Coordinated regulatory variation associated with gestational hyperglycaemia regulates expression of the novel hexokinase HKDC1

Cong Guo1,2, Anton E. Ludvik3, Michelle E. Arlotto4, M. Geoffrey Hayes3, Loren L. Armstrong3, Denise M. Scholtens5, Christopher D. Brown6, Christopher B. Newgard4,7,8, Thomas C. Becker4,8, Brian T. Layden3,9, William L. Lowe3 & Timothy E. Reddy2,10

Maternal glucose levels during pregnancy impact the developing fetus, affecting metabolic health both early and later on in life. Both genetic and environmental factors influence maternal metabolism, but little is known about the genetic mechanisms that alter glucose metabolism during pregnancy. Here, we report that haplotypes previously associated with gestational hyperglycaemia 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; in addition, 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.
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 trimester. In 5–10% of pregnancies, however, glucose homeostasis is not maintained, resulting in hyperglycaemia and gestational diabetes, that is, diabetes that is first diagnosed during pregnancy. Gestational hyperglycaemia is a major health risk to the mother, and is associated with adverse fetal outcomes including increased risk of type 2 diabetes and obesity in the offspring. Despite the prevalence of health risks, the mechanisms leading to gestational hyperglycaemia remain largely unknown.

We recently identified a strong association between genetic variants in the first intron of HKDC1 and 2-h plasma glucose levels during an oral glucose tolerance test performed in a multiethnic cohort of 4,437 mothers at ~28 weeks gestation (F-test, $P = 8.26 \times 10^{-13}$, $\beta = 0.167 - 0.229 \mu$mol l$^{-1}$). 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-h plasma glucose, suggesting that the association with HKDC1 is largely pregnancy specific. In the 100 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 (refs 8,9; Supplementary Data 1). However, none of the coding single-nucleotide polymorphisms (SNPs) were associated with 2-h plasma glucose at genome-wide significance ($P < 1 \times 10^{-8}$).

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 elements. 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% haplotypes for the regulatory elements sequenced by the 1KGP (Supplementary Table 2).

To determine whether genetic variation in the identified regulatory elements influences endogenous HKDC1 expression, we compared the variants associated with gestational hyperglycaemia to those associated with gene expression changes in an expression quantitative trait locus (eQTL) study performed in primary human livers. To do so, we imputed the tag variants from both the GWA and eQTL studies to the variants identified in the 1KGP. That analysis revealed four significant eQTLs for HKDC1 ($\log_{10}$Bayes Factor > 2.5) within DNase hypersensitive sites (DHS) that were also associated with maternal 2-h 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 levels. 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-h maternal glucose levels were also eQTLs for those genes (Supplementary Fig. 3). Together, these results suggest that genetic variants associated with gestational hyperglycaemia 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. 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 the 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 that do not segregate in the population on expression 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. None 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 $r = 0.64$, $P < 0.05$, Fig. 1c), 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. 1b, 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-fold$^{17–19}$. 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 on the basis of each variant within the haplotype and plotted those against the observed luciferase values. In all the six elements, predicted luciferase
values were positively and significantly correlated with observed luciferase values (Pearson $r = 0.22$, $P < 0.001$, 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 there are strongly coordinated regulatory effects across the genomic region associated with 2-h plasma glucose levels, and that exceptions to that coordination likely do not contribute to the expression of HKDC1.

**Table 1 | Association of functional genetic variants with regulatory activity ($\beta$-luciferase ($\beta$-luc), HKDC1 expression ($\beta$-eQTL) and gestational hyperglycaemia ($\beta$-GWA).**

| Region | SNPId | Non-Risk | Risk | MAF   | $\beta$-Luc | 95% CI Luc | $\beta$-eQTL | $\beta$-GWA | Log$_{10}$(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  | 0.21       | 0.72 to 3.70   | 0.061  | 0.496 | 0.01957 | 0.619E 0.064 |
| VI     | rs7089277   | G        | T    | 0.07  | 0.85       | 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 | 4.1E – 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    | rs7540517   | T        | C    | 0.06  | 1.55       | 1.02 to 2.09   | 0.061  | 0.481 | 0.01399 | 0.630E 0.004 |
| VII    | rs2394529   | G        | C    | 0.5   | 1.61       | -2.11 to -0.69 | -0.308 | 7.043 | 3.0703 | 1.88E – 12  | 0.954 |
| VII    | rs9645501   | G        | A    | 0.34  | -2.08      | -2.81 to -1.33 | -0.287 | 4.449 | 2.6947 | 8.61E 0.559 |
| VIII   | rs147449838 | A        | G    | 0.01  | 1.23       | -1.91 to 0.54  | 0.001  | 0.001 | 0.001  | 0.001  |
| VIII   | rs200216341 | G        | A    | 0.01  | -1.31      | -2.00 to -0.63 | 0.007  | 0.007 | 0.007  | 0.007  |
| IX     | rs1983128   | G        | A    | 0.01  | -4.45      | -5.94 to -2.96 | -0.168 | 1.180 | 0.55676 | 0.237E 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 | -0.23218 | 0.0843 | 0.077 |

* Indicates parameters and F values that were not estimated. All $\beta$s are relative to the risk allele. Bold values indicate $\hat{\beta}$ estimates that are significantly different from 0 for luciferase reporter assays ($t$-test, Bonferroni-adjusted $P < 0.05$, $n = 12$), eQTL ($\log_{10}$(Bayes Factor) > 2.5), $n = 532$ and GWA ($F$-test, $P < 10^{-8}$, $n = 4.437$). $R^2$ values are relative to rs4746822.

**HKDC1 is a novel human hexokinase gene.** Having demonstrated that alleles associated with higher maternal 2-h 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 diabetes23,24.

To determine whether 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-h 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 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. That result agrees with RNA-seq data showing that half of the hexokinase mRNA expression in HepG2 cells is not significantly associated with 2-h 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-h glucose ($P = 8.02 \times 10^{-5}$). The two SNPs in region X have large luciferase $\beta$s, but were not significantly associated with 2-h 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-h plasma glucose levels, and that exceptions to that coordination likely do not contribute to the expression of HKDC1.

The identified regulatory variants have coordinated effects. For the regulatory variants identified to collectively contribute to maternal glycaemia, we hypothesized that there would be coordination between allele-specific regulatory element activity and maternal 2-h glucose levels. Supporting our hypothesis, all four variants that were significant eQTLs for HKDC1 in liver were associated with maternal 2-h glucose levels with three reaching genome-wide significance in the imputed GWA data (Table 1). All four alleles that were associated with increased 2-h 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-h 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
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,d).

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 to 50 mM (ref. 25). The predominant hexokinase in INS-1 cells is GCK, and the cells have little or no detectable low Km 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.

On the basis of the extent of nucleotide and amino-acid similarity between HKDC1 and members of the hexokinase family, (Supplementary Data 8 and 9, Supplementary Tables 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 phenotype.26–28 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 expression.26 Such coordinated disruption of regulatory haplotypes may also explain how modest effects of individual regulatory variants 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 noncoding 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 a strong association between variants and heterogeneous effect sizes may be overcome with high-throughput empirical investigation of allele-specific regulatory activity in large populations.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 glycaemia 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 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 with HK1 in our in vitro assays. One possibility is that additional cofactors 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 the role of HKDC1 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 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 disease.34

Methods

Luciferase reporter assays. The selected regions were amplified from genomic DNA from 1KGP subjects and In-Phusion cloned at the Nhel cut site into pGL4.13 luciferase expression vector (Promega). The construct was then transformed into TOP-10 competent cells and plated onto Luria-Bertani 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 Luri-Berti media overnight. Plasmids were extracted using the PureYield Plasmid Miniprep System (Promega). Constructs were sequenced using Sanger sequencing and variants were confirmed in dbSNP.35 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 per well. After 48 h, 100 ng of each construct per well was transfected with Fugene HD (Promega) at a 5:1 Fugene:DNA ratio. At least 16 biological replicates for each construct were transfected. After 24 h, 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 programming environment (http://cran.r-project.org) was used to determine the regulatory effect of individual SNPs where (normalized luciferase intensity) = \( \beta_0 + \beta_1 \, SNP_1 + \beta_2 \, SNP_2 + \ldots + \epsilon \). Raw data for luciferase reporter assays are provided as Supplementary Data Sets.

Quantitative PCR with reverse transcription. RNA was isolated using the RNeasy kit (Qiagen). Two micrograms of RNA per reaction was reverse transcribed using the SuperScript Vilo cDNA Kit (Life Technologies). All quantitative PCR (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 for 10 s and the primer annealing temperature (30 s) for 40 cycles. Calculations were performed using the \( \Delta\Delta C_t \) method using \( P_\alpha \)-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 bichromonic acid assay (Thermo Scientific). The lysate was then mixed 1:1 with loading buffer supplemented with 5% 2-mercaptoethanol and incubated at 76 °C for 15 min. Then, 10 μg of protein along with the Precision Plus Protein Dual Color Standard (Bio-Rad) was loaded into each lane of a Mini-Protean II (Bio-Rad) and run at 100 V for 1 h. The membrane was blocked with 5% skimmed milk for 30 min and washed three times in 0.1% Tween-20 in phosphate-buffered saline. The membrane was incubated overnight in anti-HKDC1 antibody produced in rabbit (Sigma) diluted 1:500 in skimmed milk for 75 min at room temperature (Santa Cruz Biotechnology). The membrane was washed three times and incubated at room temperature in 10 μl SuperSignal West

© 2015 Macmillan Publishers Limited. All rights reserved.
Pico Chemiluminescent Substrate (Thermo Scientific) for 15 min. After exposure and imaging, the membrane was stripped using Restore Western Blot Stripping Buffer (Thermo Scientific) and blotted using 1:250 anti-rabbit antibody conjugated to horseradish peroxidase (Santa Cruz Biotechnology) and diluted in skimmed milk.

**Adenovirus construction and transduction.** A 2.789 bp BamHI/NotI fragment from plasmid RC221178 (Origene, containing 2,750 bp HKDC1 open reading frame, was cloned into the BamHI/NotI sites of the adenoviral shuttle vector pShuttle (Agilent). HKDC1 adenovirus (AdHKDC1) was linearized with NotI and transformed into BS5183-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 pReceiver-B01 plasmid containing either p150 tissue culture plates. Before complete lysis, media was removed infected cell line in 2 ml freeze/thaw buffer (10 mM Tris/HCl, pH 8.0, 1 mM MgCl2), lysed by two freeze/thaw cycles and the virus purified by ultracentrifugation on a CsCl gradient. Purified virus was de-salted with a 7 kMWCO column (Thermo Scientific) equilibrated with freeze/thaw buffer, and glycerol was added to a final concentration of 10%. Viruses were titred by measuring OD260 (1 OD260 equivalent of 107 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 × 105 per well into 12-well plates. INS-1 832/13 cells were kindly provided by Christopher Newgard (DMP1, Duke University)37. After 24 h, the cells were transfected with HKDC1 or GFP expressing adenovirus in triplicate. Cells were grown for another 24 h and pelleted with centrifugation.

**Liver eQTL analyses.** Liver gene expression and genotype data were generated as follows38. 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-transformed. Missing data were imputed by k nearest neighbours (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 quantities of the standard normal distribution (R function qnorm). This matrix of processed gene expression values was then subjected to PCA (R package pcaMethods, function pca). We controlled for the effect of the 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 ancestry, the PCs were used to weight the expression quantities 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 multiclust, function multiclust,G = 1:min(4,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 follows39. 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. A total of 29,705 SNPs were removed from the analysis because they had a no-call rate in >10% of the samples. Hence, our final marker set comprised 583,073 SNPs. Imputation was performed using MACH (http://www.broadinstitute.org/mach/) and the 1KGP reference panel, as per impute2 recommendations, in 5 Mb segments using the commands: impute_v2.2.8.x86_64_dynamic/impute2 -p rephase.g -r 20000 impute_v2.2.8.x86_64_dynamic/impute2 -r 20000 -e 20000 -c 20000 The eQTL scan was conducted with Bayesian regression using all SNPs within 1 Mb of the gene using the command: snpsel_v2.5-beta4 -use raw phenotypes -bsim 1 -method expected -pheno1 -prior qt mean_b 0 -prior qt V_b 0.02 -prior qt a 3 -prior qt b 2

**References**

1. Lain, K. Y. & Catalano, P. M. Metabolic changes in pregnancy. *Clin. Obstet. Gynecol.* 50, 938–948 (2007).
2. Metzger, B. E. Long-term outcomes in mothers diagnosed with gestational diabetes mellitus and their offspring. *Clin. Obstet. Gynecol.* 50, 972–979 (2007).
3. Weight in infancy and death from ischaemic heart disease. *Lancet* 2, 1335 (1989).
4. Curhan, G. C. et al. Birth weight and adult hypertension, diabetes mellitus, and obesity in US men. Circulation 94, 3246–3250 (1996).

5. Rich-Edwards, J. W. et al. Birth weight and the risk for type 2 diabetes mellitus in adult women. Ann. Intern. Med. 130, 278–284 (1999).

6. Hayes, M. G. et al. Identification of HKDC1 and BACE2 as genes influencing glycemic traits during pregnancy through genome-wide association studies. Diabetes 62, 3282–3291 (2013).

7. Scott, R. A. et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat. Genet. 44, 991–1005 (2012).

8. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. Predicting the functional effect of amino acid substitutions and indels. PLoS ONE 7, e46688 (2012).

9. Ng, P. C. & Henikoff, S. Predicting deleterious amino acid substitutions. Genome Res. 11, 863–874 (2001).

10. Uhlen, M. et al. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012).

11. Irwin, D. M. & Tan, H. Molecular evolution of the vertebrate hexokinase gene family: identification of a conserved fifth vertebrate hexokinase gene. Comp. Biochem. Physiol. Part D Genomics Proteomics 3, 96–107 (2008).

12. Consortium, E. F. et al. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012).

13. Bernstein, B. E. et al. The NIH Roadmap Epigenomics Mapping Consortium. Nat. Biotechnol. 28, 1045–1048 (2010).

14. Genomes Project, C. et al. Identification of HKDC1 and BACE2 as genes influencing glycemic traits during pregnancy through genome-wide association studies. Proc. Natl Acad. Sci. USA 109, 19498–19503 (2012).

15. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823 (2013).

16. Malis, P. et al. RNA-guided human genome engineering via Cas9. Science 339, 823–826 (2013).

17. Cardenas, M. L., Cornish-Bowden, A. & Ureta, T. Evolution and regulatory role of the hexokinases. Biochim. Biophys. Acta 1401, 242–264 (1998).

18. Wilson, J. E. Hexokinases. Rev. Physiol. Biochem. Pharmacol. 126, 65–198 (1995).

19. Smerno, S. et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature 507, 371–375 (2014).

20. Sherry, S. T. et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 29, 308–311 (2001).

21. He, T. C. et al. A simplified system for generating recombinant adenosviruses. Proc. Natl Acad. Sci. USA 95, 2509–2514 (1998).

22. Hohmeier, H. E. et al. Isolation of INS-1-derived cell lines with robust ATP-sensitive K+ channel-dependent and -independent glucose-stimulated insulin secretion. Diabetes 49, 424–430 (2000).

23. Fendler, W. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539 (2011).

24. Kent, W. J. et al. The human genome browser at UCSC. Genome Res. 12, 996–1006 (2002).

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 1IK2BX001587 (B.T.L.) We thank Greg Crawford for kindly providing access to DNase-seq data produced in his lab.

Author contributions

C.G. performed all the reporter assays, knockdown and overexpression assays, cellular hexokinase assays and related analyses. 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.E.R. designed and implemented adenoviral over-expression 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.E.R. and W.L.L. performed all reporter assays, knockdown and overexpression assays, cellular hexokinase assays and related analyses. 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.E.R. designed and implemented adenoviral over-expression 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.E.R. and W.L.L. supervised aspects of the manuscript. All the authors contributed to the experimental design and writing of the manuscript.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interest.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Guo, C. et al. Coordinated regulatory variation associated with gestational hyperglycaemia regulates expression of the novel hexokinase HKDC1. Nat. Commun. 6:6069 doi: 10.1038/ncomms7069 (2015).