An in vivo cis-Regulatory Screen at the Type 2 Diabetes Associated TCF7L2 Locus Identifies Multiple Tissue-Specific Enhancers

Daniel Savic1*, Graeme I. Bell1,2, Marcelo A. Nobrega1*

1 Department of Human Genetics, University of Chicago, Chicago, Illinois, United States of America, 2Department of Medicine, University of Chicago, Chicago, Illinois, United States of America

Abstract

Genome-wide association studies (GWAS) have repeatedly shown an association between non-coding variants in the TCF7L2 locus and risk for type 2 diabetes (T2D), implicating a role for cis-regulatory variation within this locus in disease etiology. Supporting this hypothesis, we previously localized complex regulatory activity to the TCF7L2 T2D-associated interval using an in vivo bacterial artificial chromosome (BAC) enhancer-trapping reporter strategy. To follow-up on this broad initial survey of the TCF7L2 regulatory landscape, we performed a fine-mapping enhancer scan using in vivo mouse transgenic reporter assays. We functionally interrogated approximately 50% of the sequences within the T2D-associated interval, utilizing sequence conservation within this 92-kb interval to determine the regulatory potential of all evolutionary conserved sequences that exhibited conservation to the non-eutherian mammal opossum. Included in this study was a detailed functional interrogation of sequences spanning both protective and risk alleles of single nucleotide polymorphism (SNP) rs7903146, which has exhibited allele-specific enhancer function in pancreatic beta cells. Using these assays, we identified nine segments regulating various aspects of the TCF7L2 expression profile and that constitute nearly 70% of the sequences tested. These results highlight the regulatory complexity of this interval and support the notion that a TCF7L2 cis-regulatory disruption leads to T2D predisposition.

Introduction

Intronic variation located in a 92-kb interval within the Transcription Factor 7-Like 2 (TCF7L2) gene locus, a transcriptional regulator of canonical Wnt signaling [1,2], is the strongest determinant for type 2 diabetes (T2D) susceptibility identified to date [3,4,5]. Indeed, associations have been reported in populations across from the globe [4] and variation in this locus remains the strongest genetic determinant of T2D risk in humans [6]. As both in vitro and in vivo functional analyses support a role for TCF7L2 in glucose metabolism [7,8,9,10,11,12,13,14,15], this transcription factor is regarded as the candidate target gene of the association.

The underlying molecular, cellular and physiological mechanisms(s) by which TCF7L2 affects T2D risk are largely unknown. The results, to date, point to a role for variation in long-range cis-regulatory elements in T2D pathogenesis through alterations in TCF7L2 expression. In this regard, two independent studies in pancreatic islets uncovered allele-specific enhancer activity for sequences spanning single nucleotide polymorphism (SNP) rs7903146, the variant showing the strongest association to T2D [16,17]. These analyses are consistent with the recent functional studies of non-coding variants in other disease-associated GWAS loci and highlight the importance of cis-regulatory variation in affecting disease risk [18]. From the standpoint of the common disease common variant (CDCV) hypothesis, the implications for common disease risk are clear as variation in these regulatory sequences can lead to a compartmentalization of phenotypic effects as enhancer elements largely govern activity in a spatial and temporal context, mitigating pleiotropic effects [18]. This would ostensibly allow these disadvantageous non-coding variants to reach higher frequencies in human populations as compared with protein-coding variants that can elicit disruptions with broader consequences on target gene activity [18].

We previously interrogated the cis-regulatory landscape of TCF7L2 using a bacterial artificial chromosome (BAC) enhancer-trapping strategy and identified widespread enhancer activity that we localized to the association interval [12]. However, a systematic fine-mapping analysis of the association interval is still lacking. To follow-up on this BAC survey, here we characterized the long-range cis-regulatory landscape of this T2D-associated genomic locus through an in vivo fine-mapping approach. We demonstrate that the association interval harbors a wide variety of tissue-specific enhancers, including a subset that drives expression in peripheral tissues involved in glucose homeostasis, adding support to a potential regulatory defect in T2D etiology.
**Methods**

**Ethics Statement**

All mice were housed at the University of Chicago. Veterinary care was available on a 24-hour basis. Mice were monitored daily for any signs of illness or discomfort. All experiments were conducted in strict accordance with institutional rules and approved by the University of Chicago Institutional Animal Care and Use Committee, protocol number 71636 (M.A.N.).

**Molecular cloning**

Conserved sequences within the association interval were cloned (Table S1) with Gateway technology (Invitrogen) in a custom vector containing a heat shock minimal promoter (Hsp68) driving lacZ expression. The transgenic lacZ plasmids were linearized, resuspended in 1x microinjection buffer and used for pronuclear injection into fertilized oocytes (CD-1) using standard protocols approved by the University of Chicago Institutional Animal Care and Use Committee.

**Mouse in vivo Transgenic Reporter Assays**

Animals were sacrificed using carbon dioxide gas followed by cervical dislocation. Embryos were harvested at embryonic day 15.5 (E15.5) or 16.5 (E16.5). Pancreatic staining was also performed postnatally on day 0 (P0) and 6 (P6). Following harvesting and dissection, embryos and tissues were placed into cold 100 mM phosphate buffer, pH 7.3 (PBS), followed by an hour of incubation with 4% paraformaldehyde at 4°C. Tissues were then washed with 1x PBS and further washed two additional times for 20 min using lacZ wash buffer (2 mM MgCl2; 0.01% deoxycholate; 0.02% NP-40; 100 mM phosphate buffer, pH 7.3), and stained for 16–20 hours at room temperature with lacZ staining solution (1 mg/ml X-gal; 4 mM potassium ferrocyanide; 4 mM potassium ferricyanide; 20 mM Tris-HCl, pH 7.5 in wash buffer). After staining, embryos and tissues were rinsed 5 times in PBS and post-fixed and stored in 4% paraformaldehyde at 4°C. Images were taken using a Leica MZ 16 F imaging system.

**Results**

**Enhancer fine-mapping at the TCF7L2 association interval**

As sequence conservation is a predictor of function [19], 13 evolutionary conserved regions (ECRs) spanning sequences exhibiting significant conservation between human and the non-eutherian mammal opossum were tested in mouse transgenic assays (Figure 1). These ECRs were cloned into a lacZ reporter construct driven by a heat shock protein 68 (Hsp68) minimal promoter. In total ~48.7 kb of sequence, or approximately 50% of the T2D-associated interval, was tested for regulatory activity in vivo (Figure 1A). For each construct, we obtained multiple transgenic lines to ensure reproducibility of enhancer patterns across independent transgenic lines. Of the 13 evolutionarily conserved regions, we observed reproducible enhancer activity in 9/13 (69%) regions tested (ECR 1, 3, 4, 5, 6, 8, 9, 11 and 13). The enhancers mapping to these intervals exhibit a diverse array of spatial expression patterns that were reproducible across independent transgenic lines (Figure 1B). Interestingly, five regions (ECR 1, 4, 5, 11 and 13) exhibited regulatory potential in tissues with known roles in controlling glucose homeostasis such as the stomach (ECR 1), bone (ECR 4 and 13) and brain (ECR 5 and 11). Beta-galactosidase staining was also localized to spinal nerves (ECR 3), walking pads (ECR 6) and the vasculature of both the limbs (ECR 8) and brain (ECR 9).

**Functional analysis of sequences spanning SNP rs7903146**

We next investigated the regulatory potential of sequences containing SNP rs7903146. For this analysis, we utilized the previous results of the construct that encompassed the protective C allele at SNP rs7903146 from ECR 5 (ECR 5-C) and engineered an identical construct that spanned the risk T allele at SNP rs7903146 (ECR 5-T; Figure 2A). In order to dissociate any regulatory effects exhibited by a conserved sequence situated downstream of this SNP, we further generated a shorter construct (ECR 5B) that was restricted to this region of conservation (Figure 2A). As SNP rs7903146 resides within a prime-specific short interspersed nuclear element (SINE) and ‘enhancer boosting’ properties were demonstrated for repetitive elements [20], this approach also allows for the identification of potential ‘enhancer boosting’ activity. All constructs were sequence-verified prior to pronuclear injections. For each of these new constructs, we obtained multiple transgenic lines (Figure 2B, C, D).

All constructs exhibited reproducible expression within the forebrain across multiple transgenic lines, suggesting that the conserved sequence downstream of SNP rs7903146 governs this activity (Figure 2B, C, D). However, we observed no allelic-specific (comparing ECR 5-C and 5-T) or ‘enhancer boosting’ effects (comparing ECR 5B and 5-C/T) as forebrain expression across all constructs was largely consistent. Although some pancreatic expression was present in mice harboring the longer constructs encompassing SNP rs7903146 (ECR 5-C and 5-T) and this pattern was absent in transgenic lines containing the shorter construct ECR 5B, this activity was not highly reproducible across multiple independent transgenic lines, nor did this expression exhibit allelic-specific effects. This conclusion is further corroborated by analyses at postnatal developmental stages in transgenic animals as consistent allelic differences were not observed at postnatal days 0 or 6 (Figure S1).

**Discussion**

Our screen defines the fine-scale regulatory landscape of the TCF7L2 T2D-associated region. To our knowledge, this is the first detailed mapping study conducted on this GWAS-associated region. Importantly, the enhancers identified from this fine-mapping scan recapitulate various aspects of BAC enhancer activities that we previously uncovered at this locus [12]. While significant attention has been given to the role of TCF7L2 in pancreatic islets, our data delineates several regions harboring regulator activity in peripheral metabolic tissues. In particular, we identified elements that govern expression within the bone, brain and stomach.

The localization of bone enhancers is not surprising, given the well established role of the canonical Wnt signaling in bone formation [21]. For instance, ablation of Wnt regulators leads to bone mass defects in mice [22,23,24,25] while disruptions of Wnt signaling antagonists generates opposing phenotypes [26,27]. As a previously unappreciated role for bone in the regulation of glucose homeostasis has been recently established, this organ has become an interest to the diabetes community [28,29]. Consequently, investigations of this transcription factor in bone-mediated glucose metabolism are clearly warranted.

The fine-mapping of a stomach enhancer within the associated interval is of relevance as this tissue secretes the orexigenic hormone ghrelin that is involved in energy and glucose homeostasis [30,31]. A recent study further uncovered a correlation between a putative TCF7L2 neuroendocrine splicing variant and the anorexigenic peptide CART [32], supporting a role for this
The canonical Wnt regulator in satiety. The extent of TCF7L2 involvement in energy metabolism and the further implications this may have for T2D risk is a disease mechanism demanding more scrutiny.

The canonical Wnt signaling pathway is also implicated in diverse neurological disorders such as autism, Alzheimer’s disease and schizophrenia [33]. Consistent with a neurological function, the TCF7L2 T2D-associated interval has been implicated in schizophrenia risk [34]. Indeed, we observed behavioral phenotypes in mice with altered Tcf7l2 levels [35]. As the historical comorbidity between T2D and schizophrenia is well documented, this may point to a common disease etiology [36,37]. In light of these observations, our identification of several brain enhancers, and in particular a forebrain enhancer situated in the vicinity of SNP rs7903146, may be of interest.

We did not observe robust allelic-specific enhancer activity for SNP rs7903146. Although we cannot exclude potential effects at other developmental stages, the use of a non-native promoter element (Hsp68) or potential complex long-range interactions (i.e. enhancer-enhancer) that was not assessed by our assay may explain these results. The previous localization of a pancreatic enhancer using an in vivo BAC transgenic strategy supports these conclusions [12]. Alternatively, as the allelic-specific properties at this locus were uncovered through cell-based luciferase and open chromatin assays [16,17], our results may primarily reflect a limitation in assessing quantitative differences with qualitative approaches. This carries broader implications for GWAS loci in general as potential causal variants are likely to constitute a modest effect on disease risk [38,39,40] and therefore may generate finer regulatory defects that are difficult to assess in vivo.

**Figure 1. Enhancer screen of evolutionary conserved sequences within TCF7L2 association interval.** (A) The TCF7L2 gene is shown above with the 92-kb T2D-associated interval highlighted in red. A red asterisk marks SNP rs7903146. Sequence conservation between human-opossum is given (ECR genome browser, [43]). The 13 evolutionary conserved regions (ECRs) harboring conservation down to opossum are highlighted in grey and numbered below. Regions exhibiting reproducible enhancer activity are marked in red. (B) Reproducible expression profiles from ECRs at embryonic day 15.5 (E15.5) are shown. Expression can be seen in the stomach (ECR 1), neurons near the spine (ECR 3), limb bones and axial skeleton (ECR 4), forebrain (dorsal view, ECR 5), walking pads (ECR 6), limb vasculature (ECR 8), brain vasculature (dorsal view, ECR 9), limbs and midbrain (dorsal view, ECR 11), and phalanges (ECR 13). In ECRs 4, 6, 8, 11 and 13, both forelimb (left) and hindlimb (right) images are given. Reproducible regulatory activities were not identified for ECRs 2, 7, 10 and 12.

doi:10.1371/journal.pone.0036501.g001
Our enhancer screen was also largely restricted to one embryonic developmental stage, E15.5. Despite this limitation, our previous BAC results suggest that the regulatory landscape at this locus is faithfully maintained into adulthood [12]. Another potential concern is the use of evolutionary conserved sequences in our in vivo assays. It has become common knowledge that a number of cis-regulatory regions, including those involved in embryonic development, lack evolutionary conservation across distant species. Our in vivo scan is far from being exhaustive, and we most likely missed other cis-regulatory elements embedded in the ~50% of sequences within the T2D-associated region that we did not test in this study. Nevertheless, our results conclusively indicate that the T2D-associated interval contains at least nine tissue-specific regulatory elements. Despite the strong concordance with the endogenous TCF7L2 expression profile [12], we cannot definitively rule out the possibility that some of these enhancers could be involved in the regulation of neighboring genes such as the upstream gene VTI1A. Genetic variation in a number of these enhancers, common or rare in populations, may lead to alterations in TCF7L2 or neighboring gene expression, leading to various phenotypic consequences. To that end, the same genomic interval has now also been associated with increased risk to schizophrenia [34], colorectal cancer [41] and coronary artery disease [42].

Our results highlight the complex regulatory nature of the 92-kb T2D-associated region of TCF7L2 and support the hypothesis that sequence variation within distal cis-regulatory elements are mediators of T2D susceptibility. The identification of several enhancers that drive expression in diverse metabolic domains further points to a possible disease etiology involving peripheral metabolic tissues.

Figure 2. Functional analysis of SNP rs7903146. (A) The association interval within the TCF7L2 gene locus is highlighted in red above. A red asterisk marks SNP rs7903146. Sequence conservation between human-opossum (ECR genome browser, [43]) is given for the entire association interval while the sequence tested is highlighted in green. Within the tested region, sequence conservation between human-opossum is given below (ECR genome browser). Positions of regions spanning SNP rs7903146 (5-C and 5-T) as well as a shorter sequence limited to the downstream ECR (5B) are shown below. (B)-(D) Images of pancreas (attached to the stomach) (left) and brain (right, dorsal view) obtained from independent transgenic lines (rows) are shown at E15.5 or E16.5 (ECR 5-C row 3 only). (B) and (C) Regions 5-C and 5-T exhibits inconsistent pancreatic staining while maintaining reproducible forebrain expression. (B) Construct 5B harbors forebrain expression but no pancreatic staining.

doi:10.1371/journal.pone.0036501.g002
Supporting Information

Figure S1  Postnatal analyses of pancreatic expression. Stable transgenic lines were stained for pancreatic beta-galactosidase activity on postnatal days 0 (P0, top panel) and 6 (P6, bottom panel). Pancreatic images for sequences spanning the protective C allele at SNP rs7903146 (5-C) and risk T allele at SNP rs7903146 (5-T) are shown. (TIFF)

Table S1  Primer sequences for amplification of evolutionary conserved regions within the TCF7L2 association interval. The evolutionary conserved region (ECR) is numbered in the first column. Subsequent columns give the primer pair sequences (in 5’ to 3’ orientation) for each ECR. (TIFF)

Author Contributions

Conceived and designed the experiments: DS GIB MAN. Performed the experiments: DS. Analyzed the data: DS GIB MAN. Contributed reagents/materials/analysis tools: DS GIB MAN. Wrote the paper: DS.