Identification and characterization of a FOXA2-regulated transcriptional enhancer at a type 2 diabetes intronic locus that controls GCKR expression in liver cells

Maykel López Rodríguez¹, Dorota Kaminska², Kati Lappalainen⁴, Jussi Pihlajamäki⁵,⁶, Minna U. Kaikkonen⁴† and Markku Laakso¹,²∗†

Abstract

Background: Genome-wide association studies (GWAS) have identified more than 100 genetic loci associated with type 2 diabetes (T2D). However, the underlying biological mechanisms for many of these associations remain unknown. GWAS signals close to the glucokinase regulatory protein gene (GCKR) have been reported for lipid and glucose metabolism traits and the risk of T2D. We investigated the regulatory function of an intronic locus at GCKR represented by the lead single nucleotide polymorphism (SNP) rs780094.

Methods: We used ENCODE project histone modification and transcription factor binding data to determine the regulatory features of a GCKR intronic locus formed by the high linkage disequilibrium rs780094(C/T), rs780095(G/A), and rs780096(G/C) SNPs. Characterization of the transcriptional activity of this region was assessed by luciferase reporter assays in HepG2 cells and mouse primary hepatocytes. ChIP-qPCR was used to determine the levels of haplotype specific transcription factor binding and histone marks. A CRISPR-dCas9 transcriptional activator system and qPCR were used to activate the locus and measure GCKR expression, respectively. Differential haplotype expression was measured from human liver biopsies.

Results: The ENCODE data suggest the existence of a liver-specific intragenic enhancer at the locus represented by rs780094. We observed that FOXA2 increased the transcriptional activity of this region in a haplotype specific way (CGG > TAC; rs780094, rs780095, and rs780096). In addition, the CGG haplotype showed higher binding to FOXA2 and higher levels of the H3K27Ac histone mark. Epigenetic activation of this locus increased the expression of endogenous GCKR in HepG2 cells, confirming that GCKR is the direct target gene of the enhancer. Finally, we confirmed that the CGG haplotype exhibits higher levels of transcription in human liver.

Conclusions: Our results demonstrate the existence of a liver-specific FOXA2-regulated transcriptional enhancer at an intronic T2D locus represented by rs780094, rs780095, and rs780096 SNPs that increases GCKR expression. Differential haplotype regulation suggests the existence of cis regulatory effects that may contribute to the associated traits at this locus.

Keywords: rs780094, rs780095, rs780096, GCKR, Transcriptional enhancer, FOXA2, Type 2 diabetes

* Correspondence: markku.laakso@uef.fi
† Equal contributors
¹ Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Yliopistonranta 1 C, 70211 Kuopio, Finland
² Department of Medicine, Kuopio University Hospital, P.O. Box 100, FI 70029, KYS, Kuopio, Finland
Full list of author information is available at the end of the article

© The Author(s). 2017 Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
Background

Type 2 diabetes (T2D; MIM 125853) develops as a consequence of complex interactions between genetic and environmental/lifestyle factors. Genome-wide association studies (GWAS) have identified more than 100 common single nucleotide polymorphisms (SNPs) associated with T2D and glycemic traits [1]. Although some association signals are due to coding variants, the majority of signals do not map to protein coding sequence, which suggests that SNPs influence gene regulation.

The glucokinase regulatory protein gene (GCKR; MIM 613463) encodes for glucokinase regulatory protein (GKRP; NP_001477), the main allosteric inhibitor of glucokinase (HK4; P35557.1; GCK hereafter), and is a T2D-associated locus. GKRP inhibits GCK by a direct effect on enzyme kinetics or by sequestering GCK in the nucleus, where GCK is unable to facilitate glycolysis [2]. This inhibitory mechanism operates as a negative feedback loop in the control of glucose disposal by the liver. In the fasting state, GCK is bound to GKRP and retained in the nucleus. In the fed state, the increase in glucose concentrations induces the dissociation of GCK from GKRP, which activates GCK and facilitates its release into the cytoplasm [2, 3]. The subsequent increase in fructose-6-phosphate, attributable to glycolysis, results in the re-formation of the GKRP–GCK complex and enzyme inhibition [2, 3]. Despite the inhibitory effect of GKRP on GCK, this nuclear sequestration during fasting is proposed to protect GCK from degradation, which in turn guarantees an effective pool of the enzyme to be released into the cytoplasm in response to glucose [4].

Given the major role of hepatic GCK in glucose metabolism in the liver, disturbances in GKRP–GCK regulation at the transcriptional or post-transcriptional level may directly affect glucose metabolism and produce subsequent changes in other metabolic traits.

GCKR lies in a large region of linkage disequilibrium (LD) on chromosome 2, spanning about 417 kb, 16 genes, and several correlated variants [5]. Genetic fine-mapping of this region has localized a GWAS signal to GCKR rather than to other genes in the LD block [5]. These studies also identified the nonsynonymous rs1260326 SNP (C/T, P446L substitution) as the strongest signal for fasting blood glucose and total triglycerides in this region [5]. Functional studies on this variant demonstrate that the P446L amino acid substitution results in lower GCK sequestration capacity and impaired response to fructose-6-P [6, 7], which is thought to affect glucose levels indirectly by affecting the cytoplasmic availability and activity of GCK [6, 7]. Thus, rs1260326 has been established as a functional SNP which is likely to have a causal relationship with GCKR-related traits.

Despite the functional evidence that the rs1260326 variant impacts on both the kinetics and cellular localization of GKRP, the high LD in the region warrants the investigation of other variants which could contribute to molecular mechanisms associated with multiple traits. One such variant is an intronic SNP, rs780094, which was originally identified to be associated with fasting serum triacylglycerol, insulinemia, and the risk of T2D [8]. As expected for SNPs in high LD, rs780094 and rs1260326 ($r^2 = 0.94$) overlap in their phenotypical associations [9]. Thus, their independent effects cannot be accurately assessed based on association analysis. While a molecular mechanism has been elucidated for the P446L variant [6, 7], no functional role has been reported for rs780094. Given its location at the non-coding region, we hypothesized that this locus could regulate GCKR transcription. Another two SNPs, rs780095 and rs780096, are in high LD ($r^2 = 0.86$) and are located very close (132 and 165 bp, respectively) to rs780094. Therefore, it is likely that these variants overlap in their functional roles and thus their haplotype combinations may influence the metabolic susceptibility.

In the present study we demonstrate that the locus containing rs780094 functions as a transcriptional enhancer that is activated by the Forkhead box protein A2 (FOXA2; Q9Y261) transcription factor (TF) in a haplotype specific fashion. We show that epigenetic activation of the enhancer induces GCKR expression and that GCKR is a direct target of this regulatory region. In addition, we show an association between haplotype expression and the levels of free fatty acids (FFAs) in human liver biopsies. Our results suggest a novel mechanism for the regulation of GCKR expression that could have a combined effect with the P446L variant on the phenotypical associations at the GCKR locus.

Methods

Culture and stimulation of HepG2

HepG2 cells (ATCC, HB-8065) were cultured in Dulbecco’s modified Eagle medium (DMEM; 4.5 g/L glucose, 2 mM L-glutamine, 100 U/ml penicillin, 100 μg/ml streptomycin; LONZA) supplemented with 10% fetal bovine serum (FBS; Gibco). Cells were seeded at 0.2–0.3 × 10^6 cells/ml in the appropriate vessel depending on the assay performed. For insulin stimulation in ChIP-qPCR experiments, the cells were serum-starved overnight and insulin (human recombinant insulin, Sigma-Aldrich; 100 nM) was added for 15 min before crosslinking. The cells were transfected using Lipofectamine 3000 (ThermoFisher) according to the manufacturer’s indications. The transfection conditions were optimized before conducting the over-expression experiments.

Preparation and stimulation of mouse primary hepatocytes

Mouse primary hepatocytes (MPHs) were prepared from 7–9-week-old C57BL male mice livers as previously
described [10]. Briefly, mice were anesthetized before dissection and the skin was opened by making a longitudinal middle ventral incision followed by two transversal incisions below the level of the liver. The vena cava superior was cannulated with a cannulation system (Fine Science Tools) coupled to a peristaltic pump (BioRad), a ligature was placed around the vena cava inferior, and the vena porta was cut open. Perfusion was performed with 250 ml of pre-perfusion solution (0.04% NaHCO₃, 10 mM HEPES, 0.5 mM EDTA in HBSS; pH adjusted to 7.4) followed by 35 ml of perfusion solution (10 mM HEPES, 80 U/ml collagenase [Worthington] in William’s E medium [Sigma-Aldrich]; pH adjusted to 7.4). Both the pre-perfusion and the perfusion solutions were prewarmed to 37 °C and circulated through an oxygen atmosphere and a 37 °C water bath before reaching the liver at 3 ml/min. After the perfusion, the liver was removed and placed on a petri dish containing a sufficient amount of plating medium (William’s E medium with 20 ng/ml dexamethasone [Sigma-Aldrich], ITS [5 mg/l insulin/5 mg/l transferrin/5 μg/l sodium selenite; Sigma-Aldrich], 10 μg/ml gentamicin [Sigma-Aldrich], and 10% FBS) and torn into pieces with sharp forceps to extract the cells. The cell suspension was filtered through a cell strainer (40 μm; Falcon) and centrifuged at 50 × g for 2 min. Cells were seeded at 0.5 × 10⁶ cells/ml on collagen-coated 96-well plates. Five hours after plating, medium was changed to maintenance medium (seeding medium without FBS), and the following day cells were transfected with Targefect F1 plus Virofect enhancer transfection reagents (Targeting Systems) according to the manufacturer’s indications. DNA constructs were expressed for 24 h in basal William’s E medium with 20 ng/ml dexamethasone and plaforms expressing human FOXA2, FOXA1, RXRA, or MAFK or the control vector pCMV6-XL5 encoding the luciferase gene hRluc (Renilla reniformis) and plasmids expressing human FOXA2, FOXA1, RXRA, or MAFK or the control vector pCMV6-XL5 without TF insert (SC122913; SC118299; SC118696; SC108256; pCMV6-XL5; OriGene) (molar ratios: 1:10, pGL4.75 to PGL4.10-TK; 1:1, luciferase constructs or PGL4.10-TK to TF plasmids or pCMV6-XL5; 1:1 FOXA2 to MAFK). The luciferase activity was measured 48 h post-transfection using the Dual-Glo Luciferase Assay System (Promega) in the Cytation 3 Multi-Mode Reader (BioTek). MPH were transfected with 60 ng of plasmid DNA and 24 h post-transfection luciferase activity was measured with the Dual Luciferase Assay System (Promega) in a Fluostar OPTIMA plate reader (BMG LABTECH). In all cases, the firefly luciferase activity was normalized to the Renilla signal and reported as proportions relative to the control vector. In HepG2 cells, we performed five or six independent experiments with three or more technical replicates (Fig. 2a, b, d; Additional file 2: Datasets). For the experiments in MPHs, four biological replicates corresponding to four livers were used (with two technical replicates each; Fig. 2c).

**Genotyping of HepG2 cells and human samples by sequencing**

gDNA from HepG2 cells or blood (from human donors) was extracted using the PureLink genomic DNA Mini Kit (Invitrogen). A 3-kb region from gDNA was amplified by PCR with Phusion High-Fidelity DNA polymerase (ThermoFisher) and specific primers (Additional file 1: Table S1) in a PRISM 2720 thermal cycler (Applied Biosystems). PCR amplicons were purified using the ChIP DNA Clean and Concentrator Kit (Zymo Biosciences). PCR amplicons were sequenced on the Nextera XT DNA Library Preparation Kit (Illumina) and ran on a MiSeq desktop sequencer (Illumina). The sequence reads were aligned to the human genome (hg19) using the Burrows–Wheeler Aligner (BWA) [13], and the allelic distributions were obtained with the Integrative Genomics Viewer (IGV) [14, 15].

**ENCODE ChIP-seq data analysis**
The processed data for HepG2 ChIP-Seq (bam and bigwig files) were downloaded from the ENCODE database under accession numbers GSM733743, GSM803499, GSM803461, GSM803403, GSM803452, GSM935610, GSM803432, GSM733693, GSM935610, and GSM9 35305. The allele-specific ChIP-Seq data were visually quantified from bam files using the IGV [14, 15].

Luciferase reporter transcriptional assays

DNA fragments containing the rs780094-rs780095-rs780096 region (Fig. 1a) were amplified by PCR from human homozygous genomic DNA (gDNA) selected from the METSIM study [11] with Phusion polymerase and specific primers (Additional file 1: Table S1). We selected the two most common haplotypes formed by these SNPs in the METSIM population [11] (CGG; TAC; rs780094-rs780095-rs780096; Additional file 2: Datasets). Amplified DNA was subcloned into BamHI and Sall restriction enzyme sites downstream of the firefly luciferase gene in a pGL4.10 vector (Promega) modified with a minimal TK (thymidine kinase) promoter as previously described (pGL4.10-TK) [12]. In total, four DNA inserts were amplified, two 638-bp inserts (one per haplotype) and two 3-kb inserts (one per haplotype). The luciferase constructs prepared with the inserts were verified by sequencing. HepG2 cells were transfected with 100 ng of plasmid DNA on 96-well plates. The luciferase constructs or the control vector pGL4.10-TK were co-transfected with pGL4.75 plasmid (Promega) encoding the luciferase gene hRluc (Renilla reniformis) and plasmids expressing human FOXA2, FOXA1, RXRA, or MAFK or the control vector pCMV6-XL5 without TF insert (SC122913; SC118299; SC118696; SC108256; pCMV6-XL5; OriGene) (molar ratios: 1:10, pGL4.75 to PGL4.10-TK; 1:1, luciferase constructs or PGL4.10-TK to TF plasmids or pCMV6-XL5; 1:1 FOXA2 to MAFK). The luciferase activity was measured 48 h post-transfection using the Dual-Glo Luciferase Assay System (Promega) in the Cytation 3 Multi-Mode Reader (BioTek). MPH were transfected with 60 ng of plasmid DNA and 24 h post-transfection luciferase activity was measured with the Dual Luciferase Assay System (Promega) in a Fluostar OPTIMA plate reader (BMG LABTECH). In all cases, the firefly luciferase activity was normalized to the Renilla signal and reported as proportions relative to the control vector. In HepG2 cells, we performed five or six independent experiments with three or more technical replicates (Fig. 2a, b, d; Additional file 2: Datasets). For the experiments in MPHs, four biological replicates corresponding to four livers were used (with two technical replicates each; Fig. 2c).
values were normalized to account for the allelic imbalance at rs780094 (C = 2XT).

**ChIP-qPCR**

ChIP experiments were performed in HepG2 cells with minor changes from a previously published protocol [16]. Briefly, the cells were crosslinked in 1% formaldehyde (Sigma-Aldrich) for 10 min at room temperature and crosslinking was stopped with 0.125 M of glycine. Chromatin was fragmented by sonication using Bioruptor UCD-300 (Diagenode). Antibodies (FOXA2, RRID:AB_2262810; H3K27Ac, RRID:AB_2118291; 1–3 μg per ChIP reaction) were coupled to magnetic beads (Millipore) overnight at 4 °C. For immunoprecipitation, a 100-μl aliquot of sonicated chromatin was incubated overnight with antibody-coupled magnetic beads, washed, and eluted by incubation at 65 °C.

![UCSC genome browser image of GCKR in liver-derived HepG2 cells. Normalized ChIP-Seq counts are shown for H3K27ac and H3K4me1 enhancer marks and P300, RXRA, FOXA2, and MAFK TF binding. The GWAS catalog SNPs and reporter elements used in the transcriptional activity assays are depicted above the figure. The TF binding sites located within the 638-bp region around rs780094 are shown in close-up in the insert below the image.](image)

**Fig. 1** rs780094, rs780095, and rs780096 GCKR SNPs reside in a liver-specific enhancer region. **a** UCSC genome browser image of GCKR in liver-derived HepG2 cells. Normalized ChIP-Seq counts are shown for H3K27ac and H3K4me1 enhancer marks and P300, RXRA, FOXA2, and MAFK TF binding. The GWAS catalog SNPs and reporter elements used in the transcriptional activity assays are depicted above the figure. The TF binding sites located within the 638-bp region around rs780094 are shown in close-up in the insert below the image. **b** The location of FOXA2 and MAFK consensus motifs in HepG2 cells overlapping rs780094 as defined by Factorbook [54]. Position weight matrix (PWM) modeling shows that rs780094 is likely to disrupt the binding of MAFK, whereas both alleles are predicted to allow binding of FOXA2. **c** The location of MAFK consensus motifs overlapping rs780095. The disruption of MAFK motif is predicted by PWM modeling.
for 1 h with brief vortexing every 15 min. To completely re-vortex crosslinking, the eluted chromatin was incubated overnight at 65 °C with proteinase K (ThermoFisher). The fragmented chromatin was purified using a ChIP DNA Clean and Concentrator Kit (Zymo Research). qPCR was conducted using a custom TaqMan SNP Genotyping Assay for rs780094 (ThermoFisher). This assay uses allele-specific probes coupled to different fluorescent dyes in the same reaction mix for the quantitative detection of the alleles in a single sample. The rs780094 assay amplifies a 70-bp region around this SNP (Additional file 1: Table S1). We used the ddCt method to calculate the haplotype-specific enrichment for both input and immunoprecipitated chromatin over an intronic region of GRB10 (rs6943153; MIM 601523) for which no TF binding or H3K27Ac histone marks have been reported. Fold enrichment was determined by dividing the enrichment of the immunoprecipitated chromatin by the enrichment of the respective input. FOXA2 and H3K27Ac ChIP-qPCR experiments were performed in triplicate (three technical replicates in each experiment for the determination of basal H3K27Ac levels; two, two, and three technical replicates for FOXA2 binding experiments; three, two, and three technical replicates for H3K27Ac in the CRISPR assays).

**CRISPR-dCas9-VPR-mediated transcriptional activation**

To determine the direct effect of the enhancer on GCKR expression we used a CRISPR-dCas9-derived transcriptional activator system [17, 18]. This system comprises two plasmids, one expressing a nuclease-null Cas9 (dCas9) fused in tandem with three transcriptional activation domains (VP64, p65, and Rta [VPR]) [17], and the second one expressing the gRNA that directs the dCas9-VPR activator to the target site [18]. The gRNAs were

![Fig. 2](image-url)
designed using an online tool [19] and two gRNAs sequences, located adjacent to rs780094 in a SNP-free region, were selected for cloning. Briefly, forward and reverse primers were synthesized comprising the gRNA sequences provided by the online designer (Additional file 1: Table S1). The primer pairs were phosphorylated using T4 Polynucleotide Kinase (ThermoFisher) and annealed in the thermal cycler with the following program: 37 °C for 30 min; 95 °C for 5 min; ramp down to 25 °C at 5 °C /min. Double-stranded oligos were subcloned into pSPgRNA plasmid (a gift from Charles Gersbach; Addgene plasmid # 47108) digested with BbsI (ThermoFisher). The identity of the gRNA constructs was verified by sequencing. HepG2 cells were co-transfected with the activator VPR and each of the gRNA constructs or an empty gRNA vector in a 1:1 mass ratio. The constructs were expressed for 48 h and thereafter total RNA was extracted using the RNeasy Mini Kit (Qiagen). cDNA was prepared with random hexamer primers using a High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA, USA) according to the manufacturer’s instructions. GCKR mRNA and RPLP0 mRNA (housekeeping gene; MIM 180510) expression were measured by qPCR using specific TaqMan gene expression assays (ThermoFisher) in a 7500 Real-Time PCR System (Applied Biosystems). The relative gene expression was calculated using the relative ddCt-quantification. Four independent experiments were performed (with three, four, four, and three technical replicates, respectively).

Haplotype-specific expression in human liver biopsies
A total of 304 severely obese individuals participated in the ongoing KOBS study [20]. All subjects underwent Roux-en-Y gastric bypass surgery at the Kuopio University Hospital (Kuopio, Finland). The study protocol was approved by the Ethics Committee of Northern Savo Hospital District (Finland) and carried out in accordance with the Helsinki Declaration. Informed written consent was obtained from all participants. Liver tissue biopsies used in our study were collected during the operation. A total of 132 subjects heterozygous for rs780094, rs780095, and rs780096 were identified and included in this study (Additional file 2: Datasets). Total RNA from human liver tissue biopsies was extracted with the miRNeasy Mini Kit (Qiagen) and the cDNA was prepared as described above. Haplotype-specific expression was determined with the TaqMan SNP genotyping assay for rs780094, as previously reported [21]. The expression was quantified using the ddCt method and the relative expression between the haplotypes was calculated. The haplotype expressions were compared using a Mann–Whitney U test, and the associations between haplotype expression and metabolic traits were assessed by a Spearman’s rank correlation test with Bonferroni correction.

Analytical methods used in the human study
Plasma glucose, insulin, serum lipids and lipoproteins (total cholesterol, HDL cholesterol, and triglycerides), and FFAs were measured from fasting venous blood samples. Plasma glucose was measured by enzymatic hexokinase photometric assay (Konelab Systems Reagents; ThermoFisher). Plasma insulin was determined by immunoassay (ADIVA Centaur Insulin IRI, Siemens Medical Solutions Diagnostics). An oral glucose tolerance test (OGTT) was performed after a 12-h fasting period. Serum FFAs were assayed with an enzymatic colorimetric method (Wako NEFA C test kit; Wako Chemicals). Glucose, insulin, and FFA levels were determined at 0, 30, and 120 min with an OGTT.

Statistical analysis
In the luciferase assays, the technical replicates of each experiment (Fig. 2) were averaged and the mean value of the haplotypes per experiment was generated. We used the two-tailed t-test to compare the haplotypes in the same treatment, and one-way ANOVA to compare the haplotypes between the different treatments (insulin versus no insulin). Basal H3K27Ac and FOXA2 ChIP-qPCR data (Fig. 3a, b) were analyzed using technical replicates. Comparisons between the haplotypes of the same treatment were carried out with a two-tailed t-test and a
one-way ANOVA was used to compare the same haplotype between the treatments (Fig. 3b). For the comparisons of GCKR gene expression in the CRISPR activator experiments (Fig. 4a) we used one-way ANOVA with a Tukey’s post hoc test with the dCt values. The H3K27Ac ChIP-qPCR data (Fig. 4b) were compared using the enrichment values of technical replicates with a two-tailed t-test (comparison of two groups). No comparison between the haplotypes was performed for gR1 + VPR. The haplotype expression data from human liver biopsies (Fig. 5) were compared with a Mann–Whitney U test. A Spearman’s rank correlation test with Bonferroni correction was used to evaluate the association between haplotype expression and metabolic traits. In all cases, the number of biological and technical replicates per experiment is specified in the relevant section and in the corresponding figure legend.

**Results**

**Genomic and regulatory features of the rs780094 locus**

We investigated the chromatin features of the region containing rs780094, rs780095, and rs780096 by exploring the ENCODE Project data for histone modifications and TF binding [22]. These SNPs are in high LD ($r^2 = 0.86$) and located very close to each other (Fig. 1a). The data from the Roadmap Epigenomic Projects [23] suggest a high liver-specificity of this regulatory locus (Additional file 3), in line with the key role of the liver in contributing to the metabolic traits associated with rs780094. Therefore, we focused our analysis on the human hepatocellular carcinoma (HepG2) cell line. In these cells, the variants reside in a region enriched for the active, enhancer-associated H3K4me1 and H3K27ac histone marks (Fig. 1a). In addition, ChIP-seq data for TFs showed the binding of FOXA2, RXRA, and MAFK TFs to this locus, which are also bound to the GCKR promoter (Fig. 1a). An adjacent second peak for FOXA2, along with peaks for FOXA1 and p300, are located around 250 bp 3’ of rs780094 (Fig. 1a). According to the 1000 Genomes Project data [24], rs780094 is also in strong LD with the common variants rs1260326 ($r^2 = 0.91$) and rs780093 ($r^2 = 0.98$), which, however, reside outside the potential enhancer. Moreover, only the region containing rs780094, rs780095, and rs780096 has been found to bind TFs. Position weight matrix (PWM) modeling for rs780094, rs780095, and rs780096 [25] showed that rs780094 overlaps a FOXA2 motif, which is not expected to affect binding (Fig. 1b). In contrast, both rs780094 and rs780095 affect the MAFK motif, which is predicted to result in decreased binding (Fig. 1b, c). This suggests that differential binding of MAFK and possibly other factors could alter the enhancer activity of this locus. Altogether, the histone modification and TF binding data for the locus containing rs780094, rs780095,
and rs780096 suggest the existence of a highly liver-specific regulatory region consisting of two adjacent TF binding loci that could act together as a transcriptional enhancer.

### Haplotype-specific transcriptional activity of a locus defined by rs780094, rs780095, and rs780096

To investigate the transcriptional activity of the region defined by rs780094, rs780095, and rs780096 and its relationship to the TFs, we performed luciferase reporter assays in HepG2 cells and MPHs. We cloned two DNA inserts representing the two most common haplotypes formed by these variants in the METSIM study [11] (CGG/TAC; rs780094, rs780095, rs780096; Additional file 2: Data-sets) into a luciferase vector with a minimal promoter: a TAC; rs780094, rs780095, rs780096; Additional file 2: Data-sets) and then cloned two DNA inserts representing the two most common haplotypes defined by rs780094, rs780095, and rs780096 and its relationship to the TFs, we performed luciferase reporter assays in HepG2 cells and MPHs. We cloned two DNA inserts representing the two most common haplotypes formed by these variants in the METSIM study [11] (CGG/TAC; rs780094, rs780095, rs780096; Additional file 2: Data-sets) into a luciferase vector with a minimal promoter: a 638-bp insert containing rs780094, rs780095, and rs780096 and both the 5’ and 3’ TF peaks; and a 3-kb insert spanning a larger region 3’ and 5’ from rs780094 (Fig. 1a). The sequences of the 638-bp inserts differ only in rs780094 (C/T), rs780095 (G/A), and rs780096 (G/C), while the 3-kb inserts differ also in the rs780093 SNP (C/T), which, however, resides outside the potential regulatory region.

None of the inserts increased luciferase transcription compared to the control vector pGL4.10-TK when transfected alone in HepG2 cells (data not shown). This may indicate insufficiency of the endogenous TF pool to activate the luciferase constructs transfected in excess to the cells, rather than a lack of enhancer activity. To test this, we co-transfected the 638-bp constructs with either the empty vector pCMV6-XL5 or plasmids expressing the TFs that bind to this region according to the ENCODE project data (FOXA2, FOXA1, RXRA, or MAFK). When co-transfected with pCMV6-XL5, none of the haplotypes increased luciferase activity compared to the control vector pGL4.10-TK (Fig. 1a, left panel). These results are consistent with the lack of transcriptional activity when the constructs were transfected alone. Of all TFs tested, only FOXA2 increased luciferase transcription compared to the control vector (Fig. 2a, right panel; Additional file 1: Figure S1). Moreover, the haplotypes showed different transcriptional activity (CGG > TAC; 15%; p = 0.0003). Similarly, FOXA2 also increased the transcriptional activity of the 3-kb region in a haplotype-specific way (Fig. 2b, right panel; CGG > TAC; 19%, p = 0.003), while no luciferase activity was observed in cotransfection with the pCMV6-XL5 vector (Fig 2b, left panel). The magnitude of the haplotype differences as well as the overall activating effect were modest, which may be explained, at least in part, because the enhancer is taken out of its natural context for the reporter assays [26]. Yet, these results suggest the existence of transcriptional activity for this locus and a regulatory bias between the haplotypes.

Insulin has been reported to inactivate FOXA2 [27–29]. Therefore, we used insulin stimulation to investigate whether the effects on luciferase transcription were FOXA2-specific. As high FOXA2 expression is obtained only 24 h after transfection, our experiment requires long exposure to insulin. However, HepG2 cells become insulin resistant when cultured in high glucose and high insulin media for 24 h or longer [30–32]. Therefore, we opted for MPHs as a physiological model for chronic insulin response. Consistent with the results in HepG2 cells (Fig 2b), FOXA2 increased the transcriptional activity for both haplotypes compared to the control (Fig. 2c, right panel). In untreated cells, the CGG haplotype showed higher transcriptional activity than the TAC haplotype (difference 64%, p = 0.02). As expected, insulin reduced FOXA2-dependent transcriptional activity similarly for both haplotypes (a 60% reduction for CGG (p = 0.001) and a 50% reduction for TAC (p = 0.011) (difference between the haplotypes p = 0.05; Fig. 2c, right panel). This similar responsiveness to insulin between the haplotypes can be explained by the fact that both of them, although to different extents, are activated by FOXA2. Similarly to the results in HepG2 cells (Fig. 2b), the 3-kb constructs did not show transcriptional activity in co-transfection with the pCMV6-XL5 control vector (Fig 2c, left panel).

Although MAFK alone was not sufficient to activate reported gene expression (Additional file 1: Figure S1), the PWM for rs780094 and rs780095 showed that the alternative variants rs780094-C and rs780095-G are likely to disrupt the motifs for MAFK (Fig. 1b, c). Therefore, to investigate the relationship between FOXA2 and MAFK in the transcriptional activity of this locus, we co-transfected each of the 3-kb inserts with MAFK and pCMV6-XL5 or FOXA2 in HepG2 cells. Consistent with the results from the 638-bp inserts (Additional file 1: Figure S1), MAFK alone did not increase luciferase transcription from the 3-kb constructs (MAFK + pCMV6-XL5; Fig. 2d, left panel). However, co-transfection of MAFK and FOXA2 produced effective transcriptional activity for both haplotypes with respect to the control vector (Fig. 2d, right panel). Importantly, in contrast to FOXA2 overexpression, the co-transfection with MAFK led to the switching of haplotype activity to TAC > CGG, in agreement with the stronger binding motif of the TAC haplotype compared to the CGG haplotype (difference 26%, p = 0.004).

In summary, these results confirm that the GCKR intronic region containing rs780094, rs780095, and rs780096 is able to enhance transcription in a haplotype-specific way. Furthermore, our data support the notion that FOXA2 plays an important role in priming the enhancer for activity and suggest that the collaborative binding between TFs such as MAFK may determine the...
direction of the haplotype bias in response to specific cellular signals.

**Haplotype-specific H3K27Ac levels and FOXA2 binding**

HepG2 cells are heterozygous for rs780094, rs780095, and rs780096, which allows investigation of the haplotype-specific features of this region in a native chromatin context. However, HepG2 cells have also been reported to have a hyper-diploid karyotype with chromosome 2 gain [33]. We found that this gain results in a duplication of the CGG haplotype, which was taken into account in ChIP-Seq data normalization (Additional file 1: Figure S2). ChIP-seq data on HepG2 cells from the ENCODE Project suggest differential histone marks (Additional file 1: Figure S3a) and TF binding (Additional file 1: Figure S3a, b) between the rs780094 alleles. The C allele shows higher H3K4me1 and H3K4me2 histone marks as well as higher enrichment of P300, FOXA2, and FOXA1 TFs. Consistent with the PWM for rs780094 (Fig. 1b) and the reporter assay results (Fig. 2d), the T allele is suggested to have higher MAFK binding (Additional file 1: Figure S3b). Furthermore, a recent study [34] investigating enhancer SNPs capable of disrupting enhancer activity upon allelic change in HepG2 cells identified rs780094 as a dubbed deleterious enhancer (deSNP) with a high risk to cause phenotypic changes (scored 64.07 and ranked in the top 0.25% among all the tested HepG2 enhancer SNPs using unmasked genomic sequences; personal communication from Di Huang and Ivan Ovcharenko). The authors also provide evidence that deSNPs are more likely to display significant allele-specific binding, with 3.7% of H3K4me1-marked loci, 6.2% of H3K27ac-marked enhancers, 25.8% of FOXA2, 20.4% of FOXA1, and 16.5% of P300 binding sites displaying differential enrichment [34]. These results suggest that although the phenomenon of allele bias we detected for rs780094 is a rather rare event based on histone modifications (~5%), genetic variations affecting the recruitment of FOXA transcription factors affect a much larger fraction of enhancers in liver, suggesting an overall stronger impact on regulatory activity [34].

To confirm the ChIP-Seq results, we performed ChIP-qPCR experiments to determine the haplotype-specific levels of the H3K27Ac histone mark and FOXA2 binding in our batch of HepG2 cells using a custom TaqMan Genotyping Assay for rs780094. Although the genotyping assay amplifies a 70-bp region around rs780094, the resolution of the ChIP methods (~250 bp) is not sufficient to discriminate whether the differential binding is allele-specific only for rs780094, thus limiting our analysis to the haplotypes. In line with the active histone marks H3K4me1/2, the CGG haplotype also exhibited 1.4-fold higher enrichment of H3K27Ac (p = 0.001; Fig. 3a). Likewise, the CGG haplotype was shown to bind two times more endogenous FOXA2 than the TCC haplotype (Additional file 1: Figure S4). Upon overexpression of FOXA2, the binding of the CGG haplotype was further increased by 2.5-fold compared to the TAC haplotype (p = 0.01; Fig. 3b). Fluorescence imaging, western blotting, and qPCR verified the efficiency of transfection and the expression levels of FOXA2 in HepG2 cells (Additional file 1: Figure S5 and S6).

HepG2 cells become insulin resistant when cultured at high glucose [30–32]. However, they respond to acute treatment with insulin independent of glucose concentration, as indicated by substrate phosphorylation and cellular translocation of transcription factors [28, 35, 36]. Hence, we used HepG2 to investigate the specificity of FOXA2 binding between the haplotypes. Insulin stimulation led to a 2.3-fold reduction in the binding of the CGG haplotype to FOXA2 (p = 0.002; Fig. 3b), while there was also a trend towards a decrease in the binding of the TAC haplotype (Fig. 3b). These results are consistent with previous reports showing a reduction of FOXA2 binding to chromatin in response to insulin [37]. Taken together, these results confirm that the locus represented by the GWAS rs780094 lead SNP exhibits haplotype bias in histone marks and TF binding.

**CRISPR-dCas9-VPR-mediated activation of the enhancer locus regulates transcription of endogenous GCKR in HepG2 cells**

Recent evidence suggests that a significant number of enhancers do not regulate the nearest gene [38–40]. To investigate whether the enhancer indeed regulates GCKR expression, we targeted the CRISPR-dCas9-VPR (VPR) activator system to this locus [17]. Co-transfection of VPR with two guide-RNA (gRNA) plasmids [18] (VPR + gR1; VPR + gR2) targeted to regions common to both haplotypes (Additional file 1: Figure S7; Additional file 1: Table S1) in HepG2 cells increased total GCKR transcription by 1.4-fold, compared to a control gRNA plasmid without targeting gRNA (VPR + gEm) (p = 0.0001 for both VPR + gR1 and VPR + gR2; Fig. 4a). To verify that the increase in gene expression was a specific effect of the VPR activator, we also measured epigenetic changes at the targeted site. To this end, VPR + gR1 significantly increased the levels of the enhancer-associated H3K27Ac histone mark in both haplotypes compared to VPR + gEm (2.1-fold for CGG, p = 0.002; 2.4-fold for TAC, p = 0.0001; Fig. 4b; for better visualization of variations between the experiments see Additional file 1: Figure S8). Given that the CRISPR-targeted regions were identical between the haplotypes, we did not look at the differences between haplotypes upon activation with VPR + gR1. Nevertheless, in agreement with the haplotype bias in basal conditions (Fig. 3a), the CGG haplotype showed higher levels of H3K27Ac in the control,
non-targeting VPR + gEm (CGG > TAC, $p = 0.03$; Fig. 4b). These results provide direct evidence that $GCKR$ is a target gene for the enhancer.

**Haplotype-specific expression in human liver biopsies**

Having demonstrated that the enhancer regulates $GCKR$ transcription, we investigated $GCKR$ haplotype-specific expression in human liver biopsies from subjects that participated in the Kuopio Obesity Surgery (K OBS) study [20] (Additional file 1: Table S2). To determine whether the CGG and TAC haplotypes were expressed differentially, we selected 132 subjects heterozygous for both haplotypes and measured the haplotype expression levels with a Taqman genotyping assay for rs780094. This strategy allows measurement of the differences in haplotype expression within each sample. Our results indicated that the expression of the CGG haplotype was 18% higher than that of the TAC haplotype ($p = 2.92 \times 10^{-4}$; Fig. 5a). Moreover, when stratified by sex, we found that the haplotype differences were larger in men than in women (CGG > TAC; men 29%, $p = 4.24 \times 10^{-5}$, n = 44; women 12%, $p = 0.051$, n = 88; Fig. 5b). The haplotype differences were statistically significant in individuals with normal glucose tolerance (NGT; 18%, n = 75, $p = 0.001$) but only a non-significant trend was observed in individuals with T2D (17%, n = 57, $p = 0.68$) (Additional file 1: Figure S9). Additionally, we assessed the associations of the haplotype expression with metabolic traits in individuals with NGT. We found that higher expression of the CGG haplotype was associated with higher levels of fasting free fatty acids (FFAs; 0.454; $p = 4.24 \times 10^{-4}$; Fig. 5a). Moreover, when stratified by sex, we found that the haplotype differences were larger in men than in women (CGG > TAC; men 29%, $p = 4.24 \times 10^{-5}$, n = 44; women 12%, $p = 0.051$, n = 88; Fig. 5b). The haplotype differences were statistically significant in individuals with normal glucose tolerance (NGT; 18%, n = 75, $p = 0.001$) but only a non-significant trend was observed in individuals with T2D (17%, n = 57, $p = 0.68$) (Additional file 1: Table S3). These results demonstrate that the CGG and TAC haplotypes were differentially expressed in the liver and suggest that the haplotype-specific regulation of this locus affects metabolic outcomes.

**Discussion**

More than 100 common variants associated with T2D have been identified by GWAS [1]. However, the translation of these associations into molecular mechanisms has been limited. The vast majority of T2D GWAS variants map to the non-coding regions and recent studies suggest that they mainly function by altering the recruitment of transcription factors and thus leading to variable histone modifications [12, 41]. Identification of the genome-wide DNA regulatory elements by consortiums such as ENCODE [22] and FANTOM [42] has lately made it possible to explore the regulatory potential of variant loci identified by GWAS. In our study we investigated the functional role of a T2D-associated $GCKR$ intronic locus represented by the lead SNP rs780094 and proximal variants rs780095 and rs780096. We demonstrate the existence of a transcriptional enhancer in a region containing these three common SNPs that is regulated by FOXA2 and able to activate $GCKR$ expression. Moreover, we provide evidence for the haplotype-specific regulation of this region, which might contribute to the related metabolic traits and to the susceptibility for T2D.

First, we investigated the enhancer activity of the locus by luciferase reporter assays and ChIP. The associations of this locus with several metabolic traits and the risk of T2D have been previously reported for rs780094 [5, 8, 43–46]. However, the high LD and the proximity of rs780095 and rs780096 SNPs suggest that the potential functional effects might be affected by more than one variant in a haplotype-dependent way. Therefore, we selected the two most common haplotypes formed by rs780094, rs780095, and rs780096 (CGG/TAC) found in the METSIM study [11] for our transcriptional experiments. Our results show that the CGG haplotype, containing the T2D risk-increasing glucose rs780094-C allele, exhibited a higher level of transcriptional activity and higher enrichment for active chromatin marks. The transcriptional function of this site is supported by the observation that epigenetic activation of the native locus using the CRISPR-dCas9-VPR activator [17, 18] leads to an increase in total $GCKR$ expression, which also suggests that $GCKR$ is a direct target of the enhancer. Interestingly, of all the TFs tested in the luciferase experiments, only FOXA2 was able to activate the regulatory region in a haplotype-specific way (CGG > TAC). The pioneering function of FOXA2 may explain the role of this TF in activating the enhancer. Namely, FOXA2 has the ability to interact directly with the core histones to open condensed chromatin and facilitate the binding of other TFs that would otherwise not be able to access their binding sites [47, 48]. In agreement with the reported inactivating effect of insulin on FOXA2 [27, 28], we showed that insulin reduced FOXA2-dependent transcriptional activity of the enhancer in MPPs, which could be due, at least in part, to the reduction in FOXA2 binding to DNA upon insulin stimulation. Moreover, previous reports have shown that $GCKR$ transcription is activated in the fed state by high glucose via ChREBP [49, 50], whereas FOXA2 is activated by glucagon in the fasting state [37]. Thus, our results suggest a novel mechanism for the control of $GCKR$ transcription through a FOXA2-activated enhancer in a haplotype-specific way, which may also be contributed to by other TFs (Fig. 6). The relationship between glucagon, FOXA2, and the enhancer in the control of $GCKR$ transcription warrants further investigations.

The analysis of haplotype-containing TF motifs identified a MAFK motif disruption at rs780094 and rs780095 loci, which translated to decreased binding of MAFK to the rs780094-C allele in ChIP-Seq. Interestingly, co-
expression of FOXA2 with MAFK was able to reverse the transcriptional haplotype bias observed for FOXA2 (TAC > CGG), supporting a co-regulatory role of MAFK [51]. Like FOXA2, MAFK can directly bind DNA to facilitate the formation of transcriptional complexes in partnership with other proteins. A distinctive characteristic of small MAFs, in which MAFK is included, is the lack of a transcriptional activator domain, where the interacting partners and the nuclear concentration of MAFs dictate the direction of regulation (activation/repression) [51, 52]. In light of this knowledge, we can speculate that MAFK forms an interface for other transcriptional regulators to bind and control the activity of the enhancer. Altogether, this supports the current notion that collaborative binding between TFs [12] could impact the overall activity of the enhancer and the functional bias between haplotypes. Further studies are needed to address the specific cellular signals that converge in the activation of the enhancer through the network of TFs.

We also showed that the CGG haplotype exhibited higher expression compared to the TAC haplotype in human liver biopsy samples, providing evidence for a haplotype-specific effect on GCKR expression. Moreover, although the differences between the haplotypes were in the same direction, these were of higher magnitude and reached significance only in male subjects, which may have important functional relevance at the population level. Interestingly, the differences in haplotype expression followed the same trend in individuals with NGT and T2D. To our knowledge, changes in GCKR expression in T2D have not been reported. Therefore, our data suggest that the genotype may have a stronger effect on GCKR expression than the glycemic status. The effect of GCKR levels on the diabetic phenotype has been suggested by a study in which overexpression of GKRP in diabetic mice improved glucose homeostasis and increased insulin sensitivity [4]. Further studies are needed to investigate whether the haplotypes are regulated differently in T2D or whether they contribute differentially to the pathophysiology of the disease.

The overall effect of this enhancer on GCKR transcription as well as the haplotype bias are modest, which raises the question of whether such an effect is physiologically relevant. One possible answer to this question may lie in the key role that GKRP plays in controlling GCK availability and activity in the liver by a direct effect on the enzyme kinetics or sequestering it in the nucleus. Previous reports show that two-fold overexpression of rat GKRP is sufficient to reduce GCK translocation, affecting glucose phosphorylation, glycolysis, and glycogen synthesis [4]. Moreover, human GKRP is reported to be a more potent
GCK inhibitor than its rat homolog [6, 7], implying that the human GKRP–GCK system may be more sensitive to changes at the level of either GCK or GKRP. Therefore, as suggested previously [8, 50, 53], minor changes in GCKR transcription may alter the GCK/GKRP ratio, affect GCK activity, and impact glucose metabolism (Fig. 6). Another possibility is that multiple variants within the GCKR locus together contribute to phenotypic traits. Genetic fine mapping of a large LD region on chromosome 2 has identified the non-synonymous rs1260326 SNP as the strongest signal for blood glucose and triglyceride levels in this GWAS locus [5]. Functional studies on this variant demonstrated that the P446L amino acid substitution (rs1260326-T allele) had lower GCK sequestration capacity and lower response to fructose-6-P, affecting the kinetics and cellular localization of GCK [6, 7]. Therefore, it is possible that rs1260326 and the haplotypes formed by rs780094-rs780095-rs780096 account for a combined effect on GCK regulation by changing the inhibitory capacity of GRP and affecting the level of GCKR transcription, respectively (Fig. 6). The involvement of several mechanisms may thus amplify subtle functional defects from individual variants and produce a significant phenotypical effect. Last, it is also possible that, in addition to GCKR, the transcriptional enhancer regulates the expression of other genes implicated in different but functionally related pathways.

Conclusions
The active histone modification marks and the TF binding profiles at the T2D-associated GCKR intronic region defined by the high LD SNPs rs780094, rs780095, and rs780096 support the role of this region as a transcriptional enhancer. The SNPs within this locus were shown to overlap the binding motif of FOXA2 and disrupt the binding of MAFK transcription factors. Our data further demonstrate that FOXA2 governs the transcriptional activity of this region in a haplotype-specific way. The CGG haplotype, carrying the T2D-risk rs780094-C allele, exhibited higher reporter gene activity and FOXA2 binding compared to the TAC haplotype. However, co-expression of FOXA2 with MAFK inverted the transcriptional activity to favor the rs780094-T allele that allows higher binding of MAFK, suggesting that specific activating signals determine the direction of the haplotype bias. The epigenetic activation of this locus increased endogenous GCKR transcription, suggesting that GCKR is a direct target gene of the enhancer. We also showed that the CGG haplotype exhibited higher expression than the TAC haplotype in human liver biopsies, which correlated with higher levels of FFA. Our study reveals for the first time a functional mechanism for a T2D-associated locus represented by the lead SNP rs780094 and including the high LD variants rs780095 and rs780096 that might help in translating the phenotypical associations.

Additional files

**Additional file 1:** Supplementary Figures S1–S9 and Supplementary Tables S1–S3. (DOCX 1937 kb)
**Additional file 2:** Datasets. (XLSX 178 kb)
**Additional file 3:** Histone modifications at rs780094 from different tissues (data from Roadmap Epigenomics Project). (XLSX 17 kb)
**Additional file 4:** Raw sequencing data from HepG2, first read. (ZIP 436 kb)
**Additional file 5:** Raw sequencing data from HepG2, second read. (ZIP 470 kb)

**Abbreviations**
ChiP-qPCR: Chromatin immunoprecipitation followed by quantitative PCR; FBS: Fetal bovine serum; FFA: Free fatty acid; gDNA: Genomic DNA; GWAS: Genome-wide association study; IGV: Integrative Genomics Viewer; LD: Linkage disequilibrium; METSIM: Metabolic Syndrome In Men; MPH: Mouse primary hepatocyte; NGT: Normal glucose tolerance; OGTT: Oral glucose tolerance test; PWM: Position weight matrix; SNP: single nucleotide polymorphism; T2D: Type 2 diabetes; TF: Transcription factor.

**Funding**
This study was supported in part by grants from the Academy of Finland, Sigrid Juselius Foundation, Jane and Aatos Erkko Foundation, Finnish Foundation for Cardiovascular Research, Finnish Ministry of Education, Finnish Cultural Foundation, and the Post Doc Pool Foundation.

**Availability of data and materials**
The data generated in this work as well as the human genotypes and gene expression data can be found in the main manuscript or in Additional file 2: Datasets. The raw sequencing data for HepG2 cells genotyping generated with MiSeq desktop sequencer (Illumina) can be found in Additional files 4 and 5.

**Authors’ contributions**
MLR, MUK, and ML designed the study. MLR, DK, and KL performed the experiments and analyzed the data. MUK analyzed the ENCODE ChIP-seq data. DK and JP analyzed the KOBS study data. MLR, MUK, and ML wrote the manuscript in consultation with the other authors. All authors approved the final manuscript.

**Ethics approval**
All animal experiments in this study were approved by National Experimental Animal Board of Finland. The METSIM and KOBS study protocols were approved by the Ethics Committee of Oulu University Hospital District (Finland) and carried out in accordance with Helsinki Declaration. Informed written consent was obtained from all participants.

**Consent for publication**
Not applicable.

**Competing interest**
The authors declare that they have no competing interests.
40. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmida C, Suzuki T, et al. An atlas of active enhancers across human cell types and tissues. Nature. 2014;507:455–61.

41. McVicker G, van de Geijn B, Degner JF, Cain CE, Banovich NE, Raj A, Lewellen N, Myrthil M, Gilday Y, Pritchard JK. Identification of genetic variants that affect histone modifications in human cells. Science. 2013;342:747–9.

42. Lizio M, Hanishbaeger J, Shimoji H, Severin J, Kasukawa T, Saito S, Abogessaisa I, Fukuda S, Hori F, Ishikawa-Kato S, et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome Biol. 2015;16:22.

43. Sørensen V, Voight BF, Lysenkov B, Burt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Hughes TE, Groop L, Altshuler D, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science. 2007;316:1331–6.

44. Dupuis J, Langenberg C, Prokopenko I, Sørensen V, Soranzo N, Jackson AU, Wheeler E, Glazer NL, Bouatia-Naji N, Glyn AL, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nature Genet. 2010;42:105–16.

45. Tam CH, Ma RC, So WY, Wang Y, Lam VK, Germer S, Martin M, Chan JC, Ng MC. Interaction effect of genetic polymorphisms in glucokinase (GCK) and glucokinase regulatory protein (GCKR) on metabolic traits in healthy Chinese adults and adolescents. Diabetes. 2009;58:765–9.

46. Onuma H, Tabara Y, Kawamura R, Takata Y, Nishida W, Ohashi J, Kohara K, et al. The GCKR rs780094 polymorphism is associated with susceptibility of type 2 diabetes, reduced fasting plasma glucose levels, increased triglycerides levels and lower HOMA-IR in Japanese population. J Hum Genet. 2010;55:600–4.

47. Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, Zaret KS. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol Cell. 2002;9:279–89.

48. Sekiya T, Muthurajan UM, Luqier R, Tulin AV, Zaret KS. Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor Foxa. Genes Dev. 2009;23:804–9.

49. Ma L, Robinson LN, Towle HC. ChREBP*Mix is the principal mediator of glucose-induced gene expression in the liver. J Biol Chem. 2006;281:2721–30.

50. Blank V, Small Maf proteins in mammalian gene control: mere dimerization partners or dynamic transcriptional regulators? J Mol Biol. 2008;376:913–25.

51. de la Iglesia N, Mukhtar M, Seoane J, Guinovart JJ, Agius L. The role of the regulatory protein of glucokinase in the glucose sensory mechanism of the hepatocyte. J Biol Chem. 2000;275:10597–603.

52. Wang J, Zhuang J, Iyer S, Lin X, Whitfield TW, Greven MC, Pierce BC, Dong X, Kundaje A, Cheng Y, et al. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. Genome Res. 2012;22:1708–1812.

53. Machiela MJ, Chanock SJ. LDlink a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics. 2015;31(21):3555–7.