npg

ARTICLE

Characterization of the transcriptional machinery bound across the widely presumed type 2 diabetes causal variant, rs7903146, within *TCF7L2*

Qianghua Xia[1], Sandra Deliard[1], Chao-Xing Yuan[2], Matthew E Johnson[1] and Struan FA Grant*,[1,3,4]

Resolving the underlying functional mechanism to a given genetic association has proven extremely challenging. However, the strongest associated type 2 diabetes (T2D) locus reported to date, *TCF7L2*, presents an opportunity for translational analyses, as many studies in multiple ethnicities strongly point to SNP rs7903146 in intron 3 as being the causal variant within this gene. We carried out oligo pull-down combined with mass spectrophotometry (MS) to elucidate the specific transcriptional machinery across this SNP using protein extracts from HCT116 cells. We observed that poly (ADP-ribose) polymerase 1 (PARP-1) is by far the most abundant binding factor. Pursuing the possibility of a feedback mechanism, we observed that PARP-1, along with the next most abundant binding proteins, DNA topoisomerase I and ATP-dependent RNA helicase A, dimerize with the TCF7L2 protein and with each other. We uncovered further evidence of a feedback mechanism using a luciferase reporter approach, including observing expression differences between alleles for rs7903146. We also found that there was an allelic difference in the MS results for proteins with less abundant binding, namely X-ray repair cross-complementing 5 and RPA/p70. Our results point to a protein complex binding across rs7903146 within *TCF7L2* and suggests a possible mechanism by which this locus confers its T2D risk.

European Journal of Human Genetics (2015) 23, 103–109; doi:10.1038/ejhg.2014.48; published online 26 March 2014

INTRODUCTION

In recent years, genome-wide association studies (GWAS) have identified many genes associated with type 2 diabetes (T2D).[1–5] The most strongly associated T2D locus in multiple ethnicities resides within the transcription factor 7-like 2 (*TCF7L2*) gene.[6–8]

TCF7L2 (formerly known as TCF4) is a member of high-mobility group box-containing transcription factors and operates at the last key stage and is therefore one of the main effectors of the canonical Wnt signaling transduction cascade,[9] regulating the expression of downstream target genes. In the absence of β-catenin, TCF7L2 binds to Wnt-responsive elements to repress target gene transcription while β-catenin binding to TCF7L2 activates gene expression.

How genetic variation in *TCF7L2* influences the risk of T2D has remained elusive and the primary site(s) of action for the gene product in the context of the disease is still unclear. It was initially speculated that TCF7L2 operated in tandem with insulin to influence blood glucose homeostasis through the alteration of levels of glucagon-like peptide 1 in the gut,[6,10] with knockout mouse models suggesting that TCF7L2 has an indispensable role in intestinal epithelium development.[11] However, other studies have shown that the *TCF7L2* variant is associated with increased *TCF7L2* expression and decreased insulin secretion, possibly implicating the pancreatic β-cell,[12,13] although an indirect effect on insulin secretion by TCF7L2 action in another tissue cannot be excluded, including adipose[14] and liver.[15]

Resolving the underlying functional mechanism to a given genetic association in the post-GWAS era has proven extremely challenging. However, the discovery of the *TCF7L2* locus in the context of T2D presents a specific opportunity for translational analyses, as studies in multiple ethnicities[16,17] and with Bayesian modeling[18] have now strongly implicated the intron 3 SNP rs7903146 (NG_012631.1:g.53341C>T) as the causal variant within this gene.

We therefore hypothesized that protein factors which bind to the immediate intronic region harboring rs7903146:C>T modulate TCF7L2 function and thus have impact further downstream where it exerts its effect. We elected to carry out oligo pull-down combined with mass spectrophotometry (MS) to elucidate the transcriptional machinery across this intronic SNP.

MATERIALS AND METHODS

Cell culture and nuclear extracts preparation

Human HCT 116 cells, where TCF7L2 is abundantly expressed, were cultured in Dulbecco's Modified Eagle Medium (DMEM; 4.5 g/l glucose, 10% FCS, 100 U/ml penicillin and 100 μg/ml streptomycin). For the SILAC experiments, cells were grown according to the standard culture procedures in SILAC-light and SILAC-heavy labeled DMEM (ThermoFisher, Waltham, MA, USA) supplemented with 10% dialyzed fetal bovine serum, $^{13}C6$-lysine and $^{13}C6$-arginine. Cells were washed with cold phosphate-buffered saline then harvested by centrifugation for 5 min at 5000 rpm at 4 °C. In all, 10^8 cells were lysed in 250 μl low salt buffer (10 mM HEPES, 1.5 mM MgCl₂, 10 mM KCl, 0.5% NP-40, proteasome inhibitor, phosphatase inhibitor, 10% glycerol, pH 7.9) by incubation on ice for 20 min, then vortexed for 10 s per 5 min. The nuclei were

[1]Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; [2]Department of Proteomics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; [3]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; [4]Institute of Diabetes, Obesity and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
*Correspondence: Dr SFA Grant, Division of Human Genetics, Children's Hospital of Philadelphia Research Center, Room 1216, 3615 Civic Center Boulevard, Philadelphia, PA 19104, USA. Tel: +1 267 426 2795; Fax: +1 267 426 0363; E-mail: grants@chop.edu

separated from the cytoplasmic fraction by centrifugation at 10 000 rpm for 1 min at 4 °C. The nuclei pellets were washed once with cold low salt buffer and harvested by centrifugation. The nuclear proteins were then released into high salt buffer (20 mM HEPES, 1.5 mM MgCl$_2$, 420 mM NaCl, 0.2 mM EDTA, proteasome inhibitor, phosphatase inhibitor, 25% glycerol, pH 7.9), and the supernatant was subsequently collected through centrifugation at 12 000 rpm for 10 min at 4 °C.

Oligonucleotide pull-down

The 5' Dual Biotin modified oligonucleotides for rs7903146 were synthesized by Integrated DNA Technologies, Inc. (Coralville, IA, USA) The oligonucleotide sequences are as follows:

C allele:

For: 5'-ACAATTAGAGAGCTAAGCACTTTTTAGATACTATATAATTTAATT GCCGTATGAGGCACCC-3'

Rev: 5'-GGGTGCCTCATACGGCAATTAAATTATATAGTATCTAAAAAGTG CTTAGCTCTCTAATTGT-3'

T allele:

For: 5'-ACAATTAGAGAGCTAAGCACTTTTTAGATATTATATAATTTAATT GCCGTATGAGGCACCC-3'

Rev: 5'-GGGTGCCTCATACGGCAATTAAATTATATAATATCTAAAAAGTG CTTAGCTCTCTAATTGT-3'

In all, 10 μm of forward and reverse oligonucleotides were annealed before use. Four picomole of the oligonucleotides were then combined with 1 mg of nuclear extract in binding buffer (20 mM HEPES, 1.5 mM MgCl2, 150 mM NaCl, 0.2 mM EDTA, proteasome inhibitor, phosphatase inhibitor, 25% glycerol, pH 7.9) in a total volume of 1 ml. These solutions were incubated for 1 h at 4 °C on a rotator. A total of 60 μl of streptavidin was added to each tube, and the mixture was then further incubated for 30 min at 4 °C on a rotator. The beads were spun for 30 s at 800 g and then washed in 5 × 500 μl binding buffer. The supernatant was discarded, and proteins were eluted by boiling in SDS sample buffer. The eluates were then isolated via SDS-PAGE gel followed by Coomassie blue R-250 staining. The bands of interest were subsequently excised from the gel and digested by trypsin for MS.

MS

The sample was digested with trypsin and analyzed with nanoLC/MS/MS at the University of Pennsylvania Proteomics Core. The data were analyzed with Sequest (ThermoFinnigan, San Jose, CA, USA; version SRF v. 5). Scaffold (version Scaffold_3.6.0, Proteome Software Inc., Portland, OR, USA) was used to validate MS/MS-based peptide and protein identifications. For the SILAC experiments, Zoom Scan was added to the MS analytical method to monitor the labeled and unlabeled peptides. The quantification of the SILAC peaks was performed with ProteoIQ with customized modification so that the software could use the zoom-scanned SLICA peaks for quantification.

Immunoprecipitation and western blotting

Co-immunoprecipitation was performed using IP buffer (50 mM tris, 150 mM NaCl, proteasome inhibitor, phosphatase inhibitor, 10% glycerol, pH 8.0) on a rotator at 4 °C overnight. Western blotting was performed according to the standard procedures. The following antibodies were used: anti-TCF7L2 (Millipore, Billerica, MA, USA; 05–511), anti-PARP-1 (Cell Signaling, Danvers, MA, USA; 46D11), anti-DNA topoisomerase 1 (Abcam, Cambridge, MA, USA; Ab28432), and anti-RNA helicase A (Abcam, Ab26271).

Cloning, transfection and luciferase assay

The 501-bp fragments harboring rs7903146 were successfully amplified by PCR using genomic DNA from HCT116 cells. Both the A and T alleles were cloned in to the pGL4.26 vector (Promega, Fitchburg, WI, USA) in the forward strand orientation and also in the reverse orientation using restriction enzymes, *KpnI* and *XhoI*. The sequences of the oligonucleotides were as follows:

Forward 5' KPN I: 5'-cccGGTACC TCCAATTTTTTCACATGTGAAGACA TACAC-3'

501 bp Forward 3' Xho I: 5'-cccCTCGAG CATTACAAATTATTAGAACTTT CACTATGTATTG-3'

Reverse 5' KPN I: 5'-cccGGTACC CATTACAAATTATTAGAACTTTCAC TATGTATTG-3'

Reverse 3' Xho I: 5'-cccCTCGAG TCCAATTTTTTCACATGTGAAGACA TACAC-3'

All constructs were verified by sequencing. HEK-293T cells, which are typically used due to the general ease of transfection, were seeded at 2×10^5 cells per well in opaque 48-well plates 24 h before transfection. Each well was transfected with 100 ng of the PGL4.26-Luc firefly reporter construct. Firefly luciferase constructs harboring each genomic element were then co-transfected with an internal control vector, pGL4.74 (hRluc/TK, Promega), plus 100 ng TCF7L2 and/or 100 ng β-catenin. β-Catenin is a key component of the experimental design as it is a well-established and key mediator of Wnt signaling via activation of TCF7L2. These transient co-transfections were carried out in four independent transfection experiments. Twenty-four hours after transfection, the activities of firefly and renilla luciferase were measured using the Dual-Luciferase Assay Kit (Promega), according to the manufacturer's instructions. Luciferase activity ratio (firefly/renilla) for each transfection condition was normalized to the ratio obtained with the control vector.

RESULTS

PARP-1 and other peptides interact specifically with an oligonucleotide harboring rs7903146

We used the combination of oligo pull-down and MS in triplicate to identify the transcriptional machinery interacting with the genomic element within the third intron of *TCF7L2* harboring the SNP rs7903146. Nuclear lysates from HCT116 cells, where TCF7L2 is abundantly expressed, were incubated with biotin-labeled, double-stranded 60-bp oligonucleotides spanning this SNP. The DNA–protein complexes were precipitated with streptavidin-agarose beads, and the bound proteins were isolated by denaturing SDS-PAGE, followed by staining with Coomassie blue R-250.
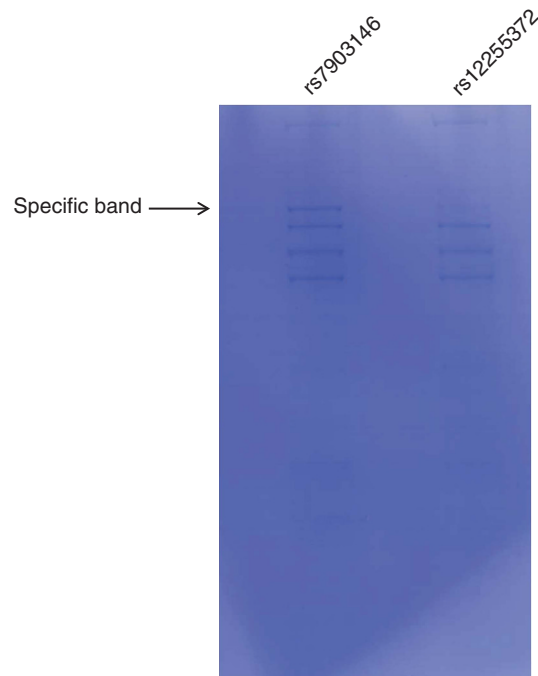

Figure 1 Oligo pull-down for protein identification. Proteins from nuclear lysates of HCT116 cells (2 mg total protein each) that bind to biotin-labeled, double-stranded oligonucleotides, stained with staining with Coomassie blue R-250. The protein band unique in the rs7903146 oligo pull-down, as compared with the non-functional proxy SNP, rs12255372, is identified by the arrow. The specific band in this run was also confirmed by MS to be the same protein complex as seen in previous runs.

As shown in Figure 1, several bands were visible in the pull-down samples. An oligo coinciding with a SNP in strong linkage disequilibrium with rs7903146 in Caucasians but not in other ethnicities and widely rejected as the causal variant,[19] namely rs12255372, did not yield this extra band to any extent. Furthermore, this particular band was also negligible with a scrambled oligo (Supplementary Figure S1).

This specific band was cut from the gel, digested with trypsin and submitted for LC-MS/MS analysis. We set a cutoff to $N = 10$ for the number of identified peptides to be considered. The identified proteins from the initial run above the cutoff are listed in Table 1, in order of peptide abundance, with PARP-1 showing the greatest abundance by far, followed by DNA topoisomerase I and ATP-dependent RNA helicase A; comparable results from subsequent runs are shown in Supplementary Table S1.

rs7903146-associated self-regulation of TCF7L2

As PARP-1, DNA topoisomerase I and ATP-dependent RNA helicase A were identified as the most abundant proteins binding at this location, we attempted to characterize their interaction with each other and with TCF7L2 at the protein level to understand how these four factors may be involved in TCF7L2 regulation. Informed by a previous study, where it was suggested that the PARP-1 and TCF7L2

Table 1 Identity of the most abundant proteins determined by mass spectrophotometry

Name	Total spectra
Poly (ADP-ribose) polymerase 1	108
cDNA FLJ53442, highly similar to poly (ADP-ribose) polymerase 1	106
DNA topoisomerase 1	42
ATP-dependent RNA helicase A	39
Isoform Beta of DNA ligase 3	29
Isoform Alpha of DNA ligase 3	29
FACT complex subunit SPT16	19
Thyroid hormone receptor-associated protein 3	18
Isoform 1 of Nucleolar RNA helicase 2	18
Isoform 2 of Nucleolar RNA helicase 2	18
Nucleolin	17
cDNA FLJ45706 fis, clone FEBRA2028457, highly similar to Nucleolin	16
DNA damage-binding protein 1	16
Isoform 1 of Splicing factor 3B subunit 3	13
Isoform 1 of Myb-binding protein 1A	12
Isoform 2 of Myb-binding protein 1A	12
NCL protein	12
X-ray repair cross-complementing protein 5	12
Putative uncharacterized protein NOP2	11
NOP2 protein	11
Isoform 2 of Putative ribosomal RNA methyltransferase NOP2	11
32 kDa protein	11
cDNA FLJ51067, highly similar to DNA damage-binding protein 1	11
116 kDa U5 small nuclear ribonucleoprotein component	10
116 kDa U5 small nuclear ribonucleoprotein component isoform b	10
Myosin-Id	10
Isoform 1 of Putative ATP-dependent RNA helicase DHX30	10
Isoform 3 of Putative ATP-dependent RNA helicase DHX30	10
Isoform 2 of Putative ATP-dependent RNA helicase DHX30	10
Putative uncharacterized protein DHX30	10
Isoform 1 of Aspartyl/asparaginyl beta-hydroxylase	10
Aspartyl/asparaginyl beta-hydroxylase isoform f	10

proteins physically interact,[20] the obvious hypothesis was that these main proteins interact to form a complex bound across the rs7903146 region.

In order to test this hypothesis, we performed co-immunoprecipitation experiments. As shown in Figure 2, these four factors exhibited interaction with each other. This supports the notion that TCF7L2, PARP-1, DNA topoisomerase I and ATP-dependent RNA helicase A form a complex to regulate TCF7L2 itself.

To test for further evidence of a feedback mechanism for TCF7L2 expression operating at this rs7903146 location, we cloned the genetic elements in both orientations upstream of the luciferase reporter gene in PGL4.26. Promoter activity was measured in 293T cells. Transfections were carried out in quadruplicate.

TCF7L2 lies at the end of the Wnt signaling pathway and is thus the effector for this cascade but of course needs to be activated first before it can exert its effect on the set of genes it regulates. Therefore, only overexpressing TCF7L2 will not be sufficient in our reporter system. β-Catenin is the key mediator to this effect and lies immediately upstream of TCF7L2, where it travels to the nucleus to physically interact with TCF7L2 to activate downstream gene expression thus its overexpression being necessary for our luciferase experiments. As such, and mirroring this point, we observed that overexpression of TCF7L2 or β-catenin alone is not sufficient to activate transcription; however, the combination of TCF7L2 and β-catenin overexpression produced a striking increase of upto fivefold in transcription levels (Figure 3). Furthermore, at this stage we also observed differences between rs7903146 alleles (Figure 3).

'X-ray repair cross-complementing protein 5' (Ku80) and 'replication protein A 70 kDa DNA-binding subunit' preferentially bind to the T allele of rs7903146

We investigated whether there was allele-specific preferential binding for some of the proteins detected across rs7903146 in order to gain further insight into the regulation of *TCF7L2*. To test our hypothesis, we performed 'two-way' oligo pull-down experiments. The nuclear extracts from the 'light' cells that were isotopically labeled in SILAC medium were used in the C allele oligo pull-down while the nuclear extracts from the 'heavy' cells that were isotopically labeled in SILAC medium were used for the pull-down of the T allele oligo.
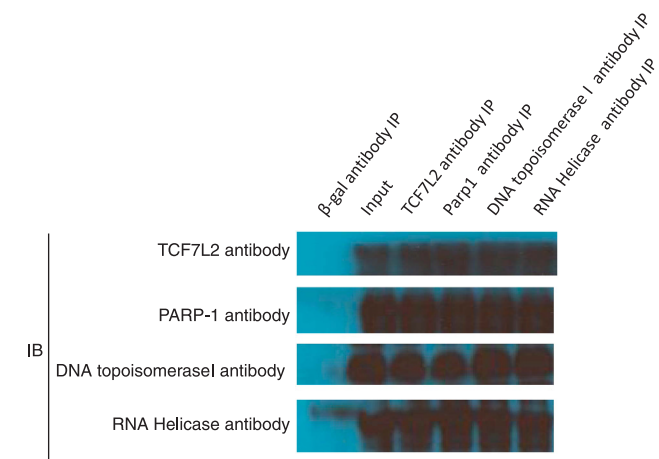

Figure 2 HCT116 nuclear extracts were used for co-immunoprecipitation to determine protein interaction. TCF7L2, PARP-1, DNA topoisomerase I and RNA helicase A interact with each other and form a complex. The first lane represents the negative control.
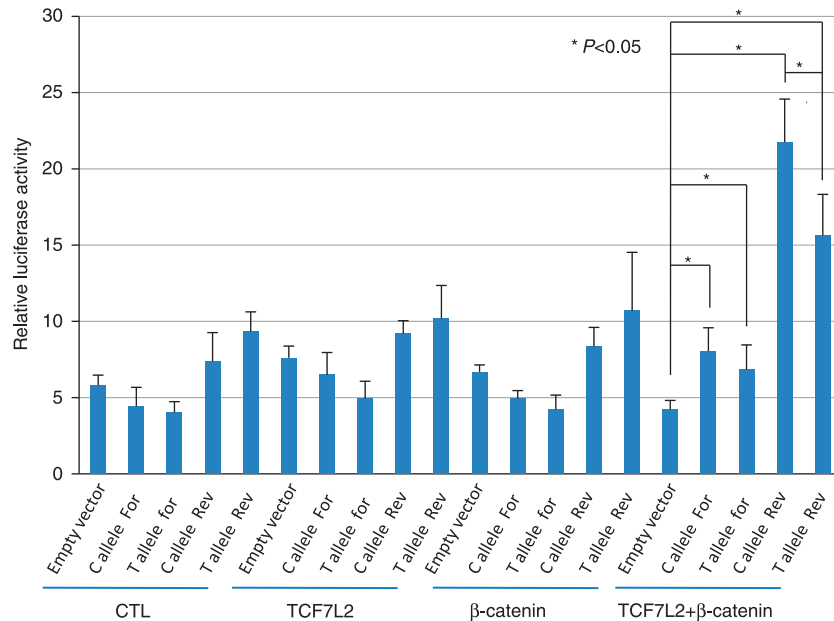
Figure 3 Outcome of firefly luciferase constructs harboring each genomic element being co-transfected with an internal control vector, plus TCF7L2 and/or β-catenin, into 293T cells. Twenty thousand 293T cells were plated in per well of 48-well dish in 0.5 ml of complete growth medium 24 h before transfection. In all, 100 ng of firefly luciferase constructs harboring each genomic element were co-transfected with 1 ng internal control vector hRluc/TK, plus 100 ng TCF7L2 or/and 100 ng β-catenin, into each well. PGL4.26 empty vector was used as the control for the basal level of luciferase activity. Dual Luciferase assay was performed 24 h after transfection, and promoter activity values are expressed as arbitrary units using a hRluc/TK reporter for internal normalization. Experiments were carried out in quadruplicate, and the SD is indicated. Significant differences in expression levels are indicated.

The SILAC-labeled extracts were then switched for a reverse experiment in order to assess the reliability and reproducibility of the approach. Interestingly, we identified two less abundant proteins, X-ray repair cross-complementing protein 5 (XRCC5; also known as Ku80) and replication protein A 70 kDa DNA-binding subunit, that were preferentially binding to the T allele over the C allele (Figure 4).

DISCUSSION

There have been many efforts to resolve the underlying causative mechanism for a given GWAS signal for a complex trait. However, in the vast majority of situations, there still remains only a list of candidate variants that could represent the causal event. The situation is somewhat more advanced for the T2D locus, *TCF7L2*, where rs7903146 is widely regarded as the causal variant through a logical process of elimination leveraging multiple ethnicities and novel statistical approaches.[16–18]

Indeed, when we first reported the association of *TCF7L2* with T2D, we noted that rs12255372 and rs7903146 both captured the association well,[6] but subsequent studies in other ethnicities observed rs12255372 was a less optimal tag-SNP and revealed rs7903146 to be clearly the best SNP to test across multiple populations.[16,17] As such, rs12255372 served as good control in this current study.

As this variant does not represent a coding variant residing in an exon, rather it resides in an intronic region, it is reasonable to presume it is involved in a regulatory process. As such, one obvious tactic to attempt to resolve its function is to determine whether a transcriptional complex binds specifically at this location, using a method such a oligo pull-down followed by detection with MS.

Through the use of this technique, we have identified PARP-1 binding across the immediate region harboring this variant. It is well established that PARP-1 has a role in DNA damage detection and repair, and recent studies have revealed important roles for PARP-1 in chromatin and transcriptional regulation.[21,22] PARP-1 is the focus of

many key oncology programs within the pharmaceutical industry, so one intriguing question raised by our findings is: will PARP-1 inhibitors also alleviate the symptoms of T2D and could rs7903146 dictate dose response? Clearly much more work is required before this can be resolved, but there are already intriguing clues in the literature to support this notion. For example, PARP activation has been shown to be present in healthy subjects at risk of developing diabetes, as well as in established type 2 diabetics.[23] PARP-1 inhibitors have also been shown to alleviate symptoms of diabetic complications.[24–26] Furthermore, PARP-deficient mice are protected from streptozotocin-induced diabetes.[27]

With respect to the connection with oncology, missense mutations in *TCF7L2* have been known for many years to be strongly linked with colorectal cancer risk.[28,29] Indeed, this connection intensified following reports that the 8q24 locus revealed by GWAS of a number of cancers, including colorectal carcinomas, was due to an extreme upstream TCF7L2-binding element driving the transcription of *MYC*.[30–32] In addition, it has been shown that when *TCF7L2* recurrently fuses with its neighboring gene, *VTI1A*, colorectal adenocarcinomas result. Curiously, many of the other GWAS-implicated loci associated with T2D are also associated with the risk for prostate cancer.[33] Although there are apparent genetic links between cancer and T2D, the mechanism is still not well understood, but the fact that we propose a role of PARP-1 in the regulation of TCF7L2 may shed some light on this striking picture emerging from GWAS findings.

Our results suggest that TCF7L2 may be also regulated by other proteins in this PARP-1 dominated complex, including DNA topoisomerase 1 and ATP-dependent RNA helicase A. DNA topoisomerase 1 is known to interact with DNA intermediates and proteins of base excision repair;[34] indeed, poly(ADP-ribose) reactivates stalled DNA topoisomerase 1 and induces DNA strand break resealing.[35] RNA helicases can unwind double-stranded DNA and RNA in a 3′ to

a

Name	Peptides	Control Ratio	H/L ratio (Log 2)	Fold change (T/C)
X-ray repair cross-complementing protein 5	60	0	0.727562458	1.655839062
Replication protein A 70 kDa DNA-binding subunit	14	0	0.433199236	1.350224438

XRCC5: GITEQQKEGLEIVK

RP-A p70 : SGGVGGSNTNWK

Scan: 3377 Retention time: 17.9587

Scan: 1616 Retention time: 11.488

b

Name	Peptides	Control Ratio	H/L ratio (Log 2)	Fold change (T/C)
X-ray repair cross-complementing protein 5	52	0	-0.637174506	0.642970964
Replication protein A 70 kDa DNA-binding subunit	13	0	-0.670010915	0.628501932

XRCC5: GITEQQKEGLEIVK

RP-A p70 : SGGVGGSNTNWK

Scan: 3291 Retention time: 17.8378

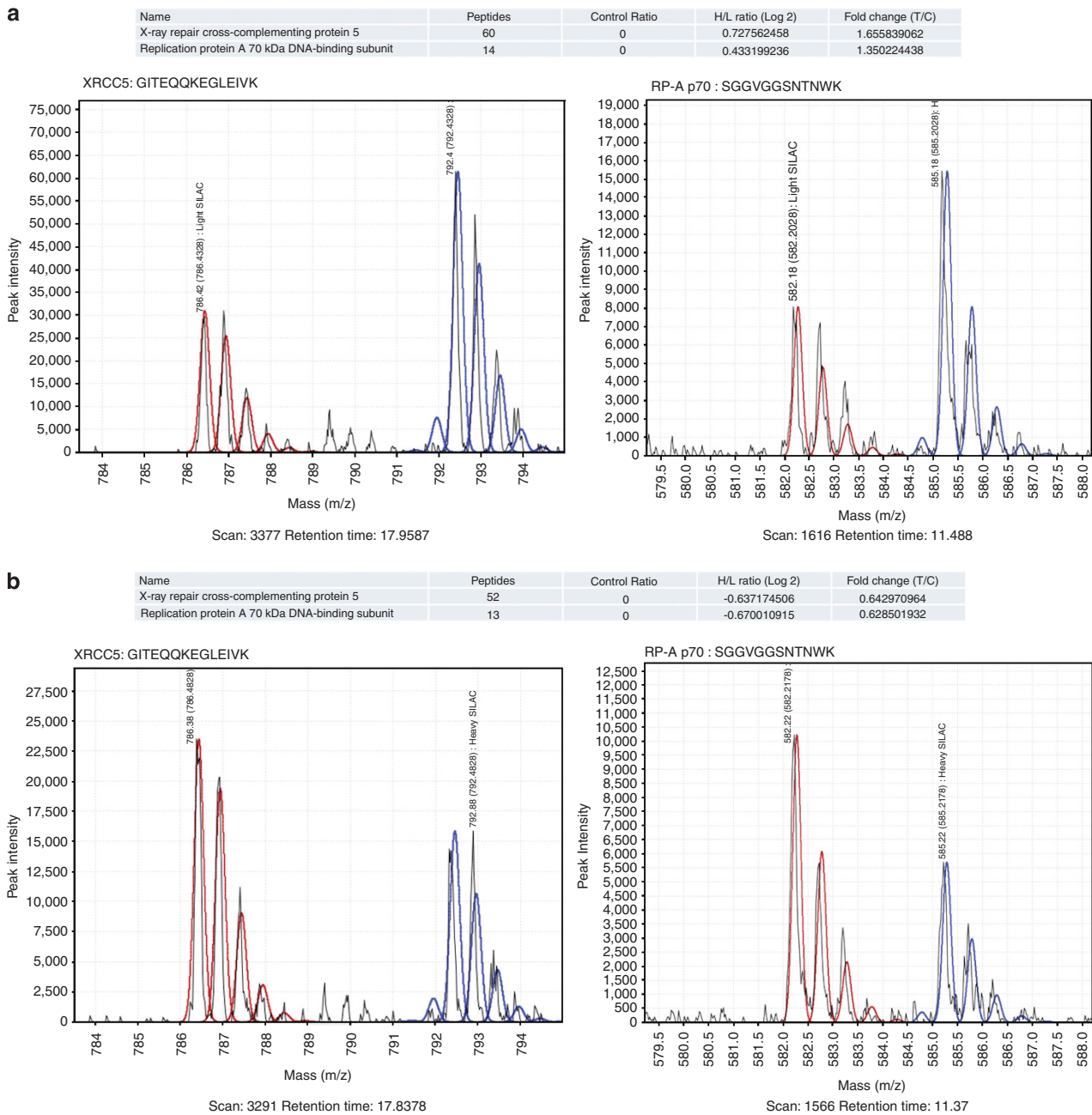Scan: 1566 Retention time: 11.37

Figure 4 Relative DNA binding affinity of XRCC5 and RP-A p70 between C and T allele of rs7903146. Peptide abundance and fold change are shown in the table. One peptide spectrum of each protein is given. (**a**) Forward labeling experiment. (**b**) Reverse labeling experiment.

5′ direction and thereby regulate the interactions between proteins and DNA or RNA;[36] indeed, RNA helicases participate in multiple biological processes, including transcription, splicing and translation.[37] The mechanism by which these three protein operate resonates with the fact that the open chromatin in islets has been reported to differ between alleles of this key SNP.[38]

The fact that the top three most abundant binding proteins physically interact with the TCF7L2 protein suggests a feedback mechanism. After all, TCF7L2 regulates *MYC* at the key cancer GWAS-implicated locus, so it is conceivable that TCF7L2 regulates *TCF7L2* at the key T2D GWAS-implicated locus. Our reporter assay results are consistent with the fact that TCF7L2 and *β*-catenin are both required for canonical Wnt pathway signaling, and our data

strongly suggest that the Wnt signaling pathway has a crucial role in the regulation of *TCF7L2* expression itself, such that a possible alteration in this feedback mechanism via rs7903146 could, at least in part, explain the functional mechanism through which this variant confers T2D risk. Interestingly, our previous TCF7L2 ChIP-seq work characterized four sites within the *TCF7L2* gene itself,[39] three of which were in intron 3 in close single-digit kilobase proximity to rs7903146, thus resonating with these results; these findings, including the fact that TCF7L2 was not detected in our MS work but did exert an effect when using a much longer DNA context with the luciferase assays, suggest that TCF7L2 is binding from a distance and interacts with this complex via DNA folding, which would need to be fully resolved in future studies using techniques such as 3C. Curiously, we

observed stronger transcriptional activity in the reverse direction than in the forward direction. Importantly, we detected differences between alleles, which may partly explain functional difference between alleles.

The consistent results from two-way pull-down suggest that XRCC5 and RP-A p70, additional logical members of the detected protein complex, may have a role in the regulation of TCF7L2 by an allele-specific binding preference. XRCC5, also known as Ku80, is a component of ATP-dependent DNA helicase II complex and is known to competitively regulate β-catenin and TCF7L2-mediated gene transactivation.[20] The XRCC5/6 heterodimer is also involved in DNA damage and repair.[40,41] The replication protein A 70 kDa DNA-binding subunit (RP-A p70), also known as replication factor A protein 1, participates in multiple biological processes in DNA metabolism, such as recombination, replication, damage and repair.[42,43]

However, it should be noted that the difference in DNA binding affinity of these two proteins, based on our MS results, between the C and T allele of rs7903146 are moderate. Indeed, it is conceivable that not all proteins in the complex mediate an effect directly so it is possible that not all would yield an allelic difference; but it is also possible that the lack of allelic difference with the more abundant proteins could be due to the sensitivity of the method we used. In fact, the odds ratio for the T-risk allele of rs7903146 is approximately 1.4 in GWAS studies. Odds ratios reported in GWAS are generally relatively modest as it is widely viewed that common diseases are caused by a large number of causal variants with small effect sizes. Therefore, the development of new approaches for sensitive, reliable and quantitative analyses in biological processes is necessary for effects of this kind. Quantitative and high-throughput proteomics may prove to be powerful tools for truly post-GWAS functional studies.

Although we ascertained this binding complex to be specific to rs7903146, where it was neither binding to a scrambled oligo nor to its strongest non-causal proxy, more work is required. More studies need to be carried out in other cell lines, tissues and settings beyond the standard culturing conditions for colorectal carcinoma cell line HCT116 used in this study. We elected to use HCT116 in this study as this colorectal carcinoma cell line has been leveraged in the past to study TCF7L2 activity,[44] and this gene has long been known to have a role in colorectal cancer as well as T2D. This cell line was considered optimal for our study for a number of reasons. First, our previously published ChIP-seq analysis of TCF7L2 occupancy across the genome in HCT116 revealed that the top enriched pathways for the genes bound by this transcription factor were related to diabetes and cardiovascular traits.[39] Furthermore, as TCF7L2 protein levels are abundant in non-transformed HCT116 cells, the gene is clearly being expressed in this setting. Also, this cell line divides rapidly, thus providing us with sufficient levels of total protein extract in a timely manner for us to carry out the oligo pull-down experiments.

HCT116 is heterozygous for rs7903146, but we do not believe this to be relevant to the experiments we carried out; after all, we are just leveraging the nuclear extract from this cell line to interact with our designed oligos. As such, we were not considering or engaging the genomic status of the cell at all. Furthermore, rs7903146 is intronic so it does not directly influence the amino-acid sequence of the TCF7L2 protein present in the cell.

Collectively, we propose a possible mechanism where TCF7L2 interacts with a protein complex, which includes PARP-1, to form a complex to regulate TCF7L2 itself. However, more detailed studies of these proteins are required to fully characterize the functional meaning of these interactions.

AUTHOR CONRIBUTIONS

QX, SD, CXY, MEJ and SFAG researched data. QX and SFAG wrote manuscript. SD, CXY and MEJ reviewed/edited manuscript.

1 Sladek R, Rocheleau G, Rung J *et al*: A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007; **445**: 881–885.
2 Wellcome Trust Case Control Consortium. Genome-wide association study of 14 000 cases of seven common diseases and 3000 shared controls. *Nature* 2007; **447**: 661–678.
3 Saxena R, Voight BF, Lyssenko V *et al*: Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007; **316**: 1331–1336.
4 Zeggini E, Weedon MN, Lindgren CM *et al*: Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 2007; **316**: 1336–1341.
5 Scott LJ, Mohlke KL, Bonnycastle LL *et al*: A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 2007; **316**: 1341–1345.
6 Grant SF, Thorleifsson G, Reynisdottir I *et al*: Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* 2006; **38**: 320–323.
7 Zeggini E, McCarthy MI: TCF7L2: the biggest story in diabetes genetics since HLA? *Diabetologia* 2007; **50**: 1–4.
8 Weedon MN: The importance of TCF7L2. *Diabet Med* 2007; **24**: 1062–1066.
9 Clevers H: Wnt/beta-catenin signaling in development and disease. *Cell* 2006; **127**: 469–480.
10 Yi F, Brubaker PL, Jin T: TCF-4 mediates cell type-specific regulation of proglucagon gene expression by beta-catenin and glycogen synthase kinase-3beta. *J Biol Chem* 2005; **280**: 1457–1464.
11 Korinek V, Barker N, Moerer P *et al*: Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nat Genet* 1998; **19**: 379–383.
12 da Silva Xavier G, Loder MK, McDonald A *et al*: TCF7L2 regulates late events in insulin secretion from pancreatic islet beta-cells. *Diabetes* 2009; **58**: 894–905.
13 Loder MK, da Silva Xavier G, McDonald A, Rutter GA: TCF7L2 controls insulin gene expression and insulin secretion in mature pancreatic beta-cells. *Biochem Soc Trans* 2008; **36**: 357–359.
14 Kaminska D, Kuulasmaa T, Venesmaa S *et al*: Adipose tissue TCF7L2 splicing is regulated by weight loss and associates with glucose and fatty acid metabolism. *Diabetes* 2012; **61**: 2807–2813.
15 Boj SF, van Es JH, Huch M *et al*: Diabetes risk gene and Wnt effector Tcf7l2/TCF4 controls hepatic response to perinatal and adult metabolic demand. *Cell* 2012; **151**: 1595–1607.
16 Helgason A, Palsson S, Thorleifsson G *et al*: Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution. *Nat Genet* 2007; **39**: 218–225.
17 Palmer ND, Hester JM, An SS *et al*: Resequencing and analysis of variation in the TCF7L2 gene in African Americans suggests that SNP rs7903146 is the causal diabetes susceptibility variant. *Diabetes* 2011; **60**: 662–668.
18 Maller JB, McVean G, Byrnes J *et al*: Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat Genet* 2012; **44**: 1294–1301.
19 Helgason A, Palsson S, Thorleifsson G *et al*: Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution. *Nat Genet* 2007; **39**: 218–225.
20 Idogawa M, Yamada T, Honda K, Sato S, Imai K, Hirohashi S: Poly(ADP-ribose) polymerase-1 is a component of the oncogenic T-cell factor-4/beta-catenin complex. *Gastroenterology* 2005; **128**: 1919–1936.
21 Tulin A, Chinenov Y, Spradling A: Regulation of chromatin structure and gene activity by poly(ADP-ribose) polymerases. *Curr Top Dev Biol* 2003; **56**: 55–83.
22 Kraus WL, Lis JT: PARP goes transcription. *Cell* 2003; **113**: 677–683.
23 Szabo C, Zanchi A, Komjati K *et al*: Poly(ADP-Ribose) polymerase is activated in subjects at risk of developing type 2 diabetes and is associated with impaired vascular reactivity. *Circulation* 2002; **106**: 2680–2686.
24 Ilnytska O, Lyzogubov VV, Stevens MJ *et al*: Poly(ADP-ribose) polymerase inhibition alleviates experimental diabetic sensory neuropathy. *Diabetes* 2006; **55**: 1686–1694.
25 Li F, Drel VR, Szabo C, Stevens MJ, Obrosova IG: Low-dose poly(ADP-ribose) polymerase inhibitor-containing combination therapies reverse early peripheral diabetic neuropathy. *Diabetes* 2005; **54**: 1514–1522.
26 Obrosova IG, Li F, Abatan OI *et al*: Role of poly(ADP-ribose) polymerase activation in diabetic neuropathy. *Diabetes* 2004; **53**: 711–720.

27 Pieper AA, Brat DJ, Krug DK *et al*: Poly(ADP-ribose) polymerase-deficient mice are protected from streptozotocin-induced diabetes. *Proc Natl Acad Sci USA* 1999; **96**: 3059–3064.

28 Yochum GS, McWeeney S, Rajaraman V, Cleland R, Peters S, Goodman RH: Serial analysis of chromatin occupancy identifies beta-catenin target genes in colorectal carcinoma cells. *Proc Natl Acad Sci USA* 2007; **104**: 3324–3329.

29 Duval A, Rolland S, Tubacher E, Bui H, Thomas G, Hamelin R: The human T-cell transcription factor-4 gene: structure, extensive characterization of alternative splicings, and mutational analysis in colorectal cancer cell lines. *Cancer Res* 2000; **60**: 3872–3879.

30 Pomerantz MM, Ahmadiyeh N, Jia L *et al*: The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat Genet* 2009; **41**: 882–884.

31 Tuupanen S, Turunen M, Lehtonen R *et al*: The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat Genet* 2009; **41**: 885–890.

32 Sur IK, Hallikas O, Vaharautio A *et al*: Mice lacking a Myc enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. *Science* 2012; **338**: 1360–1363.

33 Frayling TM, Colhoun H, Florez JC: A genetic link between type 2 diabetes and prostate cancer. *Diabetologia* 2008; **51**: 1757–1760.

34 Lebedeva NA, Rechkunova NI, Agama K, Pommier Y, Lavrik OI: Interaction of DNA topoisomerase 1 with DNA intermediates and proteins of base excision repair. *Biochemistry (Mosc)* 2009; **74**: 1278–1284.

35 Malanga M, Althaus FR: Poly(ADP-ribose) reactivates stalled DNA topoisomerase I and Induces DNA strand break resealing. *J Biol Chem* 2004; **279**: 5244–5248.

36 Weidensdorfer D, Stohr N, Baude A *et al*: Control of c-myc mRNA stability by IGF2BP1-associated cytoplasmic RNPs. *RNA* 2009; **15**: 104–115.

37 Cordin O, Banroques J, Tanner NK, Linder P: The DEAD-box protein family of RNA helicases. *Gene* 2006; **367**: 17–37.

38 Gaulton KJ, Nammo T, Pasquali L *et al*: A map of open chromatin in human pancreatic islets. *Nat Genet* 2010; **42**: 255–259.

39 Zhao J, Schug J, Li M, Kaestner KH, Grant SF: Disease-associated loci are significantly over-represented among genes bound by transcription factor 7-like 2 (TCF7L2) in vivo. *Diabetologia* 2010; **53**: 2340–2346.

40 Taccioli GE, Gottlieb TM, Blunt T *et al*: Ku80: product of the XRCC5 gene and its role in DNA repair and V(D)J recombination. *Science* 1994; **265**: 1442–1445.

41 Roberts SA, Strande N, Burkhalter MD *et al*: Ku is a 5'-dRP/AP lyase that excises nucleotide damage near broken ends. *Nature* 2010; **464**: 1214–1217.

42 Mason AC, Haring SJ, Pryor JM, Staloch CA, Gan TF, Wold MS: An alternative form of replication protein a prevents viral replication in vitro. *J Biol Chem* 2009; **284**: 5324–5331.

43 Kemp MG, Mason AC, Carreira A *et al*: An alternative form of replication protein a expressed in normal human tissues supports DNA repair. *J Biol Chem* 2010; **285**: 4788–4797.

44 Savic D, Park SY, Bailey KA, Bell GI, Nobrega MA: In vitro scan for enhancers at the TCF7L2 locus. *Diabetologia* 2013; **56**: 121–125.