Diabetologia (2016) 59:2360–2368
DOI 10.1007/s00125-016-4077-2

ARTICLE

The type 2 diabetes presumed causal variant within *TCF7L2* resides in an element that controls the expression of *ACSL5*

Qianghua Xia[1] · Alessandra Chesi[1] · Elisabetta Manduchi[1,2] · Brian T. Johnston[1] · Sumei Lu[1] · Michelle E. Leonard[1] · Ursula W. Parlin[1] · Eric F. Rappaport[3] · Peng Huang[4] · Andrew D. Wells[5] · Gerd A. Blobel[4,6] · Matthew E. Johnson[1] · Struan F. A. Grant[1,6,7]

Received: 29 April 2016 / Accepted: 22 July 2016 / Published online: 18 August 2016

Abstract

Aims/hypothesis One of the most strongly associated type 2 diabetes loci reported to date resides within the *TCF7L2* gene. Previous studies point to the T allele of rs7903146 in intron 3 as the causal variant at this locus. We aimed to identify the actual gene(s) under the influence of this variant.

Methods Using clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein-9 nuclease, we generated a 1.4 kb deletion of the genomic region harbouring rs7903146 in the HCT116 cell line, followed by global gene expression analysis. We then carried out a combination of circularised chromosome conformation capture (4C) and Capture C in cell lines, HCT116 and NCM460 in order to ascertain which promoters of these perturbed genes made consistent physical contact with this genomic region.

Results We observed 99 genes with significant differential expression (false discovery rate [FDR] cut-off:10%) and an effect size of at least twofold. The subsequent promoter contact analyses revealed just one gene, *ACSL5*, which resides in the same topologically associating domain as *TCF7L2*. The generation of additional, smaller deletions (66 bp and 104 bp) comprising rs7903146 showed consistently reduced *ACSL5* mRNA levels across all three deletions of up to 30-fold, with commensurate loss of acyl-CoA synthetase long-chain family member 5 (ACSL5) protein. Notably, the deletion of this single-nucleotide polymorphism region 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 ACSL5, an enzyme with known roles in fatty acid metabolism.

Conclusions/interpretation This 'variant to gene mapping' effort implicates the genomic location harbouring rs7903146 as a regulatory region for *ACSL5*.

Keywords ACSL5 · Chromatin conformation capture · CRISPR · Expression · TCF7L2

Qianghua Xia, Alessandra Chesi, Elisabetta Manduchi, Matthew E. Johnson and Struan F. A. Grant contributed equally to this study.

Electronic supplementary material The online version of this article (doi:10.1007/s00125-016-4077-2) contains peer-reviewed but unedited supplementary material, which is available to authorised users.

✉ Struan F. A. Grant
grants@chop.edu

[1] Divisions of Human Genetics and Endocrinology, The Children's Hospital of Philadelphia, 3615 Civic Center Boulevard, Room 1102D, Philadelphia, PA 19104, USA

[2] Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA, USA

[3] NAPCore, The Children's Hospital of Philadelphia, Philadelphia, PA, USA

[4] Division of Hematology, The Children's Hospital of Philadelphia, Philadelphia, PA, USA

[5] Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[6] Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[7] Institute of Diabetes, Obesity and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Abbreviations

3C	Chromatin conformation capture
4C	Circularised chromosome conformation capture
ACSL5	Acyl-CoA synthetase long-chain family member 5
Cas9	Clustered regularly interspaced short palindromic repeats-associated protein-9 nuclease
CRISPR	Clustered regularly interspaced short palindromic repeats
GWAS	Genome-wide association studies
PARP-1	Poly(ADP-ribose) polymerase 1
SNP	Single-nucleotide polymorphism
TAD	Topologically associating domain
TCF7L2	Transcription factor 7 like 2
TSS	Transcription start site

Introduction

Genome-wide association studies (GWASs) only report genomic signals associated with a given trait and not necessarily the precise localisation of culprit genes. One clear example is the *FTO* locus in obesity. The GWAS signal resides within an intronic region of this gene but in fact primarily influences the expression of nearby *IRX3* and *IRX5* [1, 2], suggesting that this variant is in an enhancer embedded in one gene but influencing the expression of others. The current question in the GWAS field is: how often is this the case with such association signals?

One obvious locus to consider is *TCF7L2* in the context of type 2 diabetes. Common genetic variation located within the gene encoding transcription factor 7 like 2 (TCF7L2) has been consistently reported to be strongly associated with the disease. Such reports range from 2006, when we first published the association [3], to the recent transethnic meta-analysis GWAS of type 2 diabetes [4].

Many studies have been carried out to understand the mechanism by which TCF7L2 itself plays a regulatory role in the pathogenesis of type 2 diabetes mellitus. Given that the culprit tissue for the TCF7L2 effect in this disease has still to be identified, studies have focused on multiple candidate tissues, including pancreatic islets [5], liver [6], adipose tissue [7] and intestinal cells [8]. The latter is particularly compelling given that TCF7L2 and its binding partner, β-catenin, mediate cell-type-specific regulation of the proglucagon (*GCG*) gene, which in turn is cleaved into the insulinotropic hormone glucagon-like peptide 1 (GLP-1) in the intestinal tract [8]. Furthermore, missense mutations in *TCF7L2* have long been known to cause colorectal cancer [9], while studies in *TCF7L2*$^{-/-}$ mice suggest that stem cells in the crypt of the small intestine are genetically programmed by *TCF7L2* [10]. However, a fundamental issue remains, namely that it is still unclear whether *TCF7L2* is in fact the actual principal culprit gene at this locus.

As a consequence of fine mapping efforts in African and African-American individuals, along with Bayesian refinement efforts [11–13], there is now a strong consensus in the field of type 2 diabetes genetics that the T allele of rs7903146 within *TCF7L2* is the causal variant at this locus. Given these observations, one can investigate whether the region harbouring this variant influences the promoter of *TCF7L2* itself or a promoter further away.

To address this, we used genome editing in intestinal-related human cell lines to study the role of the immediate genomic region harbouring rs7903146 in the regulation of gene expression in a controlled, homogenous manner. We assessed the impact of this editing on gene expression via microarrays and then used chromatin conformation capture (3C) to elucidate long-range interactions between this putative enhancer harbouring rs7903146 and the promoters of all genes with perturbed expression in order to identify the gene(s) in physical contact with this genomic element within *TCF7L2*.

Methods

Cell culture

NCM460 cells were purchased from INCELL Corporation (San Antonio, TX, USA) and maintained in M3 medium according to the manufacturer's protocol. The human colorectal cancer cell line HCT116 (ATCC, Manassas, VA, USA) was maintained at 37°C with 5% (vol./vol.) CO_2 incubation in DMEM, supplemented with 10% (vol./vol.) FBS (HyClone, Logan, UT, USA), 100 U/ml penicillin and 100 mg/ml streptomycin. The cells were authenticated and free from mycoplasma contamination.

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

A bicistronic vector pX330, expressing both a human codon-optimised *Streptococcus pyogenes* clustered regularly interspaced short palindromic repeats (CRISPR)-associated protein-9 nuclease (SpCas9) and a chimeric guide RNA, was purchased from Addgene (Plasmid 42230; Cambridge, MA, USA) [14]. The plasmid was linearised with BbsI, dephosphorylated and then gel-purified. We selected ten guide RNA targeting sequences according to the different protospacer adjacent motif (PAM) sites to generate the constructs. These oligonucleotides for targeting the *TCF7L2* rs7903146 region were annealed and ligated to the linearised vector. For a list of guide sequences, see electronic supplementary material (ESM) Fig. 1a. The primers used for molecular cloning are listed in ESM Table 1.

Subsequently, in order to generate 104 bp and 66 bp deletions, we generated single guide RNA (sgRNA)

expressing vectors using a similar method to that described above. We also successfully identified a homozygous 1.4 kb deletion in the NCM460 cells. HCT116 cells are heterozygous for the single-nucleotide polymorphism (SNP) rs7903146, harbouring both the C and T allele. We also identified heterozygous deletion clones, namely C/del and T/del, using a similar approach in HCT116 cells. 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. See further details in ESM Methods sections: 'Targeting the genomic region encompassing rs7903146 in HCT116 cells – 1.4 kb deletion' and 'Subsequent additional deletions'.

Array-based genome-wide expression

Total RNA was isolated from the cell pellets using TRIzol Reagent (Invitrogen, Waltham, MA, USA), according to the manufacturer's protocol, and the RNA concentrations were determined by spectrophotometry (NanoDrop ND-1000, Thermo Fisher Scientific, Waltham, MA, USA). The GeneChip Human Transcriptome Array 2.0 (Affymetrix, Santa Clara, CA, USA) was leveraged at the Genomics Analysis Core at the University of Pennsylvania, Philadelphia, PA, USA. All experiments were carried out in triplicate. See further details in ESM Methods and full results in ESM Table 2. Additionally, microarray data and metadata are deposited in the ArrayExpress repository (available at www.ebi.ac.uk/arrayexpress), with the accession E-MTAB-4839.

Circularised chromosome conformation capture and Capture C protocols

For both circularised chromosome conformation capture (4C) 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). 0.41 g 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 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 (New England Biolabs, Ipswich, MA, USA) to flank the region harbouring rs7903146 near the DpnII and NlaIII cut sites (ESM 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 used the 3C libraries generated (as described above) for the capture procedure. The 3C libraries were sonicated to an average size of 200 bp, ligated to Illumina adaptors and amplified for six cycles with index primers to generate Illumina libraries. Biotinylated DNA oligo probes (140 bp) were designed for each DpnII cut site flanking rs7903146 (ESM Table 1) and were used in the two sequential pulldowns of the Capture libraries. All 4C and Capture C libraries were subsequently sequenced on the Illumina MiSeq platform (Illumina, San Diego, CA, USA).

The sequence analysis pipeline is outlined in the ESM Methods (MiSeq Reporter Software, version 2.6). Chromatin capture data and metadata are deposited in the ArrayExpress repository (available at www.ebi.ac.uk/arrayexpress), with the accessionE-MTAB-4845

RT-PCR and quantitative PCR

Reverse transcription reactions were performed using the High Capacity RNA-to-cDNA Kit (Applied Biosystems, Waltham, MA, USA). The resulting cDNA was diluted 1:10 and used for quantitative PCR. The primer sequences used for *ACSL5* and *TCF7L2* are listed in ESM Table 1. See further details in ESM Methods.

Western blot analysis

Cell pellets were lysed with RIPA buffer. Western blotting was performed according to standard procedures [17]. The mouse primary antibody for acyl-CoA synthetase long-chain family member 5 (ACSL5) was purchased from Abcam (ab104892; Cambridge, MA, USA) and used at a dilution of 1:200. The rabbit antibody for TCF7L2 was purchased from Cell Signaling (no. 2569; Danvers, MA, USA) and used at a dilution of 1:1000. The results were analysed using ImageJ (bundled with 64-bit Java 1.8.0_77; https://imagej.nih.gov/ij/download.html).

Small interfering RNA knockdown

To knock down TCF7L2, the cells were seeded into six-well plates (Corning, Corning, NY, USA) at a density of 1×10^6 cells/well. The cells were transfected via the addition of 100 nmol/l small interfering (siRNA; siGENOME Human TCF7L2 [no. 6934] siRNA – SMARTpool; Dharmacon, Lafayette, CO, USA). *PARP-1* siRNA was purchased from Santa Cruz (sc-29437; Dallas, TX, USA). After 48 h of transfection, the cells were collected for protein extraction. All siRNA experiments used an ON-TARGETplus Non-targeting siRNA #1 (Dharmacon) as a negative control.

Poly(ADP-ribose) polymerase 1 inhibition

The poly(ADP-ribose) polymerase 1 (PARP-1) inhibitor, olaparib (also known as AZD2281, KU0059436) was purchased from Selleckchem (no. S1060; Houston, TX, USA). The HCT116 and NCM460 cells were seeded in six-well plates (Corning) at a density of 1×10^6 cells/well and treated with 50 μmol/l Olaparib in 0.1% DMSO for 48 h.

Statistical analysis

For our array-based genome-wide expression analysis, we leveraged a non-parametric permutation-based approach which corrects for multiple testing based on FDR. We used an FDR cutoff of 10% to call significantly up- and down-regulated genes in HCT116 1.4 kb deletion as compared with wild type. We then filtered the resulting lists of significant genes by a fold change greater than or equal to two (full details in ESM Methods). For follow-up additional data analyses, the two-tailed Student's *t* test was used when two groups were being compared. For 4C and Capture C analyses, we used a non-parametric permutation-based approach that accounts for multiple-testing using FDR, to identify statistically significant enrichment of reads in 5-fragment windows (full details in ESM Methods).

Results

We observed 99 genes with significant differential expression (with a false discovery rate [FDR] cut-off of 10%) and a fold-change of ≥2, of which seven were downregulated and the remaining genes were upregulated in the 1.4 kb deletion setting (ESM Table 3).

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 the promoters of these perturbed genes were making physical contact with the genomic region harbouring the variant. We employed the two techniques in parallel 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 ESM Table 1, the metrics of sequencing and processing in ESM Table 4 and the full contact data in ESM Table 5.

We focused our investigation on significant contacts to promoter regions (within 5 kb of the transcription start site [TSS]) for the 99 genes identified in the microarray analysis, as described above, to ascertain whether 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 1, ESM Fig. 2). We also detected consistent interactions with five additional promoters (*GUCY2GP*, *HABP2*, *NRAP*, *CASP7* and the miRNA, miR-4295), all residing in the same TAD, plus one promoter (*NHLRC2*) outside of the TAD boundaries (Table 1). These genes were not differentially expressed when comparing wild-type vs edited cells (ESM Table 3 and ESM Fig. 3). As for *TCF7L2*, we detected consistent interactions with the promoter region of three short alternative transcripts (Table 1), but this gene was not among the 99 identified in the microarray analyses using 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.

Table 1 Genes showing consistent interactions (within 5 kb of TSS) with the rs7903146 SNP region.

Gene	Transcript	Chromosome	TSS position	Strand
ACSL5	uc001kzs.3	chr10	114133915	+
ACSL5	uc001kzt.3	chr10	114135022	+
ACSL5	uc001kzu.3	chr10	114135955	+
ACSL5	uc009xxz.3[a]	chr10	114154675	+
ACSL5	uc010qrj.2	chr10	114169264	+
CASP7	uc010qsb.3	chr10	115469103	+
GUCY2GP	uc010qri.2	chr10	114067935	−
HABP2	uc001lai.4	chr10	115312777	+
HABP2	uc010qry.1	chr10	115312777	+
HABP2	uc010qrz.1	chr10	115312777	+
HABP2	uc021pyr.1[a]	chr10	115310589	+
miR-4295	uc021pyf.1 [a]	chr10	114393928	+
NHLRC2	uc001lax.2 [a]	chr10	115614390	+
NRAP	uc001laj.4	chr10	115348582	−
NRAP	uc001lak.4	chr10	115348582	−
NRAP	uc001lal.4	chr10	115348582	−
NRAP	uc009xyb.3[a]	chr10	115348582	−
TCF7L2	uc010qrv.2 [a]	chr10	114886572	+
TCF7L2	uc010qrw.2 [a]	chr10	114889059	+
TCF7L2	uc010qrx.2 [a]	chr10	114889391	+

Data is from five chromatin capture experiments (4C and Capture C in NCM460 and HCT116 cells, plus 4C using a different bait primer set in NCM460 cells)

Gene models are University of California, Santa Cruz (UCSC) genes, coordinates refer to the GRCh37/hg19 build

ACSL5 was the only gene identified as differentially expressed in HCT116 cells harbouring a deletion of the SNP region

[a] Interaction confirmed in one 4C experiment in human colon tissue

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

In NCM460 cells, the use of CRISPR/ CRISPR-associated protein-9 nuclease (Cas9) proved relatively challenging. However, we were able to generate a clone with a 1.4 kb deletion. The resulting cells grew extremely slowly but sufficient RNA was obtained for analysis of *ACSL5* and *TCF7L2* expression. We observed the same marked decrease in *ASCL5* gene expression (Fig. 1c), but a slight, albeit non-significant, increase in the gene expression of *TCF7L2* (Fig. 1d), contrary to the direction we observed in HCT116 cells.

Given these observations, we also assessed the production of ACSL5 and TCF7L2 proteins in the HCT116 cells. Similar to real-time mRNA observations, we saw a dramatic decrease in ACSL5 protein levels, with almost total ablation with the largest sized deletion (Fig. 2a, b). With respect to TCF7L2, we also observed a reduction in its production but not as marked as for ACSL5 (Fig. 2a, c).

To investigate whether deletion of the genomic element harbouring rs7903146 had an impact on the long-range contacts observed, we performed 4C using the HCT116 cell line harbouring the 1.4 kb 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 for establishing this long-range interaction (for an overview, see Fig. 3, also including other genes in the TAD with consistent promoter contacts to the SNP region). Additionally, all contacts for *CASP7*, *GUCY2GP* and *HABP2*, plus some contacts for *NRAP* and *TCF7L2*, were abolished by the CRISPR deletion, while the contact for the miR-4295 miRNA was not abolished; full data are available in ESM Table 5).

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 (ESM Fig. 4a–c), suggesting that these observations are largely specific to the enhancer region and are not controlled by TCF7L2.

In addition, we previously reported that a factor called PARP-1, plus its partners, binds across this region [18]. As such, in order to investigate a possible impact on ACSL5 levels in a comparable fashion to that 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 impact at both the RNA (Fig. 4a, b) and the protein (Fig. 4c–e) level in both NCM460 and HCT116.

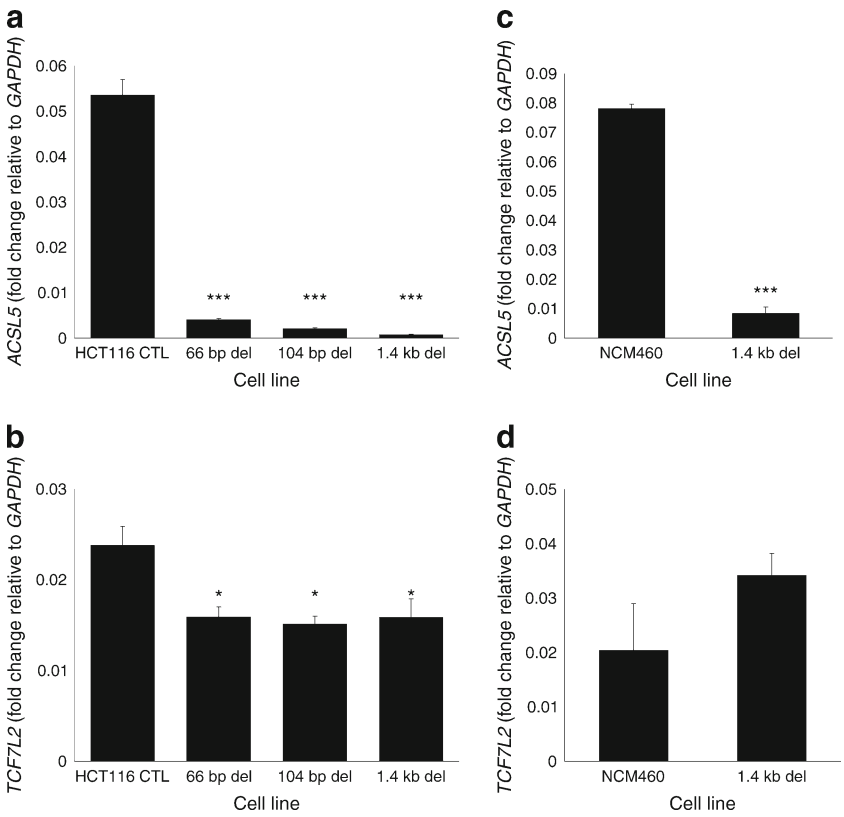

Fig. 1 Expression analysis of *ACSL5* and *TCF7L2* following gene editing of the genomic region harbouring rs7903146. Relative levels of individual gene expression in the control and targeted cells, determined by quantitative PCR with *GADPH* normalisation. (**a**) *ACSL5* and (**b**) *TCF7L2* expression in the HCT116 control and targeted cells. (**c**) *ACSL5* expression in the NCM460 control and targeted cells. (**d**) *TCF7L2* expression in the NCM460 control and targeted cells. CTL, control; del, deletion. Values represent the mean of three experiments ± SDs. *$p < 0.05$, ***$p < 0.001$

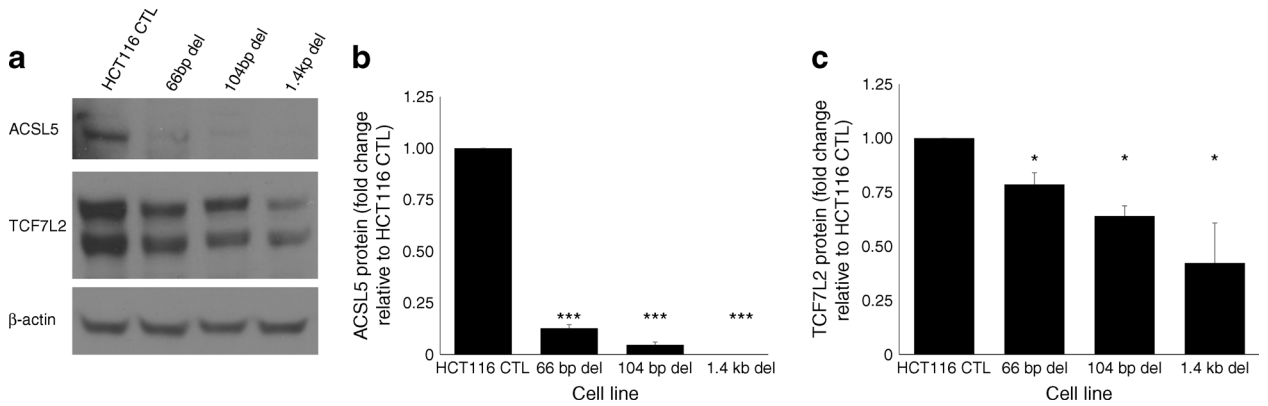

Fig. 2 Western blot analysis of ACSL5 and TCF7L2 following the CRISPR/Cas9 deletion of the region harbouring *TCF7L2* SNP rs7903146. (**a**) Levels of ACSL5 and TCF7L2 proteins in the control and targeted cells determined by western blotting, using β-actin as a loading control. A representative blot from three experiments is displayed. (**b,c**) Quantification of the three independent experiments described in (**a**) represented as bar graphs. Western blots were scanned and intensities were determined using ImageJ. Bars represent quantified western blot signals normalised to β-actin levels and HCT116 control cells. CTL, control; del, deletion. Values represent the mean of three experiments ± SD. *$p < 0.05$, ***$p < 0.001$

To investigate allelic differences, and leveraging the fact that HCT116 is heterozygous for rs7903146, we used CRISPR/Cas9 to generate clones exclusively harbouring the C or the T allele. 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 Fig. 5a, deletion of either the C or T allele reduced the expression of *ACSL5* at the mRNA level. Interestingly, the T allele clone revealed slightly stronger (but statistically significant, $p < 0.001$) 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 (Fig. 5b, c). These results are compatible with the T allele driving a stronger enhancer for *ACSL5* expression than the C allele.

Finally, to validate our results from cell line experiments, we performed 4C in frozen human colon tissue (ESM

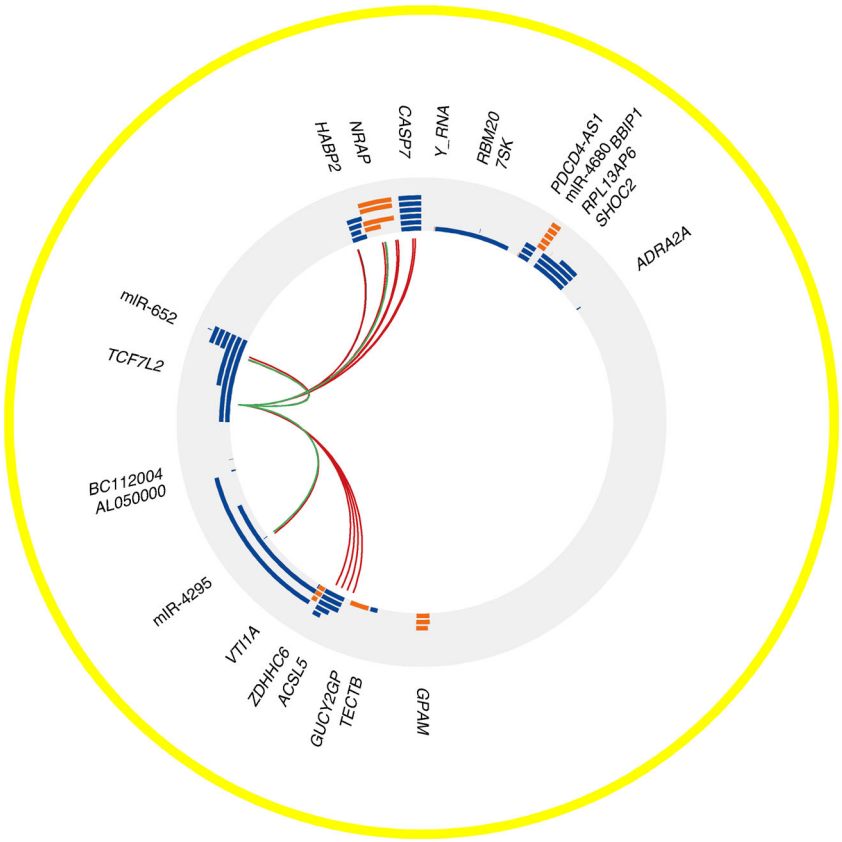

Fig. 3 Circos plot of significant contacts. The yellow circle represents the *TCF7L2* TAD region (expanded to fully contain *CASP7*; hg19 coordinates chromosome 10:112,370,010–115,490,668) with coordinates increasing clockwise starting (and ending) at 'noon' (position between *CASP7* and *Y_RNA* in the figure). For each gene, transcripts (from the University of California, Santa Cruz gene models) are depicted (without the exon/intron details) in blue on the forward strand or in orange on the reverse strand (*TCF7L2* has 30 UCSC transcripts, some hidden for ease of view). Links between the bait regions and significant 4C contacts within 5 kb of any TSS are displayed for HTC116 (red) and for HCT116 1.4 kb deletion (green). Contacts shown are only those which were also consistently detected in the HTC116 capture C, NCM460 4C and Capture C assays

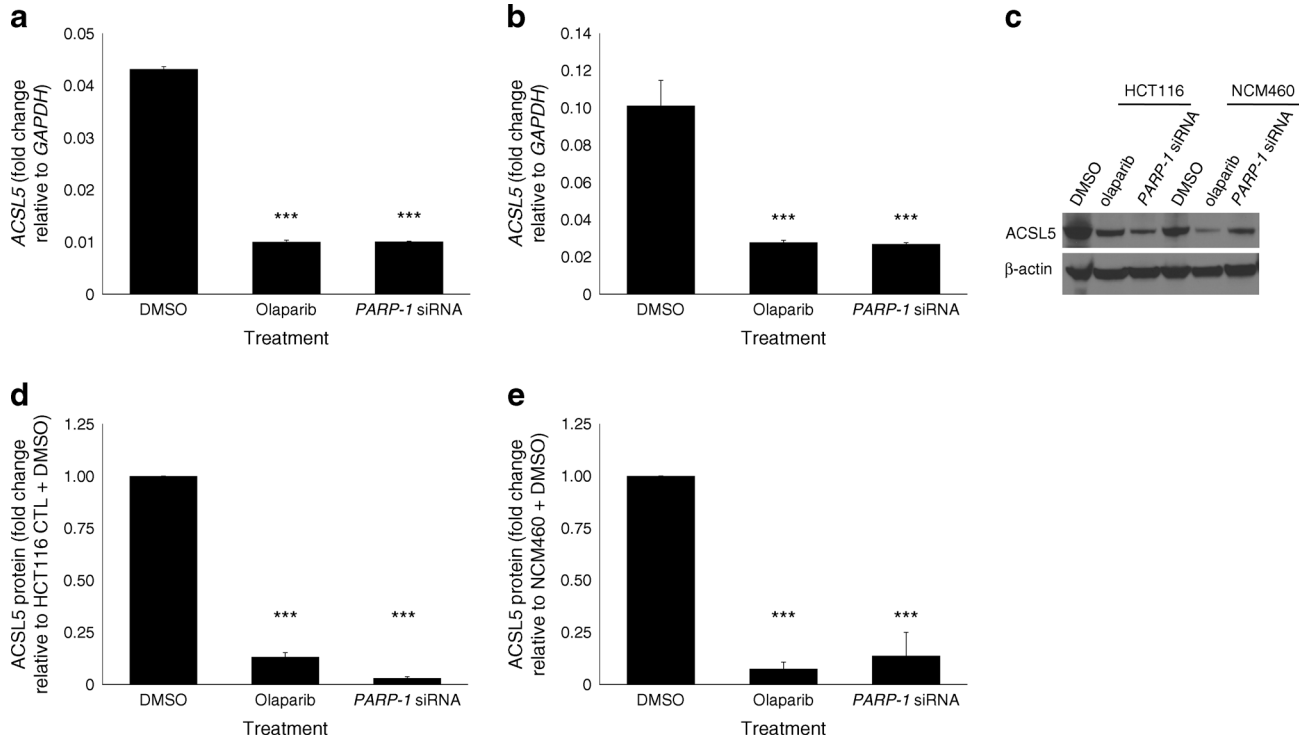

Fig. 4 PARP-1 inhibition by olaparib and siRNA. Relative levels of *ACSL5* expression in control cells (treated with DMSO) and treated (**a**) HCT116 and (**b**) NCM460 cells were determined by quantitative PCR with *GAPDH* normalisation. (**c**) ACSL5 protein levels were determined by western blotting in control cells and cells treated with the PARP-1 inhibitor, olaparib, or *PARP-1* siRNA, using β-actin as a loading control. A representative blot from three experiments is displayed. (**d,e**) Quantification of the three independent experiments described in (**c**) represented as a bar graph; (**d**), HCT116 cells and (**e**) NCM460 cells. Western blots were scanned and intensities were determined using ImageJ. Bars represent quantified western blot signals normalised to β-actin levels and HCT116/ NCM460 control cells. Values represent the mean of three experiments ± SDs. ***$p < 0.001$

Table 5). In this setting, consistent with 3C experiments in cell lines, we detected interactions between the rs7903146 SNP region and one *ACSL5* promoter (specific for a short alternative transcript), plus contacts with the *HABP2*, miR-4295, *NHLRC2*, *NRAP* and *TCF7L2* promoters (Table 1). The specific *ACSL5* promoter usage in vivo (in colon and other tissues) warrants further investigation.

Discussion

We observed that combining data resulting from the deletion of the immediate genomic region harbouring rs7903146 and the characterisation of promoter contacts strongly implicated *ACSL5* at this type 2 diabetes-associated locus. Furthermore, it resides in the same

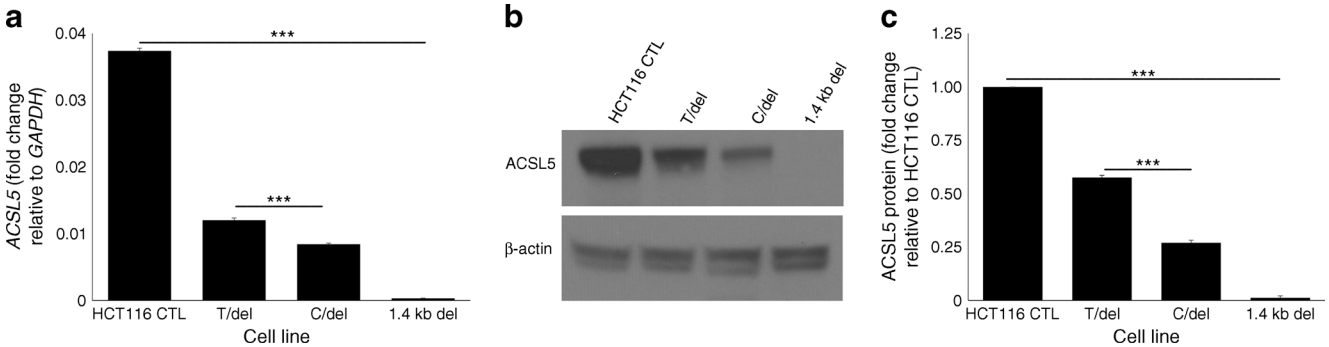

Fig. 5 Allele-specific effects on *ACSL5* gene expression. (**a**) Relative levels of *ACSL5* gene expression in control and targeted cells, determined by quantitative PCR with *GAPDH* normalisation. (**b**) Levels of ACSL5 protein in the control and targeted cells was determined by western blotting, using β-actin as a loading control. A representative blot from three experiments is displayed. (**c**) Quantification of the three independent experiments described in (**b**) represented as a bar graph. Western blots were scanned and intensities were determined using ImageJ. CTL, control; del, deletion. Bars represent quantified western blot signals normalised to β-actin levels and HCT116 control cells. Values represent the mean of three experiments ± SDs. ***$p < 0.001$

TAD as *TCF7L2*. This gene encodes an acyl-coA synthetase essential for fatty acid metabolism.

We additionally observed an impact on *TCF7L2* expression with the deletion approach, albeit relatively modest when contrasted with the consistent *ACSL5* observations across all deletions. However, results were disparate in two different cell lines, leading to uncertainty about the physiological relevance of this observation.

Furthermore, we observed some evidence for other promoters in the TAD showing consistent contact with the genomic region harbouring the variant, namely *HABP2, GUCY2GP*, *CASP7, NRAP, NHLRC2* and miR-4295. However, aside from miR-4295, when their contacts with the rs7903146 SNP region were abolished in the deletion cell line, an enhancer role of the SNP region for these genes was not supported by our expression data when comparing wild-type vs edited cells (ESM Table 3 and ESM Fig. 3). Importantly, most of the contacts detected by 3C experiments in the intestinal-related cell lines were conserved in human colon tissue, including *ACSL5* (Table 1).

The Roadmap Epigenomics Project reports epigenetic marks for rs7903146 compatible with enhancer activity in several tissues, including colon, small intestine, adipose cells and pancreatic islets, but no expression quantitative trait loci (eQTLs) have been published for this variant to date. This is likely to be because of lack of power, with limited sample sizes for many tissues, and indeed this was part of our motivation to undertake these approaches.

It should be noted that the primary focus of our CRISPR/Cas9 work involved editing the immediate genomic neighbourhood surrounding the key SNP. The rationale was that if a putative enhancer were removed, this would have a substantial impact on the expression of one or more genes and 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 (although 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 lead to a relatively modest increase in risk of type 2 diabetes.

Given the approach we employed, which required the intersection of CRISPR and 3C data, our focus was very much directed to *ACSL5*. The fact that subsequent different sized deletions yielded almost the same impact on its expression indicates that such expression changes are unlikely to be due to off-target effects.

The genomic sequence harbouring rs7903146, therefore, appears to act as an enhancer given that, when we knock out the region, ACSL5 levels are reduced. Further evidence for this comes from a previous study of human pancreatic islets, where the type 2 diabetes T risk allele was found to be enriched in open chromatin [19]. It should be noted that the immediate genomic region flanking rs7903146 is not conserved in mice, suggesting mechanistic differences between species and that a rodent model may not be optimal for investigations of the question at hand.

There is an ongoing intense discussion in the field about which tissue(s) this locus principally acts on to confer its effect on the risk of type 2 diabetes. Despite the colon-related cell lines used in this study not strictly representing the L cell (which is a limitation of the current study), in the setting we elected to work in we can observe a strong relationship between the genomic location harbouring rs7903146 and the expression of *ACSL5* with this 'variant to gene mapping' effort. Testing this in other type 2 diabetes candidate tissues is an area for future investigation. In fact, one must be aware that chromatin accessibility is tissue specific and that there have been reports in which *TCF7L2* expression in pancreatic islets runs counter to that in other tissues [20]. However, our results show that, in the cell setting we elected to use for this study, the rs7903146 region operates in a similar manner to the *FTO* locus in obesity [1, 2], where the expression of neighbouring genes is influenced by an apparent embedded enhancer in a given gene, in this case *TCF7L2*.

It is known that the *TCF7L2* locus correlates with metrics of insulin secretion and/or processing. However, other studies report an influence on insulin resistance in adipose tissue [7] and glucose clearance in the liver [6]. One could speculate that the reason this is one of the strongest loci reported to date is that it resides in a crucial region involved in multiple roles, directing different genetic influences in different tissues in a coordinated fashion.

Our data also lead us to conclude that *TCF7L2* could also play a role in the pathogenesis of type 2 diabetes. Note that although *TCF7L2* is known to have multiple isoforms, our expression data revealed no significant differences in these splice variants (ESM Table 6).

Inhibiting ACSL5 enzyme activity could be a promising avenue for improving insulin sensitivity, given that changes in NEFA levels can lead to insulin resistance. There have been two independent studies related to this in ACSL5 knockout 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 a high-fat diet [21]. In contrast, a more recent study showed that ACSL5 ablation in mice increased energy expenditure and insulin sensitivity and delayed fat absorption [22]. Further investigation is warranted to assess the effect of ACSL5 activity on insulin resistance given the observations in this current study.

In conclusion, the rs7903146 region appears to interact with the promoters of *ACSL5* and *TCF7L2*, among others. Deletion of this region in two human colon cell lines

dramatically decreased the expression of ACSL5. We also observed some effect on TCF7L2 levels, but not as strong as for ACSL5. PARP-1 inhibition also affects ACSL5 levels, although the specificity of this action is still to be elucidated. Due to the exquisite tissue specificity of enhancer–promoter interactions, it is conceivable that other target genes for which interactions were observed in our chromatin conformation experiments might be relevant in other tissues since this region might indeed be a locus control region. Given the important role of ACSL5 in lipid biosynthesis and fatty acid degradation, it is tempting to speculate that insulin sensitivity could be substantially improved by inhibiting ACSL5 enzyme activity.

Acknowledgements Some of the data were presented as an abstract at the American Society of Human Genetics annual meeting in 2015 and as an abstract at the American Diabetes Association annual meeting in 2016.

Funding This work was financially supported by Pennsylvania Department of Health award SAP4100062202, CHOP's Spatial and Functional Genomics Initiative and the Daniel B. Burke Endowed Chair for Diabetes Research.

Duality of interest The authors declare that there is no duality of interest associated with this manuscript.

Contribution statement Study concept and design: QX, AC, EM, BTJ, MEJ, and SFAG. Acquisition of data: QX, BTJ, SL, MEL, UWP, EFR, MEJ. Analysis and interpretation of data: QX, AC, EM, PH, ADW, GAB, MEJ and SFAG. Drafting and critical revision of the manuscript: QX, AC, EM, BTJ, SL, MEL, UWP, EFR, PH, ADW, GAB, MEJ and SFAG. Bioinformatics analyses: AC, EM and MEJ. Obtaining funding: ADW, GAB and SFAG. All authors contributed to and approved the final version of the manuscript. QX, AC, EM, MEJ and SFAG are the guarantors of this work and as such had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

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 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. Nat Genet 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 2 diabetes susceptibility. Nat Genet 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. J Clin Invest 117:2155–2163
6. Boj SF, van Es JH, Huch M et al (2012) Diabetes risk gene and Wnt effector 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. J Biol Chem 280:1457–1464
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 mutational analysis in colorectal cancer cell lines. Cancer Res 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. Nat Genet 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. Nat Genet 39:218–225
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. Nat Genet 44:1294–1301
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. Nat Genet 46:205–212
16. van de Werken HJ, Landan G, Holwerda SJ et al (2012) Robust 4C-seq data analysis to screen for regulatory DNA interactions. Nat Methods 9:969–972
17. Mahmood T, Yang PC (2012) Western blot: technique, theory, and trouble shooting. N Am J Med Sci 4:429–434
18. 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
19. Gaulton KJ, Nammo T, Pasquali L et al (2010) A map of open chromatin in human pancreatic islets. Nat Genet 42:255–259
20. 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 Res 21:1417–1425
21. 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 Health Dis 12:88
22. 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. Mol Metab 5:210–220