



Published in final edited form as:

Nat Commun. ; 6: 6069. doi:10.1038/ncomms7069.

Coordinated Regulatory Variation Associated with Gestational Hyperglycemia Regulates Expression of the Novel Hexokinase HKDC1

Cong Guo^{1,2}, Anton E. Ludvik³, Michelle E. Arlotto⁴, M. Geoffrey Hayes³, Loren L. Armstrong³, Denise M. Scholtens⁵, Christopher D. Brown⁶, Christopher B. Newgard^{4,7,8}, Thomas C. Becker^{4,8}, Brian T. Layden^{3,9}, William L. Lowe³, and Timothy E. Reddy^{2,10,*}

¹ Duke University Program in Genetics & Genomics, Durham, NC

² Center for Genomic & Computational Biology, Duke University School of Medicine, Durham, NC

³ Division of Endocrinology, Metabolism & Molecular Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL

⁴ Sarah W. Stedman Nutrition & Metabolism Center, and the Duke Molecular Physiology Institute, Duke University School of Medicine, Durham, NC

⁵ Department of Preventive Medicine, Division of Biostatistics, Northwestern University Feinberg School of Medicine

⁶ Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

⁷ Department of Pharmacology & Cancer Biology, Duke University School of Medicine, Durham, NC

⁸ Division of Endocrinology, Department of Medicine, Duke University School of Medicine

⁹ Jesse Brown Veterans Affairs Medical Center, Chicago, IL

¹⁰ Department of Biostatistics & Bioinformatics, Duke University Medical School, Durham, NC

Abstract

Maternal glucose levels during pregnancy impact the developing fetus, affecting metabolic health early and later in life. Both genetic and environmental factors influence maternal metabolism, but little is known about the genetic mechanisms that alter glucose metabolism during pregnancy.

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

* To whom correspondence should be addressed: tim.reddy@duke.edu.

Author Contributions

C.G. performed all reporter assays, knockdown and overexpression assays, cellular hexokinase assays, and related analysis. D.M.S. contributed to the design and analysis of allele-specific reporter assays. A.E.L. and B.T.L. purified HKDC1 and performed hexokinase assays. M.E.A., C.B.N., and T.C.B. designed and implemented adenoviral overexpression assays. M.G.H., L.L.A., and D.M.S. performed GWA imputation and related analyses. C.D.B. performed eQTL imputation and related analyses. D.M.S., B.T.L., C.B.N., T.C.B., W.L.L., and T.E.R. all supervised aspects of the manuscript. All authors contributed to the experimental design and writing of the manuscript.

Author Information

The authors have no competing financial interests with this work.

Here we report that haplotypes previously associated with gestational hyperglycemia in the third trimester disrupt regulatory element activity and reduce expression of the nearby *HKDC1* gene. We further find that experimentally reducing or increasing *HKDC1* expression reduces or increases hexokinase activity, respectively, in multiple cellular models; and that purified HKDC1 protein has hexokinase activity *in vitro*. Together, these results suggest a novel mechanism of gestational glucose regulation in which the effects of genetic variants in multiple regulatory elements alter glucose homeostasis by coordinately reducing expression of the novel hexokinase *HKDC1*.

Introduction

Regulation of glucose metabolism during pregnancy differs markedly from the non-gravid state as the mother must meet both her own and the growing fetus's energy needs. The differences are characterized by the combination of lower fasting glucose levels, increased hepatic glucose output, increased nutrient-induced insulin secretion, and significant insulin resistance by the third trimester¹. In 5-10% of pregnancies, however, glucose homeostasis is not maintained, resulting in hyperglycemia and gestational diabetes, i.e. diabetes that is first diagnosed during pregnancy. Gestational hyperglycemia is a major health risk to the mother, and is associated with adverse fetal outcomes including increased risk of type 2 diabetes (T2D) and obesity in the offspring²⁻⁵. Despite the prevalence and health risks, the mechanisms leading to gestational hyperglycemia remain largely unknown.

We recently identified a strong association between genetic variants in the first intron of *HKDC1* and 2 hr plasma glucose levels during an oral glucose tolerance test performed in a multiethnic cohort of 4,437 mothers at ~28 weeks gestation (F-test, $p = 8.26 \times 10^{-13}$, $\beta = 0.167 - 0.229$ [mmol L⁻¹])⁶. The association was replicated in a cohort of 2,192 additional European mothers and two smaller independent European and Canadian cohorts (n = 228 and 606). In a separate study of ~47,000 non-gravid individuals of European ancestry, the same region had only marginal association with 2 hr plasma glucose, suggesting that the association with *HKDC1* is largely pregnancy specific⁷. In the 1,000 Genomes Project database (1KGP), there are 60 variants in the coding exons of *HKDC1*. Seven of those variants are common, defined by a minor allele frequency > 1%. According to computational predictions of the effects of those variants on HKDC1 activity, between one and four of the common variants may impact the function of HKDC1^{8,9} (Supplementary Data 1). However, none of the coding SNPs were associated with 2 hr plasma glucose at genome-wide significance ($p < 1 \times 10^{-8}$). While the function of HKDC1 is unknown, the gene is broadly expressed including in liver and β -islet cells^{6,10}; conserved across vertebrates; and may be a novel human hexokinase based on its sequence similarity to hexokinase I (*HK1*)¹¹.

Based on these data, we hypothesized that genetic variation that alters gene regulation may contribute to the association with gestational hyperglycemia. Here we show that there are multiple variants that alter regulatory element activity in the region. Moreover, the effects of those variants are coordinated across four enhancer elements in the associated *HKDC1* locus. For each variant, alleles associated with reduced expression also associate with higher

plasma glucose in mothers. We further show that modulating HKDC1 levels alters hexokinase activity in multiple cellular models, and that purified HKDC1 protein has hexokinase properties *in vitro*. Together, our results support a novel mechanism of gestational glucose regulation in which coordinated variation across multiple enhancer elements within regulatory haplotypes reduces expression of the novel hexokinase HKDC1.

Results

Identification of regulatory variants in the *HKDC1* locus

The strongest genetic association with 2 hr plasma glucose was located within the first intron of *HKDC1* at rs4746822. That variant and variants in moderate linkage disequilibrium (LD) with rs4746822 ($r^2 > 0.3$) lie within a region of the genome exhibiting chromatin modifications consistent with active gene regulation across diverse tissues (Supplementary Figs. 1 and 2, Supplementary Table 1). To investigate the potential for regulatory variation contributing to the association with gestational hyperglycemia, we focused on a ~30 kb region defined both by the observed pattern of LD in the region and by evidence of gene regulation. As shown in Supplementary Fig. 2, there is little evidence of regulatory activity immediately flanking the target region, increasing our confidence that any regulatory variants contributing to the genetic association will be evaluated. There were three genotyped variants and nine imputed variants in the target region that were associated with 2 hr plasma glucose at genome-wide significance ($p < 1 \times 10^{-8}$, Supplementary Data 2). Together, those data led us to hypothesize that rs4746822 or variants in LD with rs4746822 influence maternal glucose metabolism by altering the activity of regulatory elements that control *HKDC1* expression.

Within the target region, we identified 11 candidate regulatory elements based on increased chromatin accessibility across 16 tissues including the metabolically relevant liver stellate cells and pancreatic islet β cells, (Fig. 1a, Supplementary Table 1)^{12,13}. The candidate regulatory regions account for 8.5 kb (28%) of the nucleotides in the 30 kb target region. Of the 425 variants in the target region identified by the 1KGP, 132 (30%) were in the candidate regulatory elements¹⁴. There were 203 total haplotypes of the individual regulatory elements, 60 of which were common in the 1KGP population (haplotype frequency $> 1\%$). The 60 common haplotypes accounted for 98% of all of the haplotypes for the regulatory elements sequenced by the 1KGP (Supplementary Table 2).

To determine if genetic variation in the identified regulatory elements influences endogenous *HKDC1* expression, we compared the variants associated with gestational hyperglycemia to those associated with gene expression changes in an expression quantitative trait locus (eQTL) study performed in primary human livers¹⁵. To do so, we imputed the tag variants from both the GWA and eQTL studies to the variants identified in the 1KGP¹⁴. That analysis revealed four significant eQTLs for *HKDC1* [$\log_{10}(\text{Bayes Factor}) > 2.5$] within DNaseI hypersensitive sites (DHS) that were also associated with maternal 2 hr glucose levels (Fig. 1b, Supplementary Fig. 2, Supplementary Data 3). We did not find evidence that the same variants also influence expression of nearby *HK1* in the same study (Supplementary Data 4), and known *HK1* eQTLs in other cell types do not overlap the variants associated with maternal glucose levels¹⁶. When we expanded our eQTL analysis to

all genes within 500 kb of *HKDC1*, we did not find any evidence that variants associated with 2 hr maternal glucose levels were also eQTLs for those genes (Supplementary Fig. 3). Together, these results suggest that genetic variants associated with gestational hyperglycemia near *HKDC1* alter *HKDC1* expression.

To identify causal variants that control *HKDC1* expression, we used luciferase reporter assays to measure allele-specific regulatory activity of the candidate regulatory elements in the region in the HepG2 liver cell line. We identified 1KGP subjects that maximized the genetic diversity in the candidate regulatory elements, and used PCR to clone the regulatory elements from those individuals upstream of a Simian Virus 40 promoter driving expression of a luciferase reporter gene. From the genomes of 19 individuals (Supplementary Table 3), we cloned 60 different naturally occurring haplotypes representing 93% of the haplotypes of the 11 regulatory elements that were identified by the 1KGP. To disentangle the effects of individual variants on expression that do not segregate in the population and to increase our power to detect regulatory effects of variants that are rare in the population, we used site directed mutagenesis to generate reporter constructs for 45 additional haplotypes (Supplementary Data 5). By balancing the number of observations of each allele, we were able to alleviate some of the loss in power due to allele frequency differences (Supplementary Table 4). In total, we assayed 105 haplotypes composed of 57 variants. Nine of the 11 regulatory elements had enhancer activity, with relative luciferase expression between 1.9 to 14-fold over control (Supplementary Fig. 4). Enhancer strength was positively and significantly correlated with average DHS peak intensity in HepG2 cells (Spearman $\rho = 0.64$, $p < 0.05$, Fig. 1b), however, we did not find significant correlation between DHS peak width and enhancer strength (Supplementary Fig. 5). We do not take these results as evidence that DHS peak intensity generally predicts enhancer strength, only that DHS sites are enriched for active regulatory elements.

To identify individual genetic variants in the regions that altered enhancer activity, we used a multiple linear regression model to estimate the effect of each variant acting independently on reporter expression (Supplementary Data 6, Supplementary Fig. 6). That analysis revealed 14 variants distributed across six regulatory elements that significantly altered gene expression (Bonferroni-adjusted $p < 0.01$, Fig. 1c, Table 1, and Supplementary Table 5). In agreement with other studies, we found that the effect of a single variant on enhancer activity is generally < 2 fold¹⁷⁻¹⁹. We next determined if the estimated effects of individual variants within an element behaved cumulatively. For each of the six elements containing regulatory variants, we quantified the expected luciferase values for each haplotype by adding the coefficients based on each variant within the haplotype and plotted those against the observed luciferase values. In all six elements, predicted luciferase values were positively and significantly correlated with observed luciferase values (Pearson $r^2 > 0.25$, $p < 0.0001$, Supplementary Fig. 7). Although we observe substantial cumulative effects, we cannot dismiss the possibility that more complex interactions also exist.

The results from the allele-specific reporter assays were positively validated in the liver eQTL study. For 10 out of 12 variants present in the eQTL analysis, the effect observed in our allele-specific reporter assays was in the same direction as predicted by the eQTL association ($p = 0.02$, binomial test). Among the validated regulatory variants were four

genome-wide significant eQTLs (Fig. 1d, Table 1). The observed effects of those four variants in our reporter assays were not specific to HepG2 cells. Specifically, in seven of the eight cases in which one of the variants significantly altered regulatory element activity in a different cell type, the direction of the effect was consistent with that observed in HepG2 cells (Supplementary Table 6). Together, these results indicate that rs10762264, rs4746822, rs2394529, and rs9645501 are regulatory variants in four separate enhancers that contribute to *HKDC1* expression in human liver.

The identified regulatory variants have coordinated effects

In order for the regulatory variants identified to collectively contribute to maternal glycemia, we hypothesized that there would be coordination between allele-specific regulatory element activity and maternal 2 hr glucose levels. Supporting our hypothesis, all four variants that were significant eQTLs for *HKDC1* in liver were associated with maternal 2 hr glucose levels with three reaching genome wide significance in the imputed GWA data (Table 1). All four alleles that were associated with increased 2 hr glucose also had decreased luciferase expression in the reporter assays and decreased *HKDC1* expression in the eQTL analysis (Fig. 1d, Table 1). We expanded the analysis to include 12 genome-wide significant SNPs associated with 2 hr glucose and found that each of the 12 SNPs associated with increased glucose also associated with decreased *HKDC1* expression ($p = 0.0002$, binomial test, Supplementary Data 7). We further expanded the analysis to include all 178 tested SNPs within the 30 kb locus and found the risk allele to decrease *HKDC1* expression in 131 cases ($p = 1.14 \times 10^{-10}$, binomial test, Supplementary Data 7). Together, those results support a model in which multiple genetically linked regulatory variants have a coordinated effect of reducing *HKDC1* expression in women with higher gestational glucose levels. Exceptions to the coordination were limited to variants that were not significantly associated with 2 hr plasma glucose levels or *HKDC1* expression and that were not in strong LD with the lead GWA variant. For rs7089277, the minor allele frequency is low (7%) and therefore the variant may have limited effects in the population, and there is only nominal association with 2 hr glucose ($p = 8.02 \times 10^{-5}$). The 2 SNPs in region X have large luciferase β 's, but were not significantly associated with 2 hr glucose ($p = 0.0183$ and 0.0184), were not in LD with rs4746822 ($r^2 = 0.077$), and did not significantly associate with *HKDC1* expression. Furthermore, reporter assays in three additional cell types found cell-dependent effects for the variants in region X (Supplementary Table 7). Together, these results indicate that there are strongly coordinated regulatory effects across the genomic region associated with 2 hr plasma glucose levels, and that exceptions to that coordination likely do not contribute to the expression of *HKDC1*.

HKDC1 is a novel human hexokinase gene

Having demonstrated that alleles associated with higher maternal 2 hr glucose levels decrease *HKDC1* expression, we next sought to understand whether *HKDC1* has hexokinase activity that may explain its contribution to glucose homeostasis. Phosphorylation of glucose by hexokinase is the first step in glucose metabolism. Screens for hexokinase activity in rat cells have identified four distinct hexokinases using electrophoresis and chromatography. Those genes have been mapped, and are known as *HK1*, *HK2*, *HK3*, and glucokinase

(*GCK*)²⁰⁻²². Reduced hexokinase activity has previously been associated with metabolic phenotypes including diabetes^{23,24}.

To determine if *HKDC1* contributes to cellular hexokinase activity, we first used siRNA-mediated knockdown to model the genetically reduced levels of *HKDC1* expression present in women with higher 2 hr glucose concentrations, and measured the effect of that reduction on cellular hexokinase activity. Targeting *HKDC1* with two different siRNAs individually and together in HepG2 cells, we reduced mRNA expression by 35–60%. Reduced *HKDC1* expression resulted in a dose-dependent decrease in cellular hexokinase activity (Fig. 2a). The siRNAs were specific and did not alter expression of the *HK1*, *HK2*, or *GCK* (Fig. 2b). *HK3* expression was not detectable in HepG2 cells, as expected¹². The magnitude of reduction in hexokinase activity indicates that about half of the hexokinase activity in HepG2 cells is due to *HKDC1*. The result agrees with RNA-seq data showing that half of the hexokinase mRNA expression in HepG2 cells is *HKDC1* (Supplementary Table 8). Providing further evidence that *HKDC1* contributes to cellular hexokinase activity, we also found that overexpression of *HKDC1* mRNA by transient transfection of an *HKDC1* expression plasmid increased cellular hexokinase activity (Fig. 2c and 2d).

To demonstrate that the activity of *HKDC1* was not specific to HepG2 cells, we transduced the rat pancreatic β cell line, INS-1, with an adenovirus that expressed *HKDC1* from a human cytomegalovirus promoter, and measured hexokinase activity in the cell lysates at glucose concentrations ranging from 0-50 mM²⁵. The predominant hexokinase in INS-1 cells is *GCK*, and the cells have little or no detectable low K_m hexokinase activity²⁵. Cells transduced with the *HKDC1* adenovirus had substantially increased *HKDC1* protein (Fig. 2e, Supplementary Fig. 8), and increased hexokinase activity across the concentration range (Fig. 2f, Supplementary Fig. 9). Moreover, the shift in the dose-response curve suggests that *HKDC1* has a lower K_m than *GCK*. Adenoviral-mediated *HKDC1* overexpression did not alter the expression of *HK2*, *HK3*, or *GCK* and decreased *HK1*, demonstrating that the virus specifically overexpressed *HKDC1* (Supplementary Fig. 10). Transduction also did not result in decreased cell density (Supplementary Fig. 11). We repeated the experiment with independently purified virus and independently grown INS-1 cells, with similar results (Supplementary Fig. 12). Together, these results demonstrate that *HKDC1* contributes to cellular hexokinase activity either through direct hexokinase activity or through modulating the activity of the other hexokinases.

Based on the extent of nucleotide and amino acid similarity between *HKDC1* and members of the hexokinase family, Supplementary Data 8 and 9, Supplementary Table 9 and 10), we hypothesized that the effects of *HKDC1* expression on cellular hexokinase activity are in part the result of the direct enzymatic activity of *HKDC1*. To test that hypothesis, we purified *HKDC1* and compared its specific activity to that of purified *HK1*. Protein purity was confirmed via a Coomassie blue stain (Fig. 2g, Supplementary Fig. 8). Results of the assays indicate that *HKDC1* protein alone has hexokinase activity, with 20% of the specific activity of *HK1* under the same conditions (Fig. 2h). Because our results show *HKDC1* hexokinase activity *in vitro* and *in vivo*, we propose renaming the *HKDC1* gene to *HK5*.

Discussion

We have shown that regulatory variation associated with gestational glucose metabolism alters expression of a novel human hexokinase. This result adds to the growing empirical evidence that regulatory variants contribute to a variety of common complex human phenotypes²⁶⁻²⁸. Moreover, our results provide empirical evidence that regulatory variants spanning multiple enhancers have a coordinated allelic effect on *HKDC1* expression. A recent analysis suggests that such coordinated effects may be a common mechanism by which regulatory variants influence gene expression²⁶. Such coordinated disruption of regulatory haplotypes may also explain how modest effects of individual regulatory variants¹⁷ can together have a sufficiently strong effect so as to be detectable in a genome-wide genetic analysis. The burden of multiple genetic variants disrupting regulatory haplotypes may therefore help explain the overabundance of non-coding genetic variants in GWA studies. Notably, however, independent effects were not observed in our analysis of the liver eQTL data. This negative result may be explained by differences in effect size, limited power between closely linked variants, and tissue-specific effects (Supplementary Fig. 13). Together, we take the results presented here to be consistent with a model of coordinated regulatory variation, and we emphasize the need for additional investigation. In part, challenges of strong association between variants and heterogeneous effect sizes may be overcome with high-throughput empirical investigation of allele-specific regulatory activity in large populations^{17,29}.

Importantly, the plasmid-based reporter assays used to detect regulatory variants here are likely biased towards detecting regulatory elements that act independent of genome context. Additional elements may be active when in the native genomic context, where interactions between multiple clustered regulatory elements contribute to effects on nearby gene regulation. Similarly, determining the direct effects variants detected here on endogenous gene expression remains a challenge. Addressing that possibility will require additional studies that include genomic context, potentially through direct modification of the genome via genome editing^{30,31}.

Studying regulatory effects in the locus associated with maternal glycemia also led us to the discovery of a novel human hexokinase that appears to have important metabolic effects during pregnancy. Previous screens to identify vertebrate hexokinases^{32,33} have identified HK1, HK2, HK3, and GCK, but did not identify HKDC1. One possible reason why HKDC1 was missed previously is that a high degree of structural similarity between HKDC1 and HK1 may have obscured separation in chromatography and electrophoresis. We also found that purified HKDC1 has reduced hexokinase activity when compared to HK1 in our in vitro assays. One possibility is that additional co-factors could be required for full HKDC1 activity in vitro, but that those factors are lost during the purification process. Such diminished hexokinase activity may have prevented earlier detection of HKDC1 as a hexokinase by less sensitive methods.

The tissues that are most relevant for HKDC1's role in meeting the metabolic demands of pregnancy are not yet known, nor is it known how *HKDC1* is regulated in response to pregnancy hormones. We expect that future evaluation of these questions will provide new

insights into the metabolic changes that occur during pregnancy. More broadly, the results from this study demonstrate the value of identifying and pursuing the targets of noncoding phenotype-associated genetic variants for revealing novel mechanisms of human disease³⁴.

Methods

Luciferase Reporter Assays

Selected regions were amplified from genomic DNA from 1KGP subjects and In-Phusion cloned at the NheI cut site into pGL4.13 luciferase expression vector (Promega). The construct was then transformed into TOP-10 competent cells and plated onto LB agar plates with ampicillin and incubated overnight at 37°C. In order to capture both haplotypes from subjects who were heterozygous in those regions, multiple colonies were selected and grown individually in LB media overnight. Plasmids were extracted using the PureYield Plasmid Miniprep System (Promega). Constructs were sequenced using Sanger sequencing and variants were confirmed in dbSNP³⁵. Site directed mutagenesis was carried out by amplifying the plasmid using primers containing the allele of choice, treating with DpnI, and transforming into TOP-10 competent cells. Primers for cloning and site directed mutagenesis are listed in Supplementary Tables 11 and 12. HepG2 cells (Duke Cell Culture Facility) were plated into white flat-bottom 96-well plates at a density of 25,000 cells/well. After 48 hr, 100 ng of each construct/well was transfected with Fugene HD (Promega) at a 5.5:1 Fugene:DNA ratio. At least 16 biological replicates for each construct were transfected. After 24 hr, luciferase and *renilla* luciferase signal were quantified using the Dual-glo Luciferase Assay (Promega) using a Victor3 1420 plate reader (PerkinElmer). Luciferase values were normalized by dividing the luciferase signal by the *renilla* signal. A linear regression model implemented in the R statistical package (R function `lm`) was used to determine the regulatory effect of individual SNPs where (normalized luciferase intensity) = $\beta_1 \text{ SNP}_1 + \beta_2 \text{ SNP}_2 + \beta_3 \text{ SNP}_3 \dots + \varepsilon$. Raw data for luciferase reporter assays are provided as supplementary datasets.

RT-qPCR

RNA was isolated using the RNeasy kit (Qiagen). 2 µg of RNA per reaction was reverse transcribed using the SuperScript Vilo cDNA Kit (Life Technologies). All qPCR reactions were performed in biological triplicates and technical duplicate using the PerfeCTa qPCR Fastmix (Quanta) on an ABI StepOnePlus cycler. Reactions were cycled at 95°C (10 s) and the primer annealing temperature (30 s) for 40 cycles. Calculations were performed using the C_1 method using β -actin as a reference control. Primer sequences are listed in Supplementary Table 13.

Western Blots

Transduced cell pellets were sonicated for 5 × 30 s pulses in 100 µL of RIPA buffer supplemented with protease inhibitors. The protein concentrations were quantified using a Bicinchoninic acid (BCA) assay (Thermo). The lysate was then mixed 1:1 with loading buffer supplemented with 5% 2-mercaptoethanol and incubated at 76°C for 15 m. 10 µg of protein along with the Precision Plus Protein Dual Color Standard (Biorad) was loaded into each lane of a Mini-Protean TGX 4-20% gel (Biorad) and run at 100 V for 1 hr at 4°C. The

membrane was blocked with 5% skim milk for 30 m and washed 3 times in 0.1% Tween-20 in PBS. The membrane was incubated overnight in anti-HKDC1 antibody produced in rabbit (Sigma) diluted 1:500 in skim milk. The membrane was washed 3 more times and incubated in horseradish peroxidase (HRP) conjugated anti-rabbit antibody diluted 1:2500 in skim milk for 75 m at room temperature (Santa Cruz Biotechnology). The membrane was washed 3 times and incubated at room temperature in 10 μ L SuperSignal West Pico Chemiluminescent Substrate (Thermo) for 15 m. After exposure and imaging, the membrane was stripped using Restore Western Blot Stripping Buffer (Thermo) and blotted using 1:500 anti β -actin antibody produced in mouse (Santa Cruz Biotechnology) and 1:2500 HRP conjugated anti-goat antibody (Santa Cruz Biotechnology) diluted in skim milk.

Adenovirus Construction and Transduction

A 2,789 bp BamHI/NotI fragment from plasmid RC221178 (OriGene), containing the 2,750 bp HKDC1 open reading frame, was cloned into the BamHI/NotI sites of the adenoviral shuttle vector pShuttle-CMV (Agilent). *HKDC1* adenoviral shuttles were linearized with PmeI and transformed into BJ5183-AD1 cells (Agilent). Recombinant clones were isolated, digested with PacI, and recombinant adenoviruses were generated by transfecting HEK293 cells (Clontech) with the PacI digested recombinant plasmid DNA and FuGene³⁶. To generate the HK1 adenovirus, a 3,600 bp DNA fragment containing the rat HK1 ORF was cloned into the shuttle vector pAC.CMV.pLpA. Recombinant HK1 virus was then generated by co-transfecting HEK293 cells with HK1 shuttle clone and plasmid pJM1⁷²⁵

After transfection, viral lysate was collected, further amplified in HEK293 cells plated in 5 \times p150 tissue culture plates. Prior to complete lysis, media was removed, infected cells harvested in 2 mL Freeze/thaw buffer (10 mM Tris/HCl, pH 8.0, 1 mM MgCl₂), lysed by 2 freeze/thaw cycles, and the virus purified by ultracentrifugation on a CsCl gradient. Purified virus was de-salted with a 7k MWCO column (Thermo Scientific) equilibrated with freeze/thaw buffer, and glycerol was added to a final concentration of 10%. Viruses were titered by measuring OD₂₆₀ (1 OD₂₆₀ = 1.1 \times 10¹² virions mL⁻¹) as well as by plaque assay in HEK293 cells.

To prepare cells for transduction, INS-1 cells were first seeded at a density of 7 \times 10⁵ per well into 12-well plates. INS-1 832/13 cells were kindly provided by Christopher Newgard (DMPI, Duke University)³⁷. After 24 hr, the cells were transduced with *HKDC1*-, *HK1*-, or *GFP*-expressing adenovirus in triplicate. Cells were grown for another 24 hr and pelleted with centrifugation.

Protein Purification

Rosetta 2 (DE3) (EMD Millipore) cells were transformed with pReceiver-B01 plasmid containing either *HKDC1* or *HK1*. A single colony was grown overnight in 100 mL TPM media (20 g L⁻¹ Tryptone, 15 g L⁻¹ Yeast extract, 8 g L⁻¹ NaCl, 2 g L⁻¹ Na₂HPO₄ 1 g L⁻¹ KH₂PO₄) augmented with 50 μ g mL⁻¹ ampicillin and 34 μ g μ L⁻¹ chloramphenicol at 37°C. The pregrowth culture was added to 500 mL of the same media and grown to an OD of ~0.6. Isopropyl β -D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM, and the cells were grown for an additional 4 hr at room temperature.

Cultures were centrifuged at $10,000 \times g$ for 30 m and resuspended in lysis buffer (50 mM NaH_2PO_4 , 300 mM NaCl, 0.25% Tween-20, 5% sucrose, 5% glycerol, 2 mM imidazole, and 10 mM 2-mercaptoethanol) supplemented with protease inhibitors (1 mM phenylmethanesulfonyl fluoride [PMSF], $2 \mu\text{g mL}^{-1}$ aprotinin, $0.5 \mu\text{g mL}^{-1}$ leupeptin, $0.7 \mu\text{g mL}^{-1}$ pepstatin A). Cells were sonicated for 5 m followed by DNase I treatment (4 U mL^{-1}) for 15 m on ice. The lysates were clarified by centrifugation at $10,000 \times g$ for 30 m and the supernatant passed over a HisTALON column (Clontech). The column was washed with 25 mL of wash buffer (Lysis buffer pH 7.0, 25 mM imidazole). The protein was eluted from the column in 1 mL fractions using wash buffer with 500 mM imidazole. DTT was added to each fraction to a final concentration of 1 mM.

Cellular Hexokinase Assays

To measure cellular hexokinase activity in knockdown and overexpression experiments, we used an absorbance-based hexokinase assay (Sciencell #8408). Cell pellets were sonicated in 200 μL of 1% Triton at 4°C for 5×30 s pulses at medium intensity using a Bioruptor (Diagenode). The lysate was centrifuged at $1000 \times g$ for 5 m to remove remaining cell debris. The assay was performed by combining 80 μL of reaction mix (42 μL assay buffer, 20 μL cofactor, 2 μL developer, 6 μL enzyme mix, and 10 μL NADP), 10 μL of cell lysate, and 10 μL of 0-500 mM glucose was in each well of clear 96-well microplates. The plates were incubated in the dark for 90 m at room temperature. Absorbance at 490 nm was quantified in technical duplicates using a Victor3 1420 plate reader (PerkinElmer).

Purified HKDC1 Hexokinase Assays

To assay hexokinase (HK) activity of purified HKDC1, 2 μg of purified protein was diluted in 20 μL of HK dilution buffer containing 20 mM KH_2PO_4 , 100 mM KCl, 1 mM MgCl_2 , 1 mM EDTA, 1 mM dithiothreitol (DTT), 60 g L^{-1} glycerol, and 1 g L^{-1} bovine serum albumin. Samples were loaded onto a microplate and mixed with 100 μL of reaction buffer containing 50 mM HEPES pH 7.4, 100 mM KCl, 8 mM MgCl_2 , 5 mM ATP, 0.5 mM NADP, 1 U mL^{-1} G6PDH (from *Leuconostoc mesenteroides*), 1 mM DTT, 1 g L^{-1} bovine serum albumin, and 10 mM D-(+)-glucose. Reactions were incubated at 37°C for 1 hr and stopped with 174 μL buffer C (0.46 mM SDS pH 8.0, 300 mM NaH_2PO_4 , pH 8.0). Fluorescence of NADPH was measured at excitation wavelength of 340 nm and an emission wavelength of 450 nm. Specific activity was determined using an NADPH standard (Supplementary Fig. 14).

Multiple Sequence Alignment

Multiple sequence alignments and percent identity matrices were constructed using Clustal Omega using default settings <http://www.ebi.ac.uk/Tools/msa/clustalo/>³⁸. Protein and DNA sequences were downloaded from UCSC genome browser³⁹.

siRNA Knockdown

HepG2 cells were plated at 1×10^5 cells/well into 24-well plates. After 48 hr, cells were transfected with either siRNAs targeting HKDC1 (Ambion Silencer Select s37044, s37046) or siRNA not known to target any genes (Ambion Silencer AM461) using 6 pM of siRNA

and 1 μ L of Lipofectamine RNAiMax transfection reagent (Life Technologies) per well. Each siRNA was transfected in biological triplicate. Cells were harvested 24 hr later for RT-qPCR and hexokinase assays as described above.

Genome Wide Association with Imputed Genotypes

Genotypes from the previously published GWA study that identified the *HKDC1* locus⁶ were imputed separately in each of the four quality-control (QC) cleaned and filtered genotyping sets using IMPUTE2 v2.3.0 and the 1KGP reference panel (December 2013 release). A cosmopolitan reference panel of unrelated individuals of the Africa (“AFR”), Americas (“AMR”), Asia (“ASN”), and Europe (“EUR”) populations was used. The strand-checking utility of SHAPEIT v2 was used to ensure consistent strand assignments between the reference data set and the QC cleaned and filtered data sets. Strand was corrected as indicated, and SNPs for which strand could not be resolved were removed. A conservative info threshold (synonymous to allelic r^2) of 0.9 was used to remove questionable imputed SNPs.

The genotype call probabilities from the filtered IMPUTE2 output were used in a linear regression model between each of the phenotypes and the genotypes probabilities under an additive model adjusting for age, BMI, and the first two principal components of ancestry. The frequentist approach in SNPTTEST v2.4.1 was used to estimate the beta and standard errors for each regression model and assess the significance of the association between the SNP and the phenotype of interest.

Liver eQTL analyses

Liver gene expression and genotype data were generated as follows¹⁵. Gene expression array probes were aligned to the human reference genome (hg19) and gene models (RefSeq). Probes lacking unique genomic or transcriptomic alignments were discarded. Furthermore, probes overlapping common polymorphisms (minor allele frequency > 5%, based on 1KGP pilot release data) were discarded. Gene expression array feature intensities were extracted from arrays, background subtracted and \log_2 transformed. Missing data was imputed by k nearest neighbors (R package impute, function impute.knn, $k = 10$). The distribution of expression measurements on each array was normalized to the average empirical distribution across all arrays (R package limma, function normalizeBetweenArrays, method quantile). For each probe and across arrays, the distribution of expression values was transformed to the quantiles of the standard normal distribution (R function qqnorm). This matrix of processed gene expression values was then subjected to PCA (R package pcaMethods, function pca). We controlled for the effect of these PCs by taking the residuals of linear models using the first 15 PCs as covariates. Values from replicate arrays were then averaged. Because this sample set included both European and African American individuals, residual expression values were then transformed to the quantiles of the standard normal distribution, within each population, and then pooled. Given the mappings of probes to gene models, if there were multiple probes that mapped to the same gene, probes were clustered [R package mclust, function Mclust($y, G=1:\min(4, \text{probe_count})$)] and the mean per cluster per individual was estimated.

These residuals were then transformed by gene to standard normal. These data were used as phenotypes for the eQTL scan.

Genotype data were processed and quality controlled as follows¹⁵. Genotyping was performed on the Illumina human 610 quad beadchip platform (GPL8887) at the Northwestern University Center for Genetic Medicine Genomics Core Facility according to the manufacturer's instructions. One sample was removed because it had a no call rate >10%. The initial marker set comprised 620,901 markers. 8,300 markers were removed because they showed significant deviation from Hardy-Weinberg equilibrium (HWE, Fischer's exact test, $p < 0.001$). 29,705 SNPs were removed from the analysis because they had a no call rate in more than >10% of the samples. Hence, our final marker set is comprised of 583,073 SNPs. Imputation was performed using impute2 and the full 1KGP reference panel, as per impute2 recommendations, in 5 Mb segments using the commands:

```
impute_v2.2.2_x86_64_dynamic/impute2 -prephase_g -Ne 20000
```

```
impute_v2.2.2_x86_64_dynamic/impute2 -Ne 20000
```

The cis-eQTL scan was conducted with Bayesian regression using all SNPs within 1 Mb of the gene using the command:

```
snptest_v2.5-beta4 -use_raw_phenotypes -bayesian 1 -method expected -pheno  
pheno1 -prior_qt_mean_b 0 -prior_qt_V_b 0.02 -prior_qt_a 3 -prior_qt_b 2
```

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was funded by the following grants: NIH DK097534 (C.G., M.G.H., D.M.S., L.L.A., W.L.L., T.E.R.); NIH T32 DK007169 (A.E.L.); NIH DK42583 (M.E.A., C.B.N.); NIH R01 MH101822 (C.D.B.); Department of Veterans Affairs Career Development Grant 11K2BX001587 (B.T.L.) We thank Greg Crawford for kindly providing access to DNase-seq data produced in his lab.

References

1. Lain KY, Catalano PM. Metabolic changes in pregnancy. *Clinical obstetrics and gynecology*. 2007; 50:938–948. doi:10.1097/GRF.0b013e31815a5494. [PubMed: 17982337]
2. Metzger BE. Long-term outcomes in mothers diagnosed with gestational diabetes mellitus and their offspring. *Clinical obstetrics and gynecology*. 2007; 50:972–979. doi:10.1097/GRF.0b013e31815a61d6. [PubMed: 17982340]
3. Weight in infancy and death from ischaemic heart disease. *Lancet*. 1989; 2:1335.
4. Curhan GC, et al. Birth weight and adult hypertension, diabetes mellitus, and obesity in US men. *Circulation*. 1996; 94:3246–3250. [PubMed: 8989136]
5. Rich-Edwards JW, et al. Birthweight and the risk for type 2 diabetes mellitus in adult women. *Annals of internal medicine*. 1999; 130:278–284. [PubMed: 10068385]
6. Hayes MG, et al. Identification of HKDC1 and BACE2 as genes influencing glycemic traits during pregnancy through genome-wide association studies. *Diabetes*. 2013; 62:3282–3291. doi:10.2337/db12-1692. [PubMed: 23903356]
7. Scott RA, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nature genetics*. 2012; 44:991–1005. doi:10.1038/ng.2385. [PubMed: 22885924]

8. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS one*. 2012; 7:e46688. doi:10.1371/journal.pone.0046688. [PubMed: 23056405]
9. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. *Genome research*. 2001; 11:863–874. doi:10.1101/gr.176601. [PubMed: 11337480]
10. Uhlen M, et al. Towards a knowledge-based Human Protein Atlas. *Nature biotechnology*. 2010; 28:1248–1250. doi:10.1038/nbt1210-1248.
11. Irwin DM, Tan H. Molecular evolution of the vertebrate hexokinase gene family: Identification of a conserved fifth vertebrate hexokinase gene. *Comparative biochemistry and physiology. Part D, Genomics & proteomics*. 2008; 3:96–107. doi:10.1016/j.cbd.2007.11.002.
12. Consortium EP, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489:57–74. doi:10.1038/nature11247. [PubMed: 22955616]
13. Bernstein BE, et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nature biotechnology*. 2010; 28:1045–1048. doi:10.1038/nbt1010-1045.
14. Genomes Project C, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; 491:56–65. doi:10.1038/nature11632. [PubMed: 23128226]
15. Innocenti F, et al. Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. *PLoS genetics*. 2011; 7:e1002078. doi:10.1371/journal.pgen.1002078. [PubMed: 21637794]
16. Montgomery SB, et al. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature*. 2010; 464:773–777. doi:10.1038/nature08903. [PubMed: 20220756]
17. Patwardhan RP, et al. Massively parallel functional dissection of mammalian enhancers in vivo. *Nature biotechnology*. 2012; 30:265–270. doi:10.1038/nbt.2136.
18. van den Boogaard M, et al. A common genetic variant within SCN10A modulates cardiac SCN5A expression. *The Journal of clinical investigation*. 2014; 124:1844–1852. doi:10.1172/JCI73140. [PubMed: 24642470]
19. Xu Z, et al. A functional variant in IL12B promoter modulates its expression and associates with increased risk of allergic asthma. *Genes and immunity*. 2013; 14:238–243. doi:10.1038/gene.2013.13. [PubMed: 23575353]
20. Gonzalez C, Ureta T, Sanchez R, Niemeyer H. Multiple molecular forms of ATP:hexose 6-phosphotransferase from rat liver. *Biochemical and biophysical research communications*. 1964; 16:347–352. [PubMed: 5871820]
21. Katzen HM, Schimke RT. Multiple forms of hexokinase in the rat: tissue distribution, age dependency, and properties. *Proceedings of the National Academy of Sciences of the United States of America*. 1965; 54:1218–1225. [PubMed: 5219826]
22. Heredia VV, Thomson J, Nettleton D, Sun S. Glucose-induced conformational changes in glucokinase mediate allosteric regulation: transient kinetic analysis. *Biochemistry*. 2006; 45:7553–7562. doi:10.1021/bi060253q. [PubMed: 16768451]
23. Fendler W, et al. Population-based estimates for double diabetes amongst people with glucokinase monogenic diabetes, GCK-MODY. *Diabetic medicine : a journal of the British Diabetic Association*. 2014 doi:10.1111/dme.12449.
24. Faletra F, et al. Congenital hyperinsulinism: clinical and molecular analysis of a large Italian cohort. *Gene*. 2013; 521:160–165. doi:10.1016/j.gene.2013.03.021. [PubMed: 23506826]
25. Becker TC, BeltrandRio H, Noel RJ, Johnson JH, Newgard CB. Overexpression of hexokinase I in isolated islets of Langerhans via recombinant adenovirus. Enhancement of glucose metabolism and insulin secretion at basal but not stimulatory glucose levels. *The Journal of biological chemistry*. 1994; 269:21234–21238. [PubMed: 8063745]
26. Corradin O, et al. Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome research*. 2014; 24:1–13. doi:10.1101/gr.164079.113. [PubMed: 24196873]
27. Musunuru K, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature*. 2010; 466:714–719. doi:10.1038/nature09266. [PubMed: 20686566]

28. Huang Q, et al. A prostate cancer susceptibility allele at 6q22 increases RFX6 expression by modulating HOXB13 chromatin binding. *Nature genetics*. 2014; 46:126–135. doi:10.1038/ng.2862. [PubMed: 24390282]
29. Kwasniewski JC, Mogno I, Myers CA, Corbo JC, Cohen BA. Complex effects of nucleotide variants in a mammalian cis-regulatory element. *Proceedings of the National Academy of Sciences of the United States of America*. 2012; 109:19498–19503. doi:10.1073/pnas.1210678109. [PubMed: 23129659]
30. Cong L, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 2013; 339:819–823. doi:10.1126/science.1231143. [PubMed: 23287718]
31. Mali P, et al. RNA-guided human genome engineering via Cas9. *Science*. 2013; 339:823–826. doi:10.1126/science.1232033. [PubMed: 23287722]
32. Cardenas ML, Cornish-Bowden A, Ureta T. Evolution and regulatory role of the hexokinases. *Biochimica et biophysica acta*. 1998; 1401:242–264. [PubMed: 9540816]
33. Wilson JE. Hexokinases. *Reviews of physiology, biochemistry and pharmacology*. 1995; 126:65–198.
34. Smemo S, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*. 2014; 507:371–375. doi:10.1038/nature13138. [PubMed: 24646999]
35. Sherry ST, et al. dbSNP: the NCBI database of genetic variation. *Nucleic acids research*. 2001; 29:308–311. [PubMed: 11125122]
36. He TC, et al. A simplified system for generating recombinant adenoviruses. *Proceedings of the National Academy of Sciences of the United States of America*. 1998; 95:2509–2514. [PubMed: 9482916]
37. Hohmeier HE, et al. Isolation of INS-1-derived cell lines with robust ATP-sensitive K⁺ channel-dependent and -independent glucose-stimulated insulin secretion. *Diabetes*. 2000; 49:424–430. [PubMed: 10868964]
38. Sievers F, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*. 2011; 7:539. doi:10.1038/msb.2011.75. [PubMed: 21988835]
39. Kent WJ, et al. The human genome browser at UCSC. *Genome research*. 2002; 12:996–1006. doi:10.1101/gr.229102. Article published online before print in May 2002. [PubMed: 12045153]

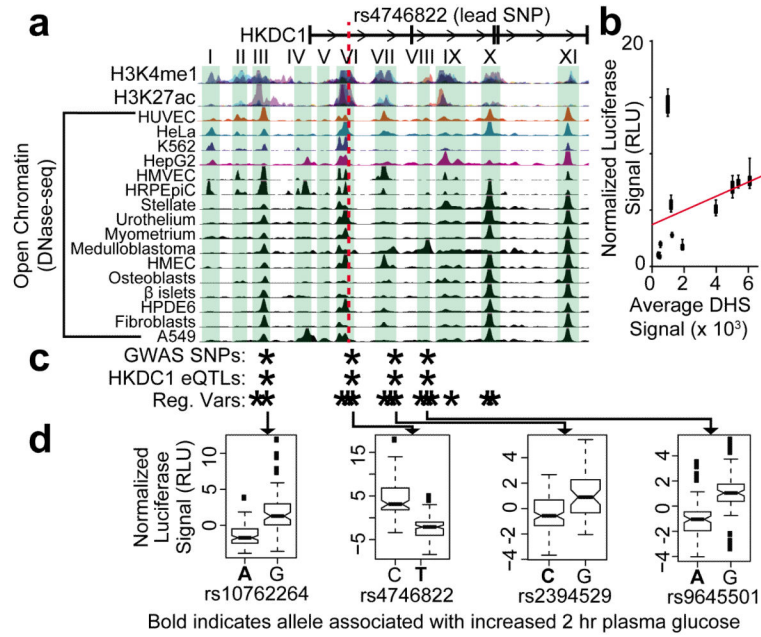


Figure 1. Coordinated allelic regulation of *HKDC1*

(a) Map of chromatin landscape and the *HKDC1* genome wide association (GWA) locus target regions. Evidence of active regulatory elements – genomic regions with the covalent histone modifications H3K4me1 and H3K27ac as measured by ChIP-seq and open chromatin measured by DNase-seq – is shown across the genomic locus associated with gestational hyperglycemia. Green boxes indicate candidate the regulatory elements whose activity was measured with luciferase reporter assays. Histone modification and open chromatin data were obtained from the ENCODE project. **(b)** For each regulatory element in **a**, the enhancer activity (y-axis) is plotted against DNase-seq signal averaged across the element (x-axis) in HepG2 cells (n= 8 to 19). Enhancer activity was determined by dividing the relative luciferase signal from the most active haplotype by that of a control vector with the same promoter but no enhancer. The red line indicates the Pearson correlation between DNase-seq signal and enhancer activity. Error bars show standard deviation (s.d.). **(c)** Coordinated regulatory variation in the *HKDC1* locus. SNPs that are significantly associated with gestational hyperglycemia (“GWA SNPs”), *HKDC1* mRNA expression (“*HKDC1* eQTLs”), or regulatory activity in allele-specific luciferase reporter assays (“Reg. Vars”) are marked with an asterisk. **(d)** Example box plots showing allele specific regulatory activity for the four SNPs that were significantly associated with gestational hyperglycemia, *HKDC1* expression, and luciferase reporter gene expression. In each example, they associated with increased 2 hr plasma glucose are shown in bold face. The bottom and top boxes are the first and third quartiles, and the band inside the box is the median. The ends of the whiskers represent the lowest and highest data points within 1.5 interquartile range of the lower and upper quartiles. Black squares represent outliers defined as 1.5 times the interquartile range above the upper quartile or below the lower quartile. The number of replicate measurements followed by each allele are as follows: 103 of rs10762264A, 79 of rs10762264G, 80 of rs4746822C, 115 of rs4746822T, 129 of rs2394529C, 47 of rs2394529G, 80 of rs9645501A, and 80 of rs9645501G.

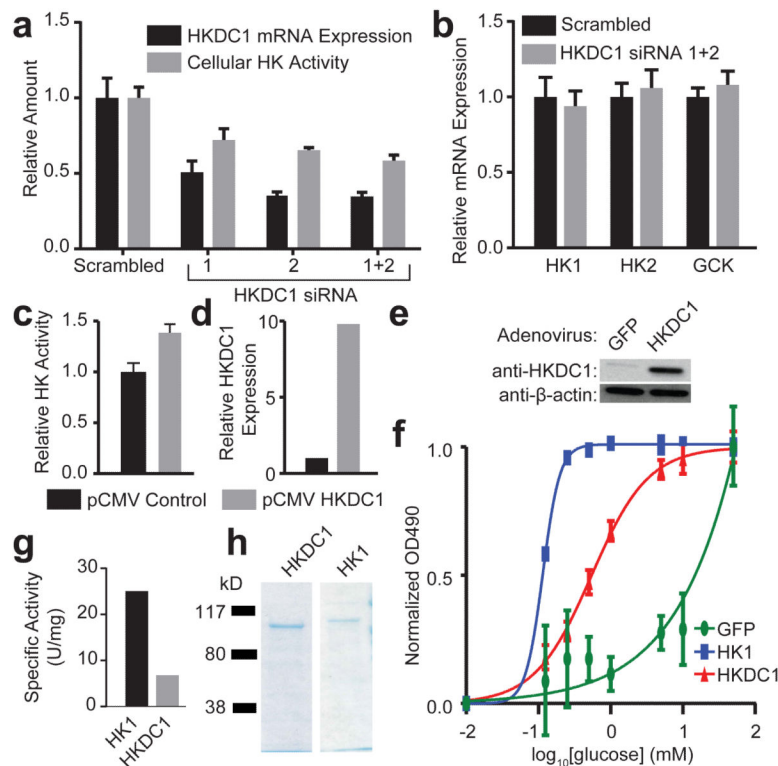


Figure 2. HKDC1 is a hexokinase

(a) Two different siRNAs were used to knockdown *HKDC1* in HepG2 cells. Quantification of hexokinase (HK) activity from whole cell lysates shows a dose-dependent decrease in HK activity with reduced *HKDC1* expression (n = 4). (b) siRNAs are specific to *HKDC1* and do not impact the expression of the other human hexokinase genes that are expression in HepG2 cells (n = 4). (c,d) Transient overexpression of *HKDC1* increases total hexokinase activity in HepG2 cells compared to controls which were transfected with a plasmid expression a truncated *HKDC1* (n = 2). (e,f) Adenoviral-mediated overexpression of *HKDC1* overexpression in INS-1 cells increased the amount of *HKDC1* protein and cellular HK activity across a range of 0-50 mM glucose shown in f. The level of hexokinase activity was determined by dividing the optical density at 490 nm (OD490) at each glucose concentration by the OD490 at 50 mM glucose in all 3 transduction conditions, respectively (n = 3). (g) *HKDC1* and *HK1* protein was expressed in bacterial cells and isolated. Purity of the isolated protein was demonstrated with Coomassie-blue staining. (h) Hexokinase assays performed on the purified protein demonstrate that both *HKDC1* and *HK1* have hexokinase activity. Specific activity was defined as micromoles of NADPH generated per hour per microgram of protein (n = 2). All error bars in figure 2 show s.d.

Association of functional genetic variants with regulatory activity (β Luciferase), HKDC1 expression (β eQTL), and gestational hyperglycemia (β GWA)

Table 1

Region	SNPId	Non Risk	Risk	MAF	β Luc	95% CI Luc	β eQTL	β GWA	$\log_{10}(\text{bf})$ eQTL	p GWA	R ²
III	rs10762264	G	A	0.48	-3.13	-4.64 to -1.62	-0.327	6.946	3.4654	3.75E-12	0.906
III	rs12241136	T	A	0.10	-2.77	-3.72 to -1.81	0.057	3.639	-0.0459	2.739E-4	0.001
VI	rs78983061	C	A	0.06	2.21	0.72 to 3.70	0.061	0.496	0.01597	0.6198	0.064
VI	rs7089277	G	T	0.07	8.59	6.07 to 11.10	0.040	3.942	-0.0701	8.07E-05	0.150
VI	rs4746822	C	T	0.49	-6.7	-9.10 to -4.29	-0.322	7.242	3.3899	4.41E-13	1.000
VII	rs4746824	C	A	0.22	1.18	0.70 to 1.66	-0.195	2.792	1.0366	0.005238	0.180
VII	rs75405157	T	C	0.06	1.55	1.02 to 2.09	0.061	0.481	0.01399	0.6306	0.004
VII	rs2394529	G	C	0.5	-1.41	-2.11 to -0.69	-0.308	7.043	3.0703	1.88E-12	0.954
VIII	rs9645501	G	A	0.34	-2.08	-2.81 to -1.33	-0.287	4.449	2.6947	8.61E-06	0.559
VIII	rs147449838	G	A	<.01	-1.23	-1.91 to -0.54	*	*	*	*	<0.001
VIII	rs200216341	G	A	<.01	-1.31	-2.00 to -0.63	*	*	*	*	<0.001
IX	rs1983128	G	A	<.01	-4.45	-5.94 to -2.96	-0.168	1.180	0.55676	0.2378	0.036
X	rs5030945	T	C	0.45	-28.14	-37.05 to -20.14	-0.011	2.359	-0.22229	0.01831	0.077
X	rs874557	A	G	0.45	-31.58	-44.26 to -18.83	-0.001	2.357	-0.22318	0.01843	0.077

Notes:

All β -s are relative to the risk allele. Bolded values indicate β estimates that are significantly different from 0 for Luciferase reporter assays (t-test, Bonferroni adjusted $p < 0.05$, $n > 12$), eQTL ($|\log_{10}(\text{Bayes Factor})| > 2.5$), $n = 532$) and GWA (F-test, $p < 1 \times 10^{-8}$, $n = 4,437$), R² values are relative to rs4746822.

* indicates parameters and p-values that were not estimated.