579	9.123422	9.168431	9.152144	2.09827485	0.065
CNTNAP3B	TC09000274.hg.1	8.884767	8.848718	8.949992	8.894492333	9.867293	9.94053	10.07314	9.960321	2.093371942	0.065
ZFY	TC0Y000230.hg.1	3.752921	3.742309	3.886209	3.793813	4.79353	4.967866	4.762388	4.841261333	2.066870978	0.065
CNTNAP3P2	TC09001097.hg.1	8.780784	8.671496	8.808092	8.753457333	9.775105	9.76791	9.794024	9.779013	2.035743318	0.065
CD22	TC19002611.hg.1	7.496711	7.472528	7.357891	7.442376667	8.511424	8.444981	8.407759	8.454721333	2.017186768	0.065
PLCX01	TC0Y000001.hg.1	7.801568	7.887932	7.78853	7.82601	8.869998	8.808192	8.831636	8.836608667	2.014746974	0.065
SIC25A4	TC04000902.hg.1	7.731592	7.727855	7.677032	7.712159667	8.78382	8.719345	8.637093	8.713419333	2.001747031	0.065
ITGB8	TC07002185.hg.1	4.169065	4.150344	4.528804	4.282737667	6.594495	6.183318	7.057103	6.611638667	5.024224743	0.066
IFT1*	TC100000639.hg.1	5.545639	5.41373	5.429467	5.462945333	8.030437	7.875544	7.211888	7.705956333	4.733840181	0.066
IFIH1	TC02002481.hg.1	7.376666	7.30839	7.387425	7.357493667	9.263865	9.169157	8.690556	9.041192667	3.712505669	0.066
STAT1	TC02002624.hg.1	9.447249	9.33216	9.350792	9.376733667	11.24442	11.13776	10.69997	11.02738333	3.139749949	0.066
TNFSF18	TC01003523.hg.1	6.867813	6.623658	6.759576	6.750349	8.637429	8.335772	8.084942	8.352714333	3.036407317	0.066

SP110	TC02002853.hg.1	7.416396	7.332067	7.327757	7.35874	9.171918	9.103541	8.546068	8.940509	2.993366646	0.066
HERC6	TC04000484.hg.1	7.133396	7.177999	7.07971	7.130368333	8.912126	8.769859	8.298826	8.660270333	2.88766223	0.066
IFI16	TC010004781.hg.1	4.244726	4.419363	4.179866	4.281318333	5.520812	5.836253	5.423042	5.593369	2.482942181	0.066
SAMD9	TC070001605.hg.1	7.478155	7.260333	7.426781	7.388423	8.843756	8.772536	8.451249	8.689180333	2.463581728	0.066
PEG10	TC070000562.hg.1	7.250421	7.302387	7.158503	7.237103667	8.323639	8.535298	8.731679	8.530205333	2.450543347	0.066
EIF2AK2	TC020004161.hg.1	10.5599	10.35248	10.53463	10.48233667	11.87697	11.8624	11.51492	11.75143	2.410100543	0.066
EIF2AK2	TC020001739.hg.1	10.1688	9.956161	10.12231	10.08242367	11.43149	11.30942	11.06964	11.27018333	2.277987241	0.066
EGR1	TC050000701.hg.1	9.109491	9.197609	8.944432	9.083844	10.36905	10.30115	10.08694	10.25238	2.247834782	0.066
SVTL4	TC0X0001792.hg.1	6.638273	6.467709	6.62708	6.577563333	7.628644	7.552915	7.888414	7.689991	2.162091631	0.066
PLSCR1	TC030001869.hg.1	7.570716	7.471206	7.613105	7.551675667	8.809301	8.721548	8.410266	8.647038333	2.136667867	0.066
FOLR1	TC11002773.hg.1	8.130079	7.988111	8.13183	8.08334	9.188298	8.992875	9.296123	9.159098667	2.10783022	0.066
EFCA2	TC010005153.hg.1	5.454544	5.287477	5.289515	5.343845333	6.59807	6.2116	6.445354	6.418341333	2.105986225	0.066
CLIP4	TC020000192.hg.1	8.236803	8.192588	8.447588	8.292326333	9.352991	9.310345	9.430964	9.364766667	2.102987585	0.066
TDRD7	TC090000490.hg.1	7.477938	7.545817	7.593023	7.538926	8.688763	8.672197	8.377446	8.579468667	2.057001244	0.066
AL353791.1	TC090000231.hg.1	6.957292	6.767798	7.142112	6.955734	7.902903	8.012329	8.057733	7.990988333	2.049474917	0.066
RRAGA	TC090000089.hg.1	6.090742	6.130761	6.302799	6.174767333	7.348462	7.052833	7.175847	7.192380667	2.024566974	0.066
GCH1	TC140001152.hg.1	7.107457	6.889773	7.038494	7.011908	8.112754	7.915864	8.042753	8.023790333	2.016540433	0.066
LINC01405	TC120002568.hg.1	7.0116	6.808451	7.004365	6.941472	7.898651	7.879567	8.072917	7.950378333	2.012384989	0.066
TMEM62	TC150000324.hg.1	7.488865	7.432105	7.395287	7.438752333	8.597935	8.445536	8.298452	8.447307667	2.011895446	0.066
DDX60	TC040001718.hg.1	6.910216	6.913398	6.95328	6.925631333	8.935765	8.822409	8.169574	8.642582667	3.287409853	0.067
IFI44	TC010000795.hg.1	4.593648	4.575769	4.677972	4.615796333	7.227855	6.958635	6.080028	6.755506	4.406733547	0.068
USP18	TC220000029.hg.1	7.258577	7.336696	7.229379	7.274884	8.818762	8.785658	8.24678	8.617066667	2.535346036	0.068
MIR181B1	TC010003671.hg.1	7.111768	7.020969	6.948081	7.026939333	7.91246	8.294816	8.472959	8.226745	2.297087268	0.068
PARP12	TC070001916.hg.1	8.093146	8.053158	8.011609	8.052637667	9.261798	9.321334	8.855282	9.146138	2.133911484	0.068
ARID5B*	TC10002092.hg.1	9.157546	8.825983	9.01428	8.99269667	10.03541	9.902898	10.28832	10.07554267	2.108581814	0.068
DDX58	TC090000999.hg.1	7.957823	7.882794	8.125931	7.98849333	9.599522	9.496011	8.981726	9.359086333	2.585130301	0.069
PARP9	TC030001705.hg.1	7.630479	7.600169	7.571609	7.600752333	8.799061	8.676964	8.334315	8.603446667	2.003738629	0.069
OAS1	TC120000884.hg.1	5.951177	6.036322	5.945669	5.977722667	7.211936	7.071826	6.695856	6.993206	2.021580052	0.071
OASL	TC120002059.hg.1	7.125969	7.208574	6.98483	7.106457667	8.28475	8.321476	7.737112	8.114446	2.011104897	0.074

Table 3. Metrics of 4C and Capture C sequencing and processing.

NCM460_CapC	Read1	Read2	Filtering	Di-Tag Count
Total Reads	30,055,019	30,055,019	Same Circularized	1,005,674
Not Truncated	24,939,663	24,612,071	Same Fragment Dangling Ends	862,153
Truncated	5,115,356	5,442,948	Same Fragment Internal	10,216,485
Too short to map	1,472,516	1,641,054	Re-ligation	3,172,248
Unique Alignments	26,550,583	26,176,013	Contiguous Sequence	19,195
Multiple Alignments	1,866,732	1,877,387	Wrong Size	120,866
Failed To Align	165,188	360,565	Total Pairs	21,515,367
Paired	21,515,367	21,515,367	Valid Pairs	6,118,746
Average length of truncated sequence	28.08	27.66	Unique Di-Tags	92,421
			On Target Di-Tags	78,464
HCT116_CapC	Read 1	Read 2	Filtering	Di-Tag Count
Total Reads	27,731,176	27,731,176	Same Circularized	261,375
Not Truncated	24,540,046	24,358,664	Same Fragment Dangling Ends	505,202
Truncated	3,191,130	3,372,512	Same Fragment Internal	6,611,220
Too short to map	933,194	1,004,265	Re-ligation	2,513,564
Average length of truncated sequence	27.91	27.94	Contiguous Sequence	28,860
Unique Alignments	17,101,688	17,056,268	Wrong Size	66,179
Multiple Alignments	1,096,277	1,011,558	Total Pairs	14,274,734
Failed To Align	8,600,017	8,659,085	Valid Pairs	4,288,334
Paired	14,274,734	14,274,734	Unique Di-Tags	88,718
			On Target Di-Tags	54,670
NCM460_4C	Read Pairs		Colon_4C	Read Pairs
Total Number of Read Pairs	23,981,159		Total Number of Read Pairs	8,431,008
Number Passing Filters	3,253,060		Number Passing Filters	4,473,537
NCM460_4C_2p	Read Pairs			
Total Number of Read Pairs	7,697,481			
Number Passing Filters	2,388,476			
HCT116_4C	Read Pairs			
Total Number of Read Pairs	17,381,980			
Number Passing Filters	2,680,316			
HCT116_del1.4_4C	Read Pairs			
Total Number of Read Pairs	7,306,205			
Number Passing Filters	948,767			

Given that TCF7L2 is a relatively ubiquitously expressed transcription factor, we investigated the possibility that TCF7L2 regulates ACSL5 expression. When we knocked down TCF7L2 gene expression with siRNA, we observed only a subtle impact on ACSL5 expression (Figure 6), suggesting that these observations are largely specific to the enhancer region and not
5 being controlled by TCF7L2.

In addition, we previously reported that a factor called PARP-1, plus its partners, bind across this region[19]. As such, in order to investigate a possible impact on ACSL5 expression in a comparable fashion to what was observed with the CRISPR generated deletion series, we used both a PARP-1 inhibitor and siRNA knockdown of PARP-1. Indeed we observed a significant
10 impact at both the RNA and the protein level in both NCM460 and HCT116 (Figure 7).

To investigate allelic differences, and leveraging the fact that HCT116 is heterozygous for rs7903146, we used CRISPR/Cas9 to generate clones harboring the C allele or the T allele exclusively. We generated cDNA from both wild type and targeted cells and carried out quantitative PCR to determine the levels of ACSL5 mRNA in these cells. As shown in Figure
15 8A, deletion of either the C or T allele reduced expression of ACSL5 at the mRNA level. Interestingly, the T allele clone revealed slightly stronger (but statistically significant) expression of ACSL5 than the C allele. Consistent with the results from quantitative PCR, protein levels of ACSL5 were also reduced in the deletion cells while the T allele clone exhibited relatively higher expression than the C allele (Figure 8B and 8C). These results are compatible with the T
20 allele driving a stronger enhancer for ACSL5 expression than the C allele.

Finally, to validate our results from cell lines experiments, we performed 4C in frozen human colon tissue (data not shown). In this setting, consistent with the cell lines chromatin conformation capture experiments, we detected interactions between the rs7903146 SNP region and one ACSL5 promoter (specific for a short alternative transcript), plus contacts with HABP2,
25 MIR4295, NHLRC2, NRAP and TCF7L2 promoters (Table 4).

Table 4: Genes showing consistent interactions (within 5kb of their transcriptional start sites) with the rs7903146 SNP region in five chromatin capture experiments (4C and Capture C in NCM460 and HCT116 cells, plus 4C with a different bait primer set in NCM460 cells).

Interactions confirmed in one 4C experiment in human colon tissue are marked with an asterisk.

- 5 Coordinates refer to the GRCh37/hg19 build and gene models are the UCSC genes. Of these, the only gene identified as differentially expressed in HCT116 cells harboring a deletion of the SNP region was *ACSL5*.

Gene	Transcript	Chr	TSS pos	Strand
<i>ACSL5</i>	uc001kzs.3	chr10	114133915	+
<i>ACSL5</i>	uc001kzt.3	chr10	114135022	+
<i>ACSL5</i>	uc001kzu.3	chr10	114135955	+
<i>ACSL5</i>	uc009xxz.3 *	chr10	114154675	+
<i>ACSL5</i>	uc010qrj.2	chr10	114169264	+
<i>CASP7</i>	uc010qsb.3	chr10	115469103	+
<i>GUCY2GP</i>	uc010qri.2	chr10	114067935	-
<i>HABP2</i>	uc001lai.4	chr10	115312777	+
<i>HABP2</i>	uc010qry.1	chr10	115312777	+
<i>HABP2</i>	uc010qrz.1	chr10	115312777	+
<i>HABP2</i>	uc021pyr.1 *	chr10	115310589	+
<i>MIR4295</i>	uc021pyf.1 *	chr10	114393928	+
<i>NHLRC2</i>	uc001lax.2 *	chr10	115614390	+
<i>NRAP</i>	uc001laj.4	chr10	115348582	-
<i>NRAP</i>	uc001lak.4	chr10	115348582	-
<i>NRAP</i>	uc001lal.4	chr10	115348582	-
<i>NRAP</i>	uc009xyb.3	chr10	115348582	-
<i>TCF7L2</i>	uc010qrv.2 *	chr10	114886572	+
<i>TCF7L2</i>	uc010qrw.2 *	chr10	114889059	+
<i>TCF7L2</i>	uc010qrx.2 *	chr10	114889391	+

Discussion

We observed that combining data resulting from the deletion of the immediate genomic region harboring rs7903146 and the characterization of promoter contacts strongly implicated ACSL5 at this T2D-associated locus. Furthermore, it resides in the same TAD as TCF7L2. This gene encodes an acyl-coA synthetase essential for fatty acid metabolism.

There was a relatively modest impact on TCF7L2 expression with the deletion approach (when contrasted with consistent ACSL5 observations across all deletions), and was in opposite directions in the two different cell lines, leading to uncertainty about the physiological relevance of this observation.

Furthermore, we did observe some evidence for other promoters in the TAD showing consistent contact with the genomic region harboring the variant, namely HABP2, GUCY2GP, CASP7, NRAP, NHLRC2 and MIR4295, but even if their contacts with the rs7903146 SNP region were abolished in the deletion cell line (except for MIR4295), an enhancer role of the SNP region for these genes was not supported by our expression data when comparing wild type versus edited cells (Table 2 and data not shown). Importantly, most of the contacts detected by chromatin conformation capture experiments in the intestinal-related cell lines were conserved in human colon tissue, including ACSL5 (Table 4).

It should be noted that the primary focus of our CRISPR/Cas9 work involved the editing of the immediate genomic neighborhood surrounding the key SNP. The rationale was that if a putative enhancer was removed, it would have a substantial impact on the expression of one or more genes; indeed that is what we observed with ACSL5, which happens to reside in the same TAD as TCF7L2. On the other hand, with respect to limiting our investigations to allele specific effects, our targeted cell line experiments showed that the change of a C to a T at rs7903146 had a relatively modest effect on ACSL5 expression (but somewhat more apparent at the protein level), which may be in line with expectations given that this susceptibility variant, and indeed the vast majority of GWAS-reported signals, only raise one's risk relatively modestly for T2D.

The genomic sequence harboring rs7903146 therefore appears to act as an enhancer, given that when we knock out the region, ACSL5 expression is reduced. Further evidence for this comes from a previous study of human pancreatic islets, where the T2D T risk allele was

found enriched in open chromatin[20]. It should be noted that the immediate genomic region flanking rs7903146 is not conserved in mouse suggesting mechanistic differences between species, and that a rodent model may not be optimal for the question at hand.

There is an ongoing intense discussion in the field about which tissue(s) this locus principally acts to confer its effect on T2D risk. Undeterred, in the setting we elected to work in, we observe a strong relationship between the genomic location harboring rs7903146 and the expression of ACSL5 with this ‘variant to gene mapping’ effort. In fact, one must be aware that chromatin accessibility is tissue specific – and there have been reports where TCF7L2 expression in pancreatic islets is counter to other tissues[21]. However, our results show that in the cell setting we elected to carry out this study, the rs7903146 region operates in a similar manner to the FTO locus in obesity[1, 2], where the expression of neighboring genes is influenced by an apparent embedded enhancer in a given gene, in this case TCF7L2.

Inhibiting the ACSL5 enzyme activity could be a promising avenue for improving insulin sensitivity, given that changes in free fatty acid levels can lead to insulin resistance. There have been two independent studies in ACSL5 knock-out mice. In the earlier study, the long chain acyl-CoA synthesis rate was reduced by 60% in the knockout mice but the absorption of dietary long chain fatty acids remained unaffected in response to high fat diet[22]. In contrast, a more recent study showed that ACSL5 ablation in mice increased energy expenditure and insulin sensitivity and delayed fat absorption[23].

In conclusion, the rs7903146 region appears to interact with the promoter of ACSL5. Deletion of the region in two human colon cell lines dramatically decreases expression of ACSL5. We also did observe some effect on TCF7L2 expression – but not as strong as for ACSL5. PARP-1 inhibition also impacts ACSL5 expression. Due to the exquisite tissue specificity of enhancer-promoter interactions, it is conceivable that other target genes for which we see interactions in our chromatin conformation experiments might be relevant in other tissues - this region might indeed be a locus control region. Given the important role of ACSL5 in lipid biosynthesis and fatty acid degradation, it appears that insulin sensitivity can be improved substantially by inhibiting ACSL5 enzyme activity.

References

- [1] Smemo S, Tena JJ, Kim KH, et al. (2014) Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* 507: 371-375
- [2] Claussnitzer M, Dankel SN, Kim KH, et al. (2015) FTO Obesity Variant Circuitry and
5 Adipocyte Browning in Humans. *N Engl J Med* 373: 895-907
- [3] Grant SF, Thorleifsson G, Reynisdottir I, et al. (2006) Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nature genetics* 38: 320-323
- [4] DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium (2014) Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type
10 2 diabetes susceptibility. *Nature genetics* 46: 234-244
- [5] Lyssenko V, Lupi R, Marchetti P, et al. (2007) Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes. *The Journal of clinical investigation* 117: 2155-2163
- [6] Boj SF, van Es JH, Huch M, et al. (2012) Diabetes risk gene and Wnt effector
15 Tcf7l2/TCF4 controls hepatic response to perinatal and adult metabolic demand. *Cell* 151: 1595-1607
- [7] Kaminska D, Kuulasmaa T, Venesmaa S, et al. (2012) Adipose tissue TCF7L2 splicing is regulated by weight loss and associates with glucose and fatty acid metabolism. *Diabetes* 61: 2807-2813
- [8] Yi F, Brubaker PL, Jin T (2005) TCF-4 mediates cell type-specific regulation of proglucagon gene expression by beta-catenin and glycogen synthase kinase-3beta. *The Journal of biological chemistry* 280: 1457-1464
20
- [9] Duval A, Rolland S, Tubacher E, Bui H, Thomas G, Hamelin R (2000) The human T-cell transcription factor-4 gene: structure, extensive characterization of alternative splicings, and
25 mutational analysis in colorectal cancer cell lines. *Cancer research* 60: 3872-3879

- [10] Korinek V, Barker N, Moerer P, et al. (1998) Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nature genetics* 19: 379-383
- [11] Helgason A, Palsson S, Thorleifsson G, et al. (2007) Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution. *Nature genetics* 39: 218-225
- 5 [12] Palmer ND, Hester JM, An SS, et al. (2011) Resequencing and analysis of variation in the TCF7L2 gene in African Americans suggests that SNP rs7903146 is the causal diabetes susceptibility variant. *Diabetes* 60: 662-668
- [13] Wellcome Trust Case Control C, Maller JB, McVean G, et al. (2012) Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nature genetics* 44: 1294-
10 1301
- [14] Cong L, Ran FA, Cox D, et al. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339: 819-823
- [15] Hughes JR, Roberts N, McGowan S, et al. (2014) Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nature genetics* 46: 205-
15 212
- [16] van de Werken HJ, Landan G, Holwerda SJ, et al. (2012) Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nature methods* 9: 969-972
- [17] Martin M (2012) Cutadapt removes adapter sequences from high-throughput sequencing reads. *embNET Journal* 17
- 20 [18] Williams RL, Jr., Starmer J, Mugford JW, et al. (2014) fourSig: a method for determining chromosomal interactions in 4C-Seq data. *Nucleic acids research* 42: e68
- [19] Xia Q, Deliard S, Yuan CX, Johnson ME, Grant SF (2015) Characterization of the transcriptional machinery bound across the widely presumed type 2 diabetes causal variant, rs7903146, within TCF7L2. *Eur J Hum Genet* 23: 103-109
- 25 [20] Gaulton KJ, Nammo T, Pasquali L, et al. (2010) A map of open chromatin in human pancreatic islets. *Nature genetics* 42: 255-259

[21] Savic D, Ye H, Aneas I, Park SY, Bell GI, Nobrega MA (2011) Alterations in TCF7L2 expression define its role as a key regulator of glucose metabolism. *Genome research* 21: 1417-1425

[22] Meller N, Morgan ME, Wong WP, Altemus JB, Sehayek E (2013) Targeting of Acyl-CoA synthetase 5 decreases jejunal fatty acid activation with no effect on dietary long-chain fatty acid absorption. *Lipids in health and disease* 12: 88

[23] Bowman T, O'Keeffe K, D'Aquila T, et al. (2016) CoA synthetase 5 (ACSL5) Ablation in Mice Increases Energy Expenditure and Insulin Sensitivity and Delays Fat Absorption. *Molecular Metabolism* doi: 10.1016/j.molmet.2016.01.001.

10

While certain of the preferred embodiments of the present invention have been described and specifically exemplified above, it is not intended that the invention be limited to such embodiments. It will be apparent to one skilled in the art that various changes and modifications can be made therein without departing from the scope of the present invention, as set forth in the following claims.

15

20

What is claimed is:

1. An isolated nucleic acid binding complex comprising a SNP containing transcription factor 7-like 2 (TCF7L2) encoding nucleic acid, and a nucleic acid selected from *RBM20*, *PDCD4*, *MIR4680*, *BBIP1*, *SHOC2*, *RPL13AP6*, *ADRA2A*, *GPAM*, *TECTB*, *MIR6715B*, *GUCY2GP*,
5 *ACSL5*, *ZDHHC6*, *VTI1A*, *MIR4295*, *LOC103344931*, *TCF7L2*, *SNORA87A*, *HABP2*, *NRAP*, *CASP7*.
2. The isolated binding complex of claim 1, wherein said complex comprises TCF7L2 and ACSL5.
3. A method for identifying agents which disrupt the binding complex of claim 1 or claim 2
10 thereby modulating TCF7L2 embedded enhancer/silencer function, comprising;
 - a) incubating said complex in the presence and absence of an effective amount of said agent, said complex comprising at least one detectably labeled protein or nucleic acid;
 - b) measuring disruption of said binding complex in the presence of said agent
15 relative to that observed in the absence of said agent, agents which disrupt said complex being identified as modulators of function of one or more said genes in claim 1.
4. The method of claim 3, wherein said method is performed in a cell and said function is selected from the group consisting of Wnt signaling, chromatin remodeling, activation of target gene expression and DNA damage detection and repair.
5. The method of claim 4, wherein said agent is selected from the group consisting of a siRNA,
20 an antisense oligonucleotide, a small molecule, and a peptide.
6. The method of claim 5 wherein said cells are selected from the group consisting of H116 cells, NCM460 cells, INS cells, PC12 cells, MIN6 cells, pancreatic beta islet cells and alpha TC6 cells.
7. A method for identifying agents which modulate ACSL5 activity, comprising;
 - a) incubating cells comprising an enhancer element within a topologically
25 associated domain containing TCF7L2 and ACSL5 in the presence and absence of an effective amount of said agent;

b) measuring one or more parameter of ACSL5 activity in the presence of said agent relative to that observed in the absence of said agent, agents which alter said parameter in treated versus untreated cells being identified as modulators of ACSL5 activity.

8. The method of claim 7, wherein said one or more parameters of ACSL5 activity is selected from modulation of one or more of fatty acid synthesis, fatty acid oxidation, glycolysis, insulin production, glucagon production, and glucose uptake.

9. A method for identifying an agent which modulates function of one or more genes selected from *RBM20*, *PDCD4*, *MIR4680*, *BBIP1*, *SHOC2*, *RPL13AP6*, *ADRA2A*, *GPAM*, *TECTB*, *MIR6715B*, *GUCY2GP*, *ZDHHC6*, *VTI1A*, *LOC103344931*, *MIR4295*, *TCF7L2*, *SNORA87A*, *HABP2*, *NRAP*, *CASP7*, comprising

a) incubating cells comprising an enhancer element within a topologically associated domain containing *TCF7L2* and at least one of said genes in the presence and absence of an effective amount of said agent;

b) measuring one or more parameter associated with fatty acid or glucose metabolism in the presence of said agent relative to that observed in the absence of said agent, agents which alter said parameter in treated versus untreated cells being identified as modulators of fatty acid or glucose metabolism.

10. The method of claim 9 wherein said at least one gene is *GUCY2GP*.

11. The method of claim 9, wherein said at least one gene is *CASP7*.

12. The method of claim 9, wherein said at least one gene is *HABP2*.

13. The method of claim 9, wherein said at least one gene is *NRAP*.

14. The method of claim 9 wherein said at least one gene is *TCF7L2*.

15. The method of claim 9, wherein *LOC103344931* is excluded.

16. The method of claim 9, wherein said parameter is selected from modulation of one or more of fatty acid synthesis, fatty acid oxidation, glycolysis, insulin production, glucagon production, and glucose uptake.

FIG. 1A

Upstream

```

1  CGTCCTCTTT GGCTGAAGGT TTTTGCCTGC AACCTTGCAT TGTAATCCAG TGACACCTGA
61 CGTATCTGTA AATTTCTTCA AATTTCTAAG TGTATTACAA CCCCCTGTGC AAAAGATGAT
121 TAATTAATTG CCTTGACAGT AAAACAAAAA AAAAAAAAAA GGTGTGGGGG TATATGGTAT
181 CCCTGATTTA CTATAGAAGA TGCAGAGAGT GAAGGGAGAT GAGGTGGGGA GGAGGGGCCC

```

Downstream

```

1  CTTGGGTTCA AATGATCCTC CTGCCTCAGC CTCTCTTAAG CAGTGCCTTT CCCCATTCTC
61 ATGGGACTTT CCAATCCAIG AGATACTTTG CTGCAGGGAA GCCCTGTCTG TCCAGGCCTG
121 TGTAATAGAC GACTTCACATGGTCCTGTGT TGTGTTTGC CTTCTGTGTG GCTAAGTTTC
181 CATGACCTGG TGGCTTGGAA GCCCCATCCC TGATTTGTGG GAGAGGCAGG GAGGCACCTT

```

Guide sequence

1.ACAGATACGTCAGGTGTCACTGG	6.AGTCGTCTATTACACAGGCCTGG
2.GGATTACAATGCAAGGTTGCAGG	7.GTGTAATAGACGACTTCACAIGG
3.TTAATCATCTTTTGCACACGGGG	8.TGTGAAGTCGTCTATTACACAGG
4.AAGAAATTTACAGATACGTCAGG	9.GTGGCTAAGTTTCCATGACCTGG
5.TGTCACCTGGATTACAATGCAAGG	10.TCCATGAGATACTTTGCTGCAGG

CRISPR/Cas9 PAM sites with the highest scores

FIG. 1B

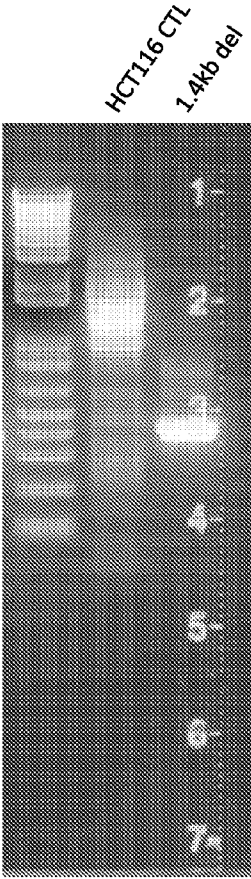


FIG. 1C

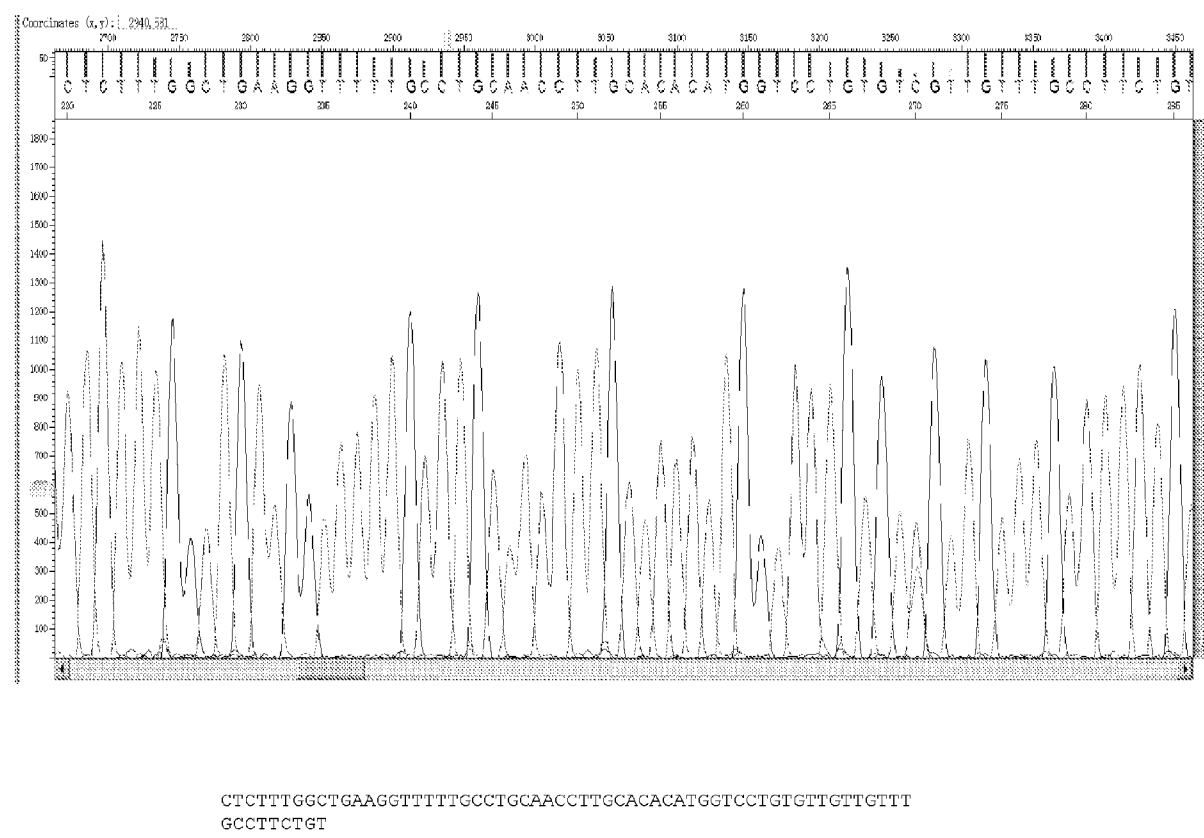


FIG. 1D

■	WT	250	CCCTGGTCTCATCAGCGATATTCGTCCTGCCTTCCTTCTGAGTAATTCCATCTTAGGAC	309
•				
•	D5	44	CCCTGGTCTCATCAGCGATATTCGTCCTGCCTTCCTTCTGAGTAATTCCATCTTAGGAC	103
■	WT	310	TGGAGTCAGGTGGAGCAAGATTCCATCGTTGGTTTCTGTTGGGCCTAGAGTGTCACTG	368
•				
•	D5	104	TGGAGTCAGGTGGAGCAAGATTCCATCGTTGGTTTCTGTTGGGCCTAGAGTGTCACTG	163
■	WT	369	AGACCTAATTCATACTTTATGAATTCTAGTACTGCTCTCGAAGGTAAGAGCCGTCCTCT	428
•				
•	D5	164	AGACCTAATTCATACTTTATGAATTCTAGTACTGCTCTCGAAGGTAAGAGCCGTCCTCT	223
■	WT	429	TTGGCTGAAGGTTTTTGCCTGCAACCTTGCA	459
•				
•	D5	224	TTGGCTGAAGGTTTTTGCCTGCAACCTTGCA	254
■				
■	WT	1816	CACATGGTCCTGTGTTGTTGTTTGCCTTCTGTGTGGCTAAGTTTCCATGACCTGGTGGCT	1875
•				
•	D5	255	CACATGGTCCTGTGTTGTTGTTTGCCTTCTGTGTGGCTAAGTTTCCATGACCTGGTGGCT	314
■	WT	1876	TGGAAGCCCCATCCCTGATTTGTGGGAGAGGCAGGGAGGCACCTTGTAGCGCACTAGGCG	1935
•				
•	D5	315	TGGAAGCCCCATCCCTGATTTGTGGGAGAGGCAGGGAGGCACCTTGTAGCGCACTAGGCG	374
■	WT	1936	TTGGGCCTGAACAAGTCTGTGTGCTTCCAATGTCTTTGTGGGAGGTTACGAGTCCTTC	1995
•				
•	D5	375	TTGGGCCTGAACAAGTCTGTGTGCTTCCAATGTCTTTGTGGGAGGTTACGAGTCCTTC	434
■	WT	1996	TTATTATATAATAGTATCTTGTCTTAGCTTGGTGCC	2031
•				
•	D5	435	TTATTATATAATAGTATCTTGTCTTAGCTTGGTGCC	457

Figure 1E

```

1   TCCAATTTTT TCACATGTGA AGACATACAC AAAAGTTTTA TTGGAGGGTT GCACAIGTGA
61  AAGAAAAAGG GAGAAAGCAG GATTGAGCAG GGGGAGCCGT CAGATGGTAA TGCAGATGTG
121 ATGAGATCTC TGCCGGACCA AAGAGAAGAT TCCTTTTTAA ATGGTGACAA ATTCATGGGC
181 TTTCTCTGCC TCAAAACCTA GCACAGCTGT TATTTACTGA ACAATTAGAG AGCTAAGCAC
241 TTTTITAGATA CTATATAATT TAATTGCCGT ATGAGGCACC CTTAGTTTTT AGACGAGAAA
301 CCACAGTTAC AGGGAAGGCA AGTAACTTAG TCAATGTCAG ATAAC TAGGA AAAGGTTAGA
361 GGGGCCCTGG ACACAGGCCT GTGTGACTGA GAAGCTTGGG CACTTCACTG CTACATTTCA
421 TCTCTTCGCT ATAAACATTT TAGCTTTTTG TGTTTGCTGA CTGGCAACAA TACATAGTGA
481 AAGTTCTAAT AATTTGTAAT G
      rs7903146

```

Guide sequence:

```

11. AACAGCTGTGCTAGGTTTTGAGG 16. ACTTGCCTTCCCTGTAACTGTGG
12. CAGTAAATAACAGCTGTGCTAGG 17. TATAATTTAATTGCCGTATGAGG
13. AAAACTAAGGGTGCCTCATAACG 18. AGACGAGAAACCACAGTTACAGG
14. GTTTCTCGTCTGAAAAC TAAGG 19. GACGAGAAACCACAGTTACAGG
15. GGTTTCTCGTCTGAAAAC TAAGG 20. AGAAACCACAGTTACAGGGAAGG

```

CRISPR/Cas9 PAM sites with the highest scores

FIG. 1F

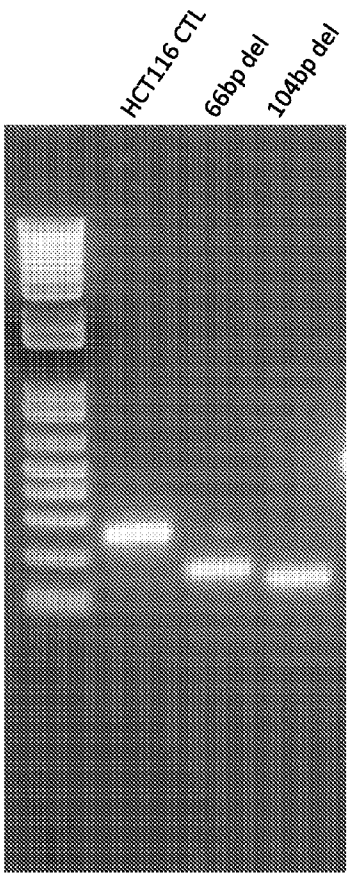


FIG. 1G

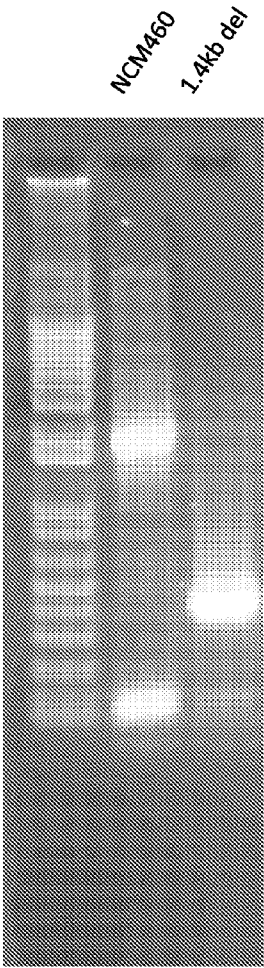


FIGURE 2

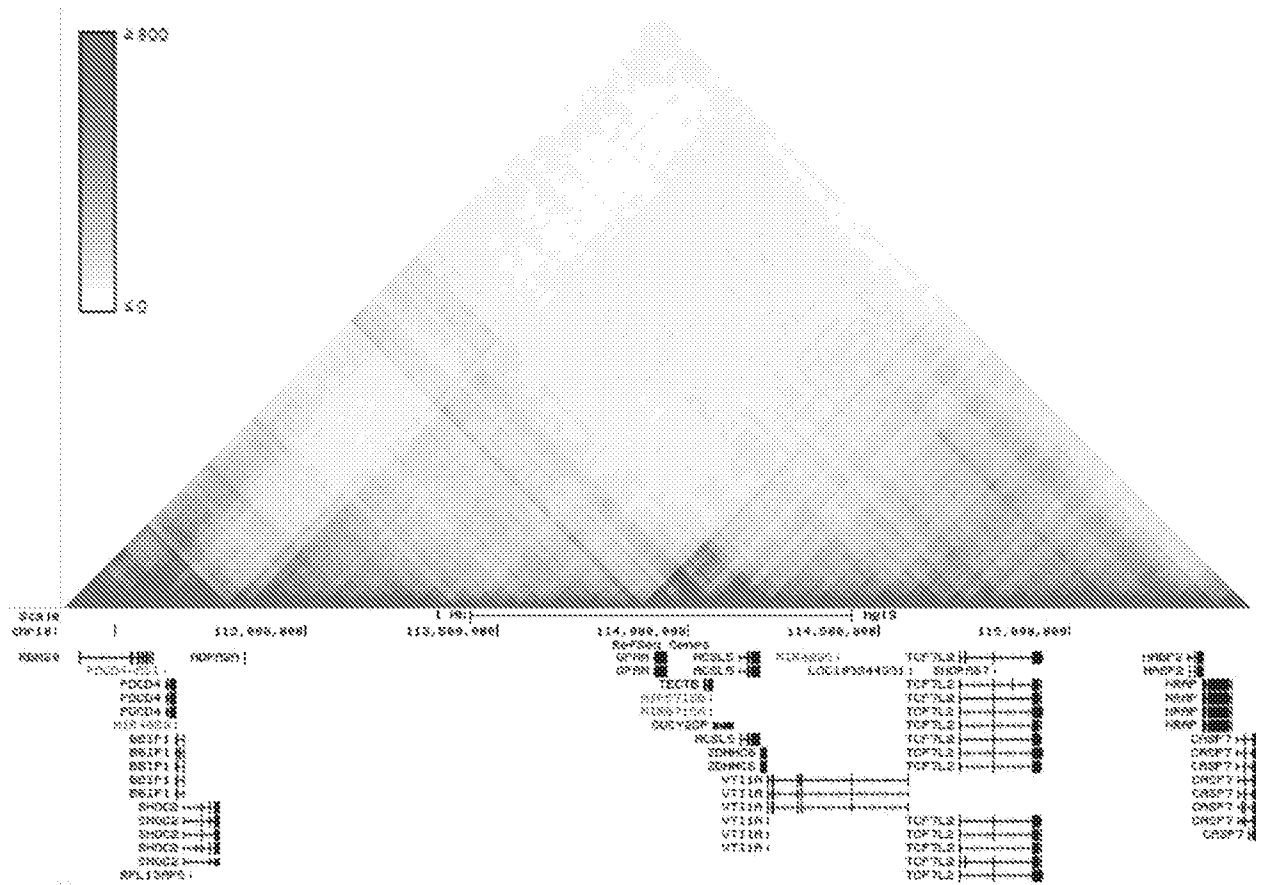
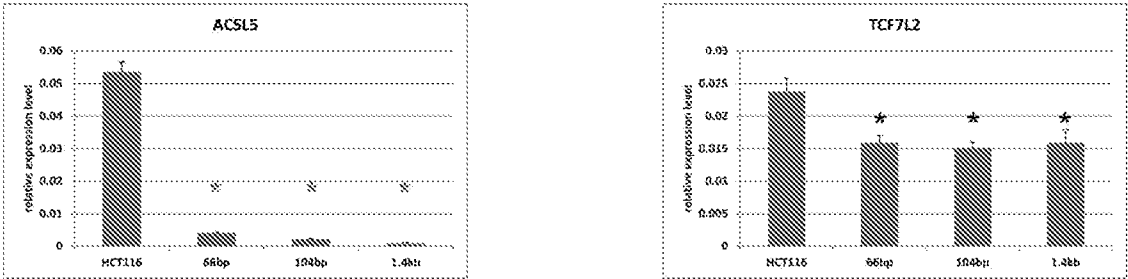


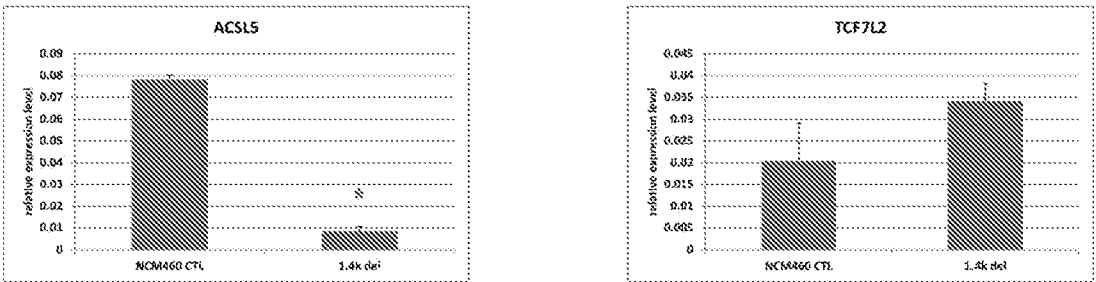
FIG. 3A



* P-value: < 0.001

* P-value: < 0.05

FIG. 3B



* P-value: < 0.001

FIG. 4A

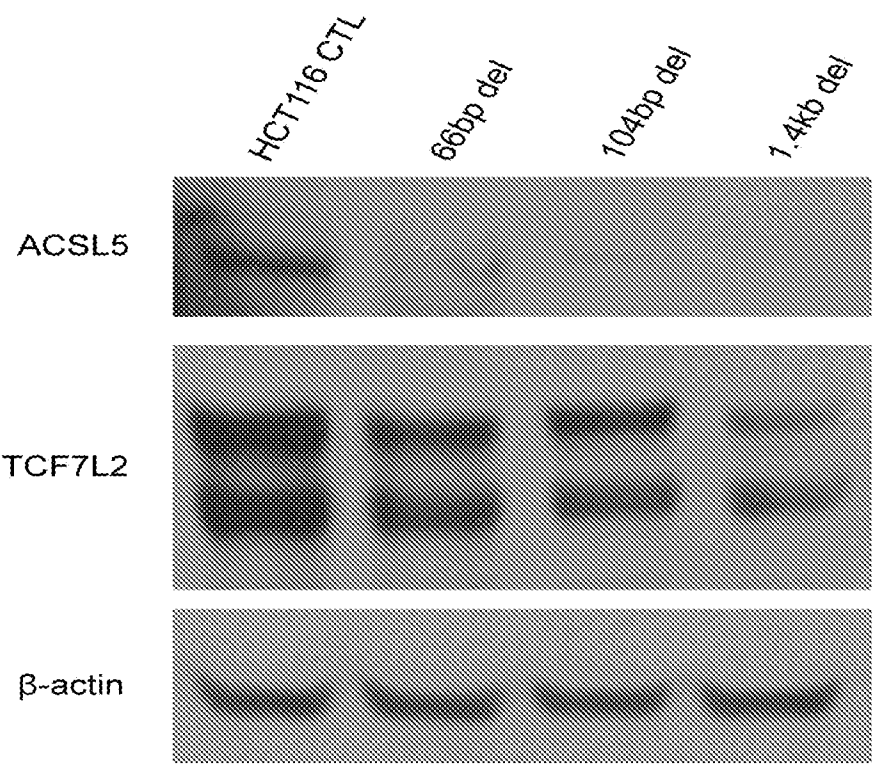
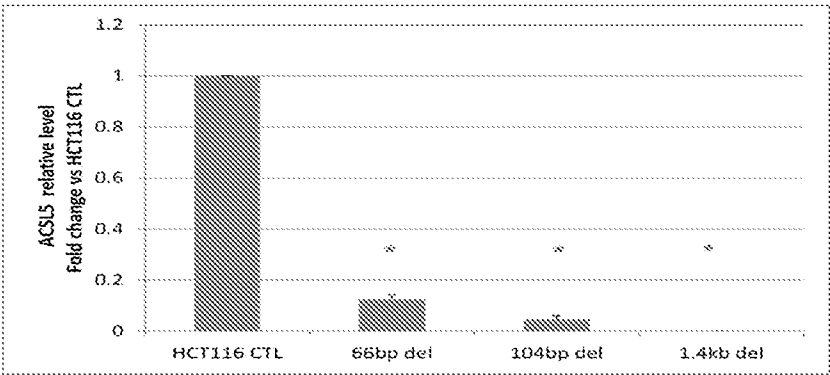
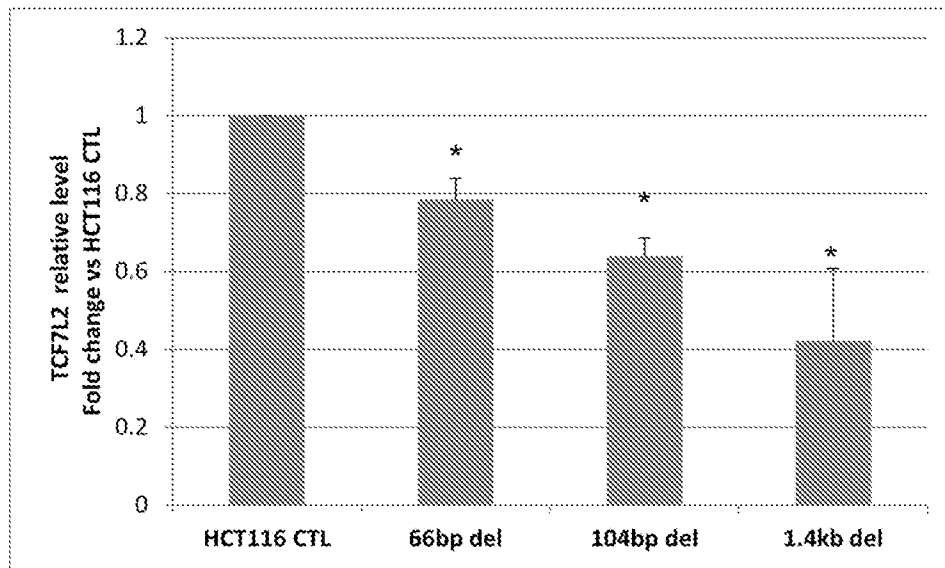


FIG. 4B



* P-value < 0.001

FIG. 4C



* *P*-value: < 0.05

FIGURE 5

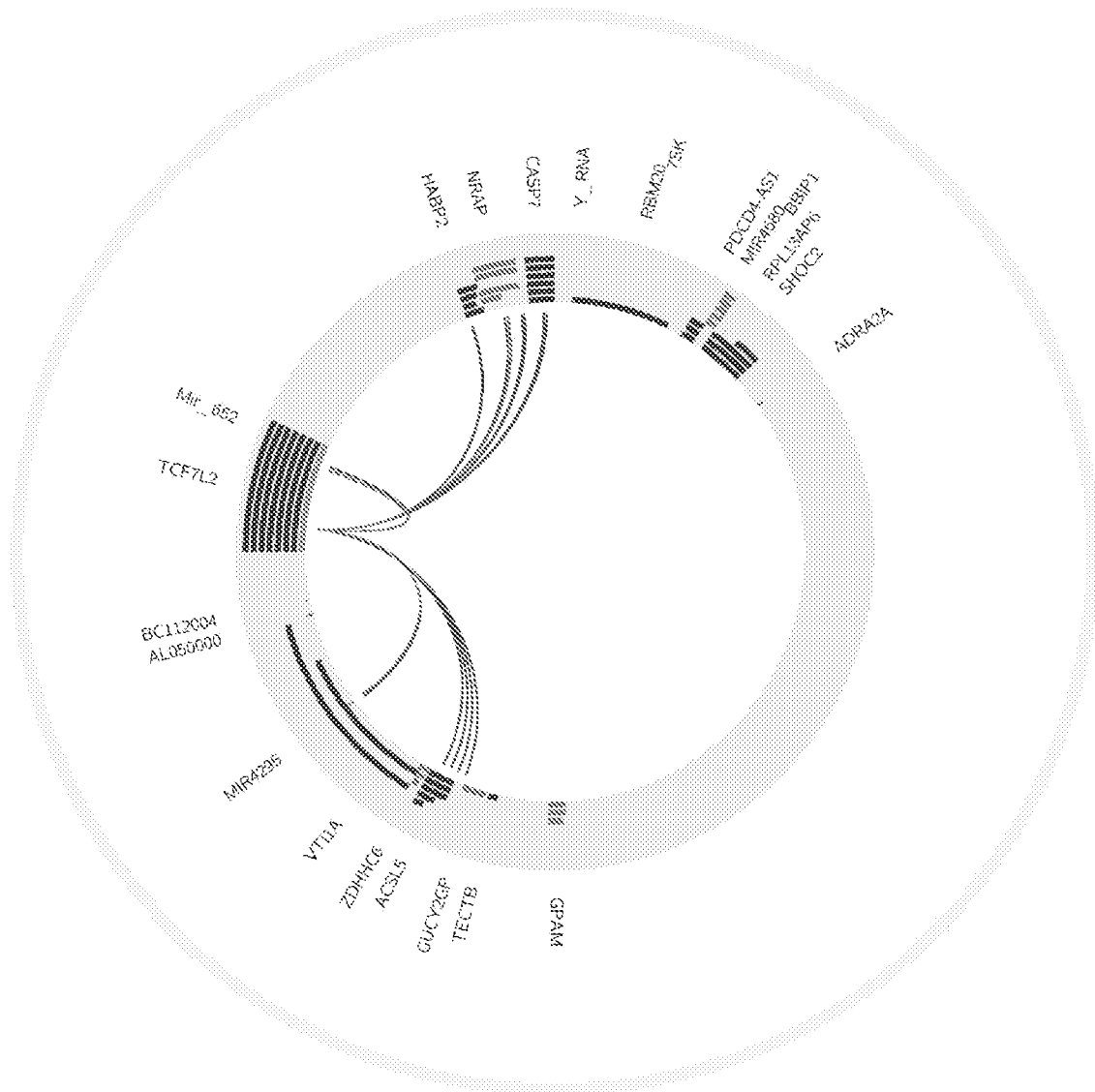


FIG. 6A

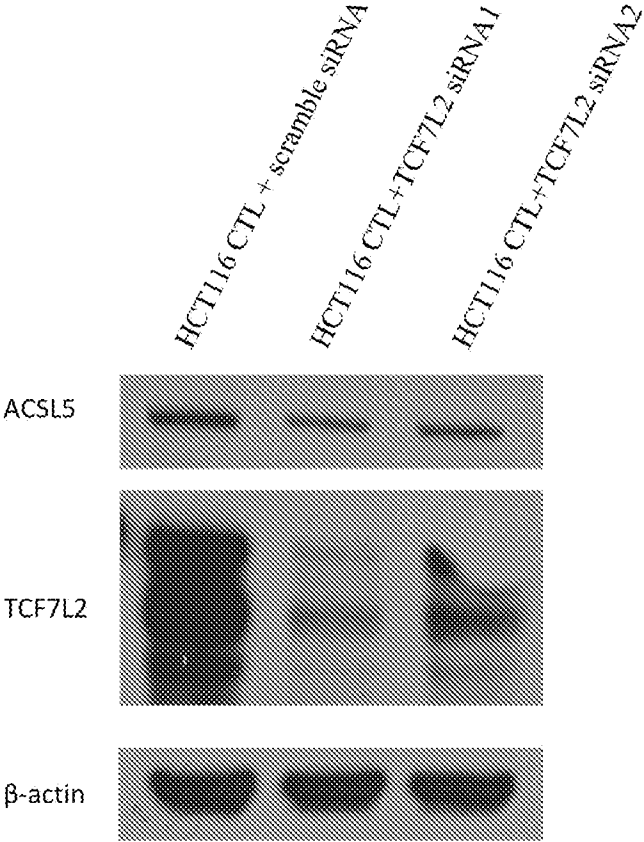


FIG. 6B

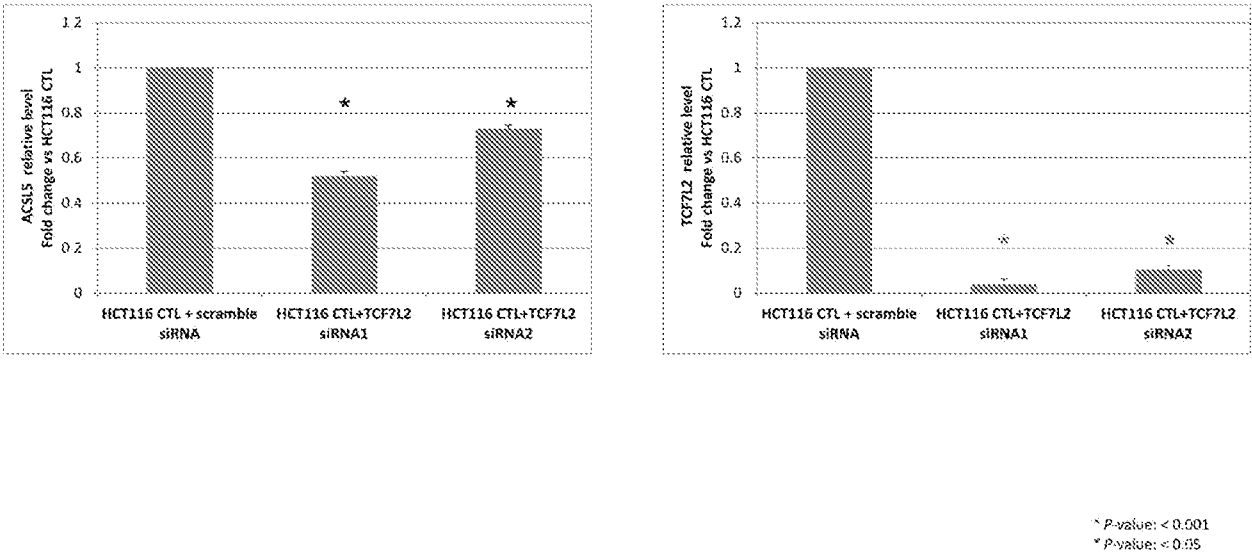


FIG. 7A

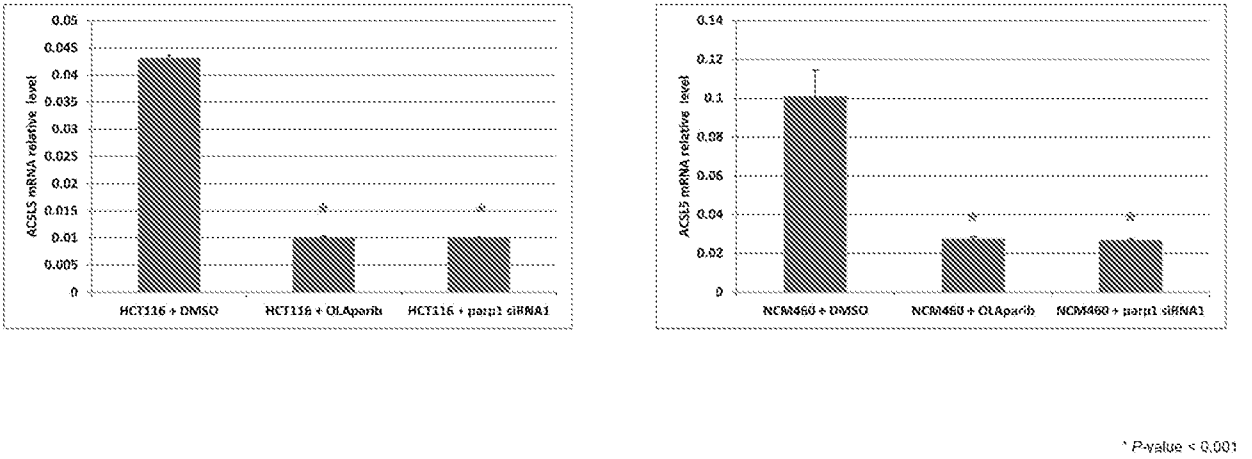


FIG. 7B

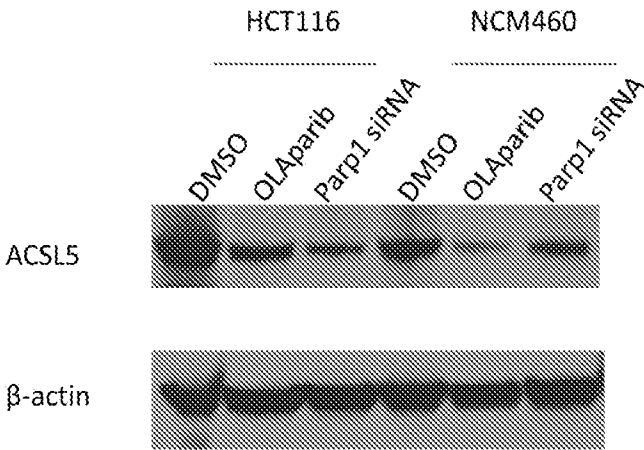
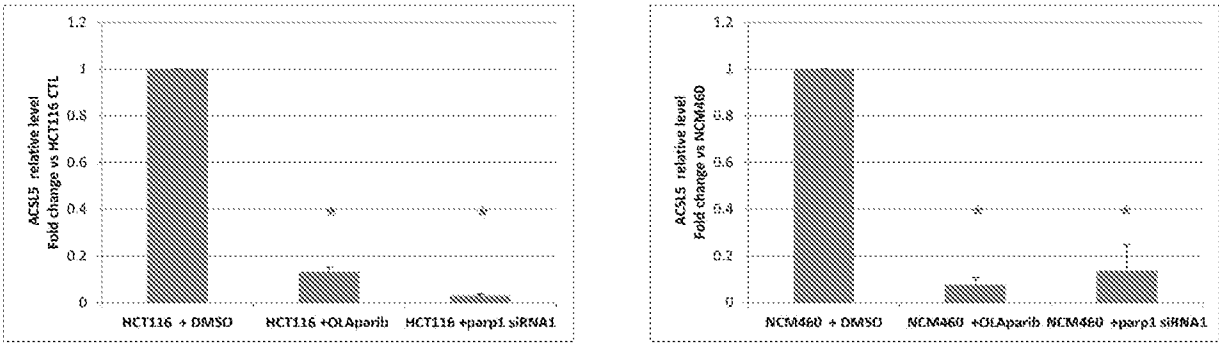
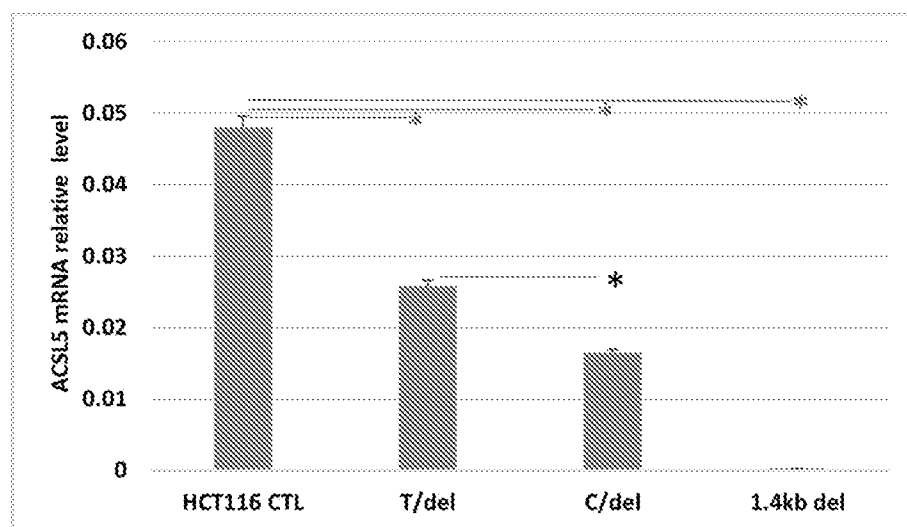


FIG. 7C



* P-value < 0.001

FIG. 8A



* *P*-value: < 0.001

* *P*-value: < 0.01

FIG. 8B

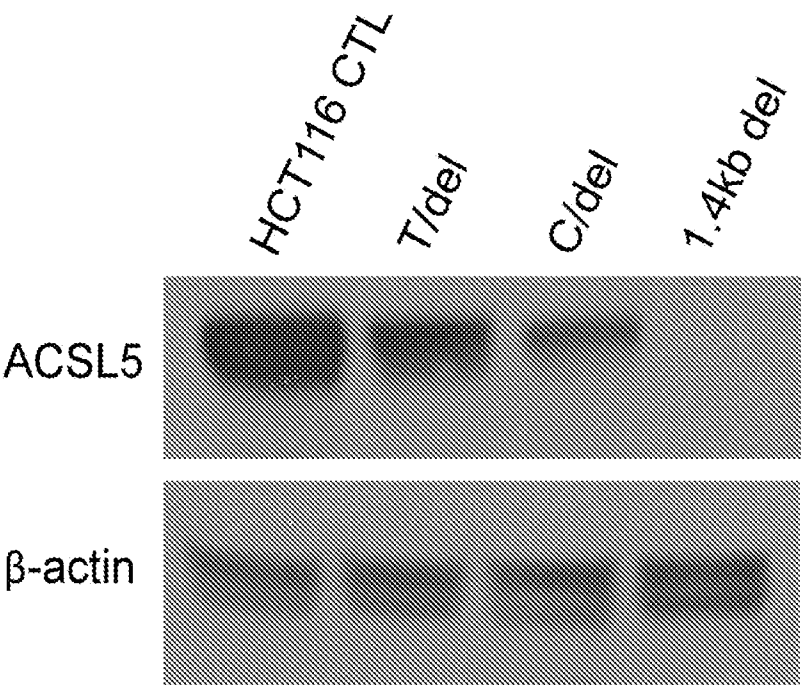
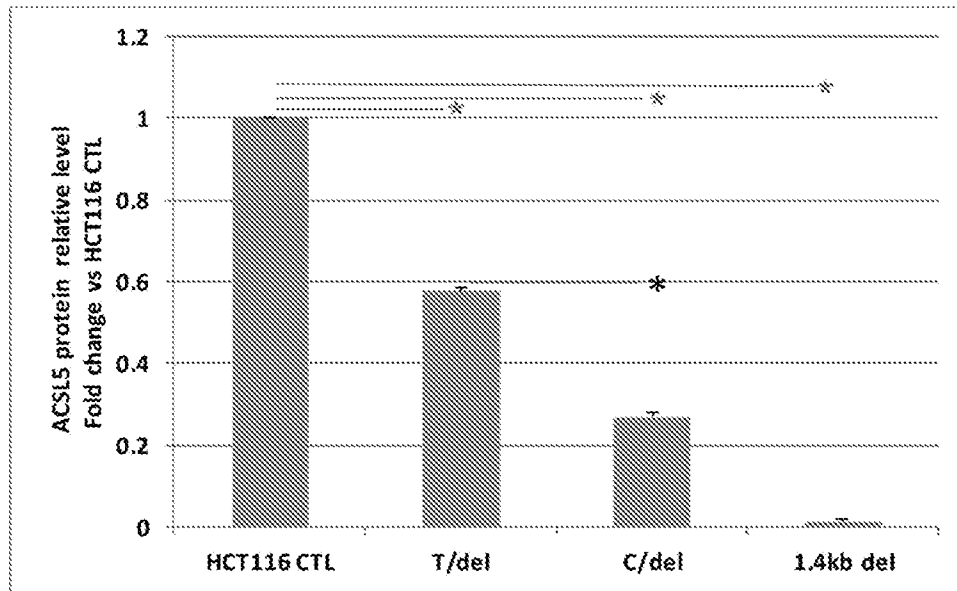


FIG. 8C



* *P*-value: < 0.001

* *P*-value: < 0.01

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2016/037242

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
- a. ☐ forming part of the international application as filed:
☐ in the form of an Annex C/ST.25 text file.
☐ on paper or in the form of an image file.
- b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
- c. ☒ furnished subsequent to the international filing date for the purposes of international search only:
☒ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).
☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).
2. ☒ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2016/037242

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - A61K 31/7088; C12Q 1/68; C40B 30/04 (2016.01)

CPC - A61K 31/7088; C12Q 1/68; C12Q 2600/136; C12Q 2600/158 (2016.08)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC - A61K 31/7088; C12Q 1/68; C40B 30/04

CPC - A61K 31/7088; C12Q 1/68; C12Q 2600/136; C12Q 2600/158

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

USPC - 435/6; 506/9; 514/44R (keyword delimited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

PatBase, Google Patents, Google Scholar, PubMed

Search terms used: TCF7L2 transcription factor 7-like ACSL5 RBM20 PDCD4 MIR4680 BBIP1 SHOC2 RPL13AP6 ADRA2A GPAM TECTB MIR6715B GUCY2GP ZDHHC6 VT11A MIR4295 LOC103344931 SNORA87A HABP2 NRAP CASP7

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JOHNSON et al. "Two novel type 2 diabetes loci revealed through integration of TCF7L2 DNA occupancy and SNP association data," BMJ Open Diabetes Res Care, 01 December 2014 (01.12.2014), Vol. 2:e000052, Pgs. 1-9. entire document	1-16
A	WO 2014/153039 A2 (THE CHILDREN'S HOSPITAL OF PHILADELPHIA) 25 September 2014 (25.09.2014) entire document	1-16
A	US 2012/0149016 A1 (GRANT) 14 June 2012 (14.06.2012) entire document	1-16
A	US 2011/0269688 A1 (HAKONARSON et al) 03 November 2011 (03.11.2011) entire document	1-16
A	XIA et al. "Characterization of the transcriptional machinery bound across the widely presumed type 2 diabetes causal variant, rs7903146, within TCF7L2," Eur J Hum Genet, 26 March 2014 (26.03.2014), Vol. 23, Pgs. 103-9. entire document	1-16



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

23 August 2016

Date of mailing of the international search report

22 SEP 2016

Name and mailing address of the ISA/

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, VA 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774