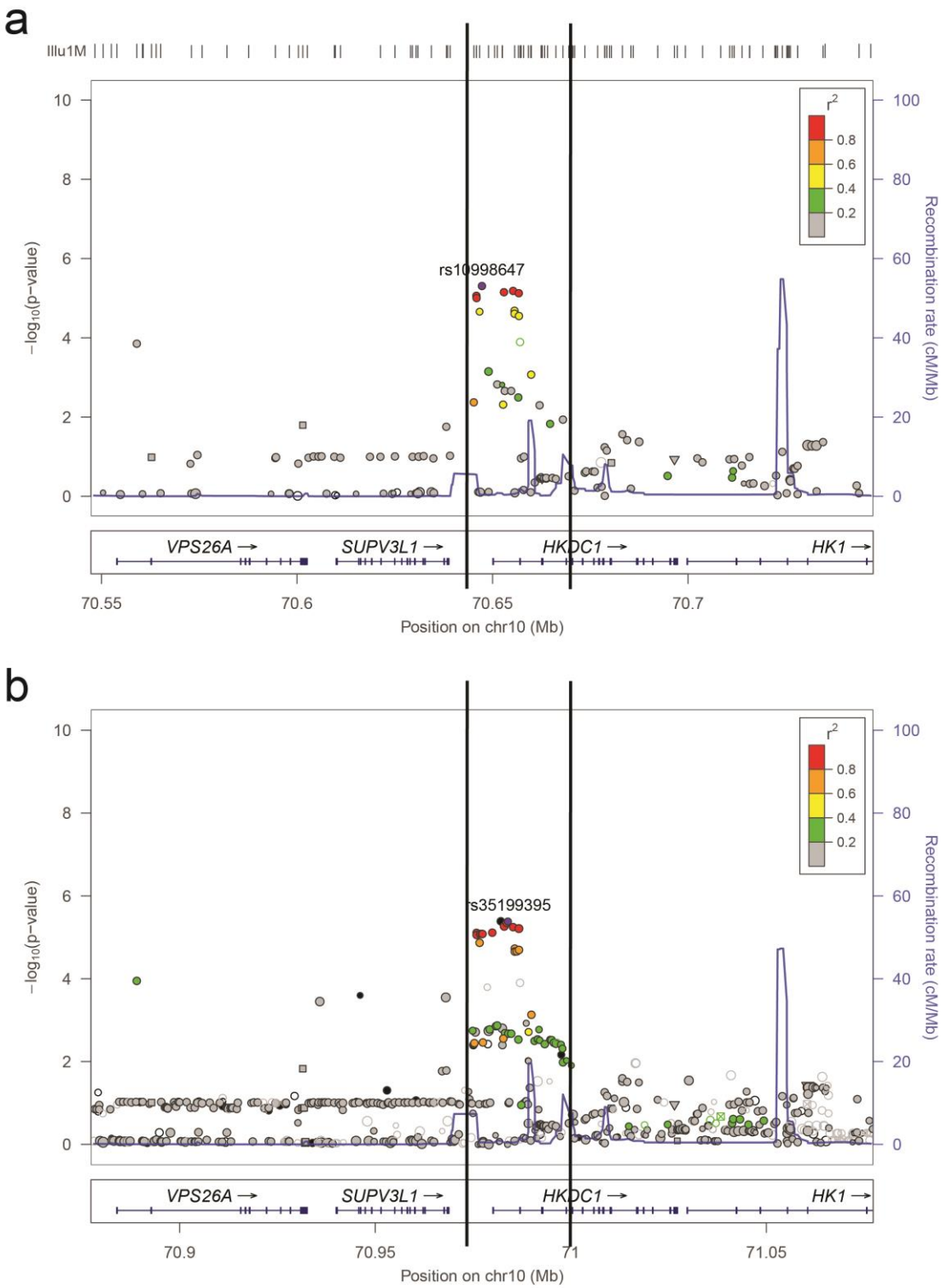
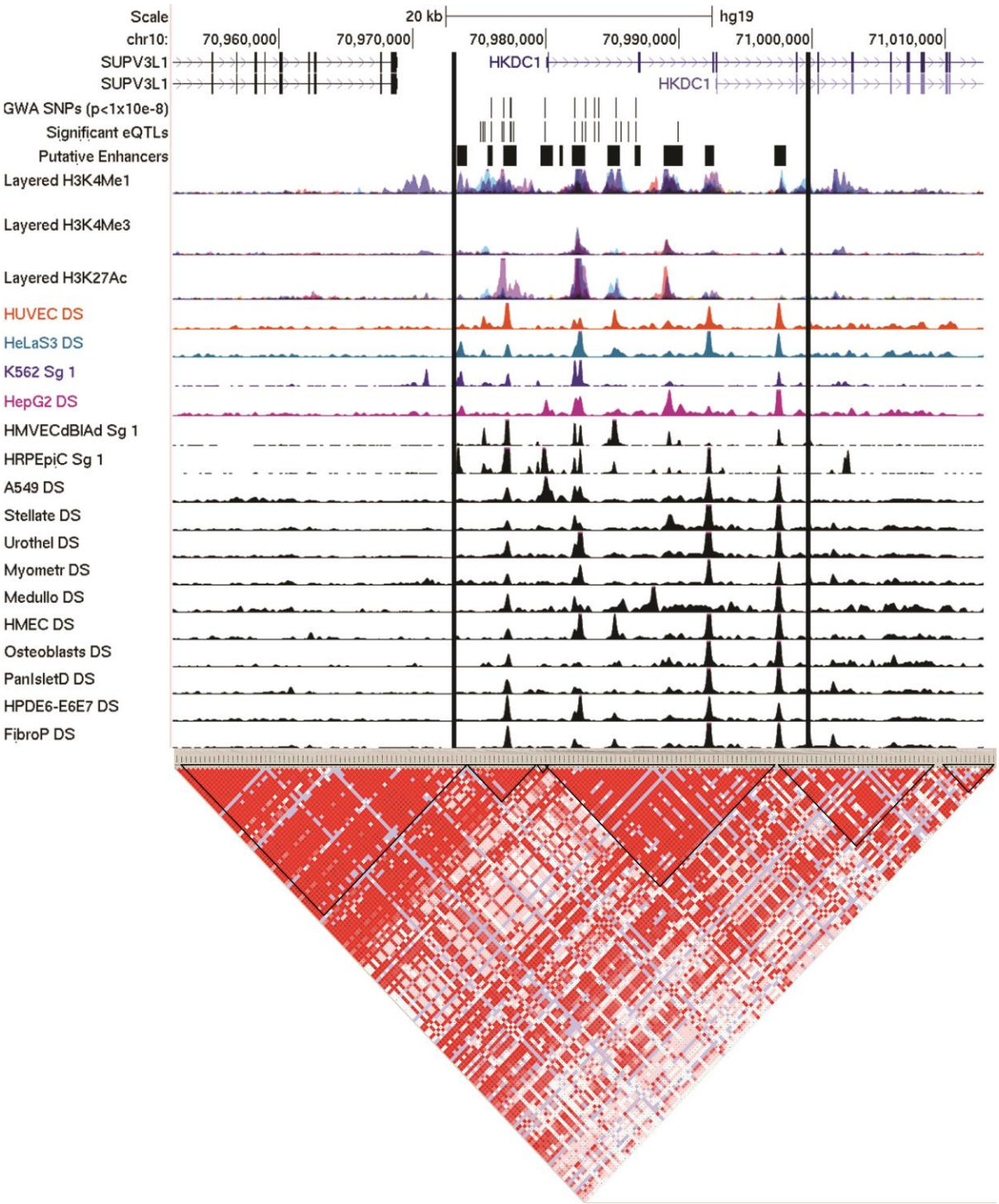
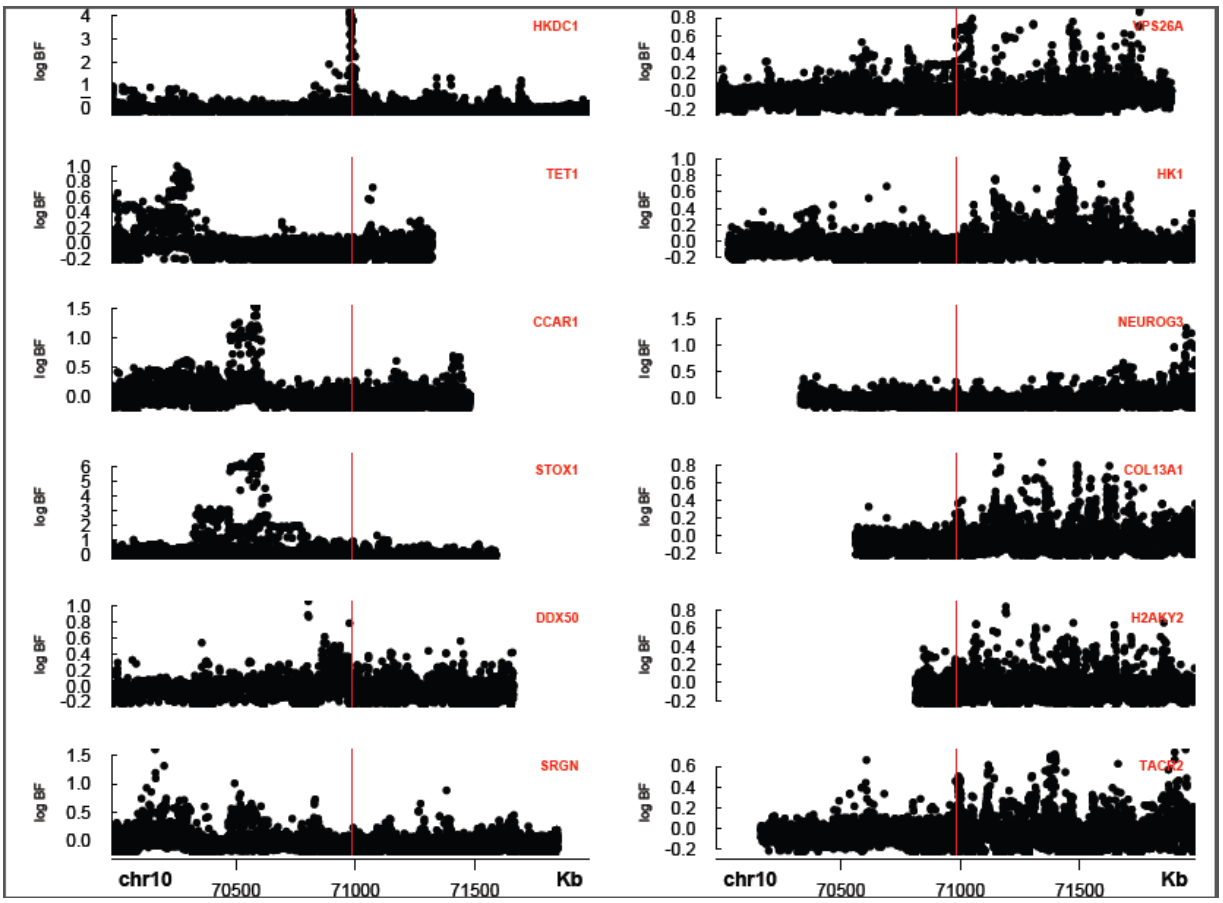
## Supplementary Figures



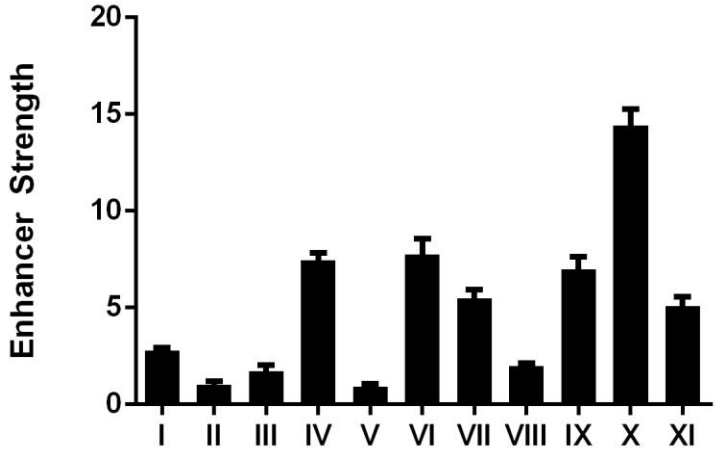

**Supplementary Figure 1:** Loci associated with 2 hr glucose during pregnancy imputed to HAPMAP **(a)** and 1000 Genomes **(b)**. Peak of association is in the first intron of HKDC1. Black bars delineate a 30kb block with the strongest association (F-test, $p < 1 \times 10^{-5}$, n = 1,367) and LD ($r^2 > 0.3$).



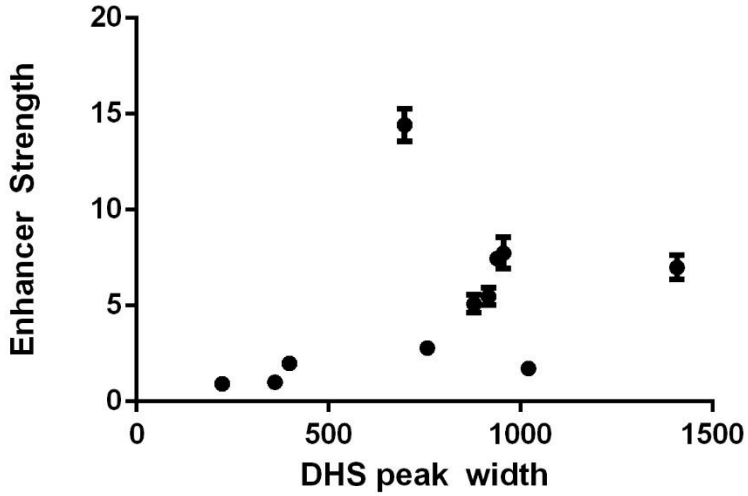

**Supplementary Figure 2:** Significant HKDC1 eQTLs [$\log_{10}$(Bayes Factor) > 2.5] and GWA SNPs (F-test, $p < 1 \times 10^{-8}$) in the maternal 2 hr glucose level associated locus. An enrichment of open chromatin is located in a 30kb block between the 1$^{st}$ and 5$^{th}$ exons of HKDC1, flanked by two regulatory "deserts" on either side. Open chromatin for each cell line is denoted by DNaseI hypersensitivity (ENCODE). Putative enhancer marks are marked by H3KMe1 and H3K27AC. Black bars delineate the same 30 kb block in Supplementary Figure 1.



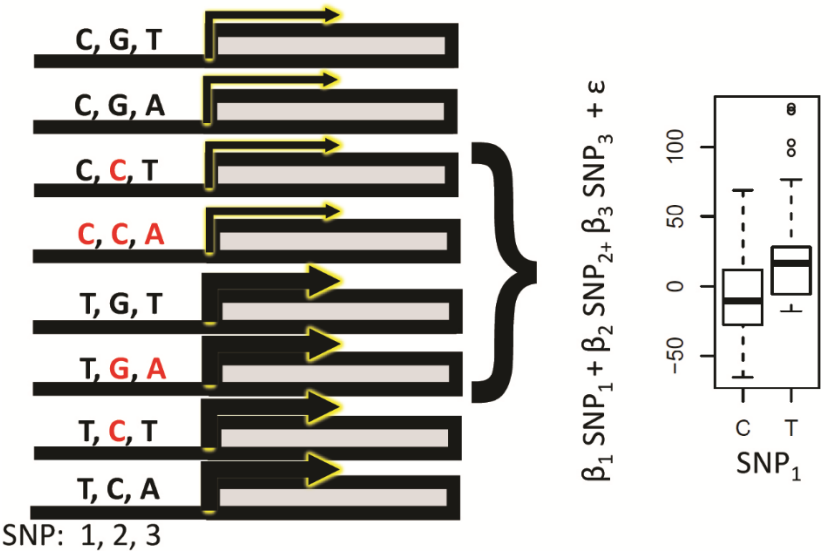

**Supplementary Figure 3:** eQTL analysis in primary liver for all genes within 500 kb of HKDC1. The red line represents the lead GWAS SNP rs4746822.



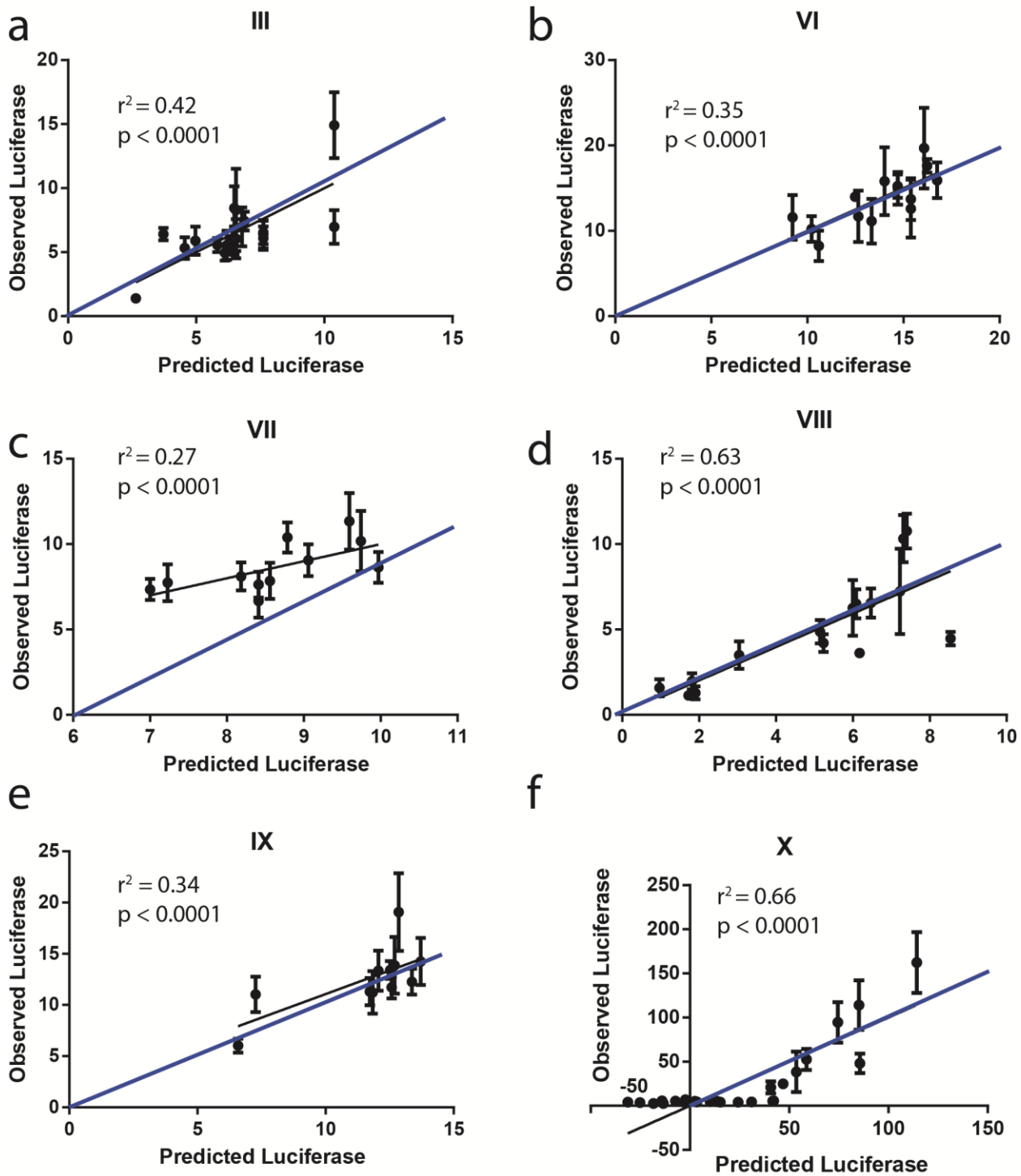

**Supplementary Figure 4:** Enhancer strength of selected regulatory elements. Firefly luciferase intensity was normalized by dividing by the *Renilla* luciferase intensity. Enhancer strength was determined by dividing the normalized luciferase intensity of each construct by the normalized intensity of the empty vector. Error bars show s.d (n = 8 to 19).



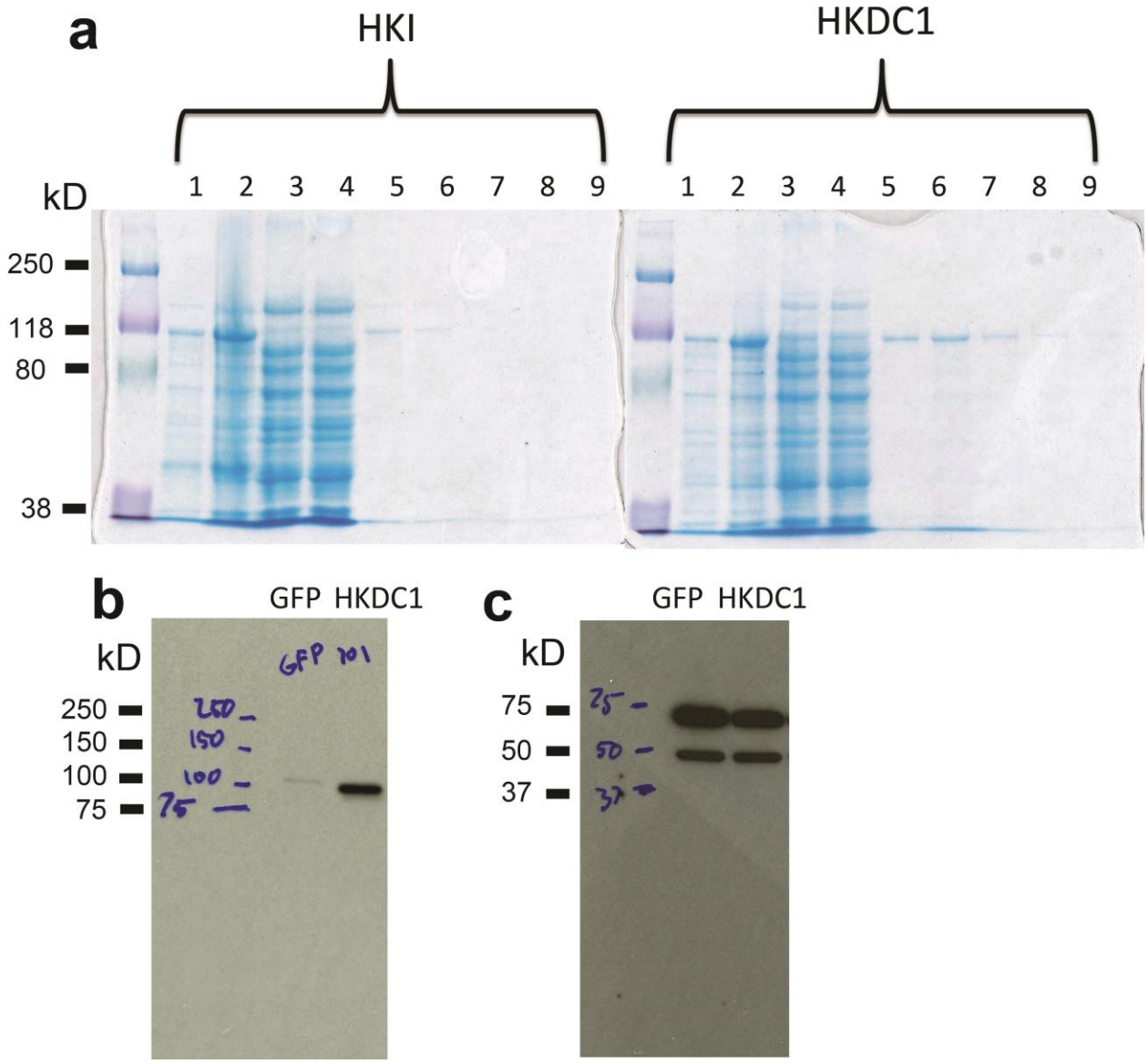

**Supplementary Figure 5:** Enhancer strength vs DHS peak width. Spearman ρ =0.5 ,p = 0.117 (n = 8 to 19) . Error bars show s.d.



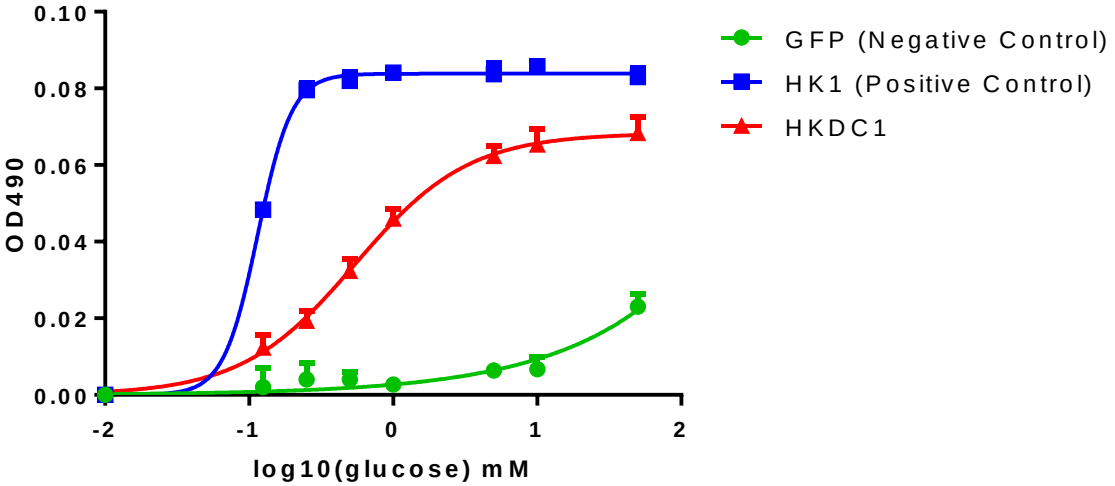

**Supplementary Figure 6:** Experimental design for luciferase reporter assays. Multiple haplotypes of each regulatory element were cloned upstream of the promoter and site directed mutagenesis was used to segregate alleles which do not segregate naturally in the population. A linear regression model was used to determine the effect of each SNP on luciferase expression. 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. Circles represent outliers defined as 1.5 times the interquartile range above the upper quartile or below the lower quartile.



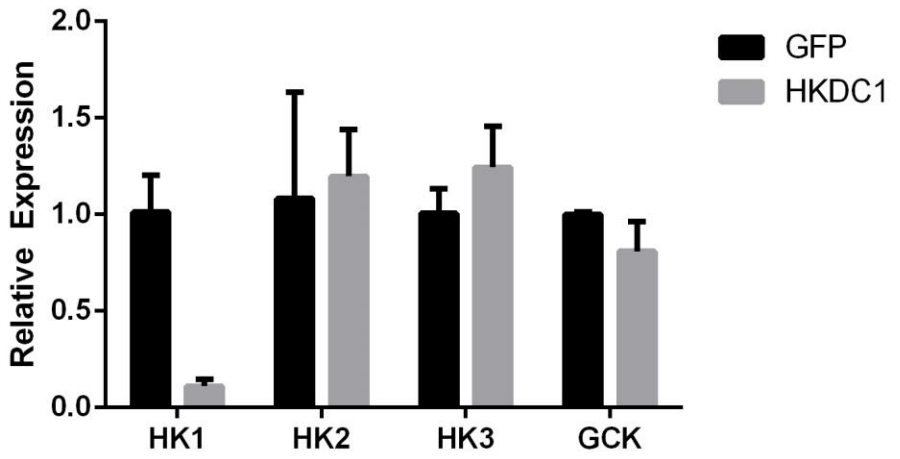

Supplementary Figure 7: Plots of predicted luciferase values versus observed luciferase values. In **a-f**, predicted luciferase values are positively and significantly associated with observed luciferase values, suggesting that the effects of individual variants may combine additively within haplotypes. The blue line is y = x and the black line is the regression line. Error bars show s.d.



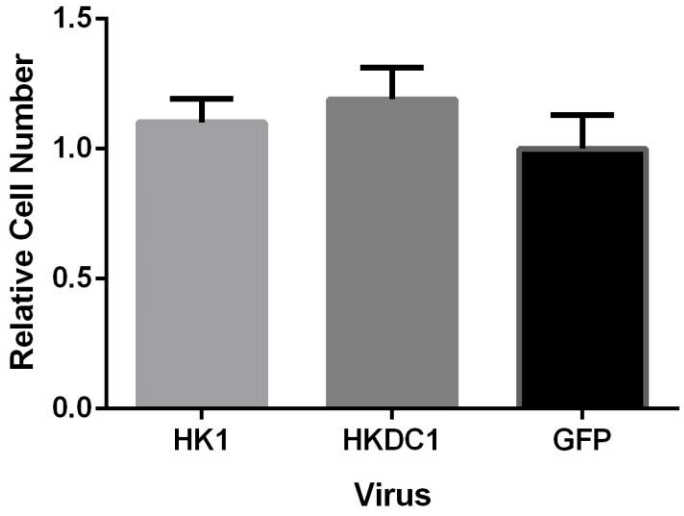

**Supplementary Figure 8:** Full blot and gel images. **(a)** Coomassie stain for HKDC1 and HK1. The numbers representing the following: 1:Preinduction, 2:Induction, 3:Clarified Lysate, 4:Column Flow through, 5:Fraction 1, 6:Fraction 2, 7:Fraction 3, 8:Fraction 4, 9:Fraction 5. Full western blot images are shown for INS-1 cells induced with GFP and HKDC1 adenovirus using an HKDC1 **(b)** and anti-β Actin antibody **(c)**.



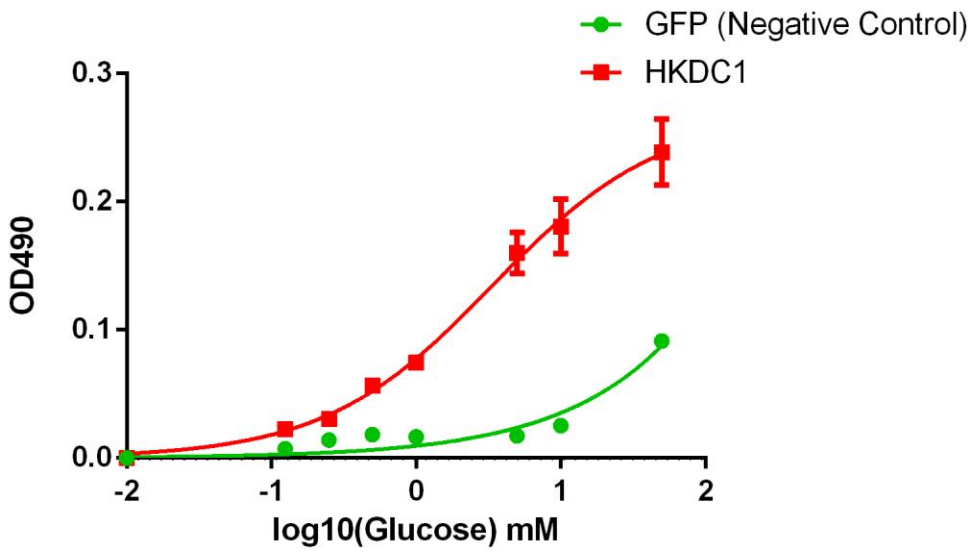

**Supplementary Figure 9:** Un-normalized HKDC1 virus transduction results. HKDC1 overexpression in INS-1 cells via adenovirus shows increased HK activity between a range of 0-50mM glucose. Error bars show s.d.



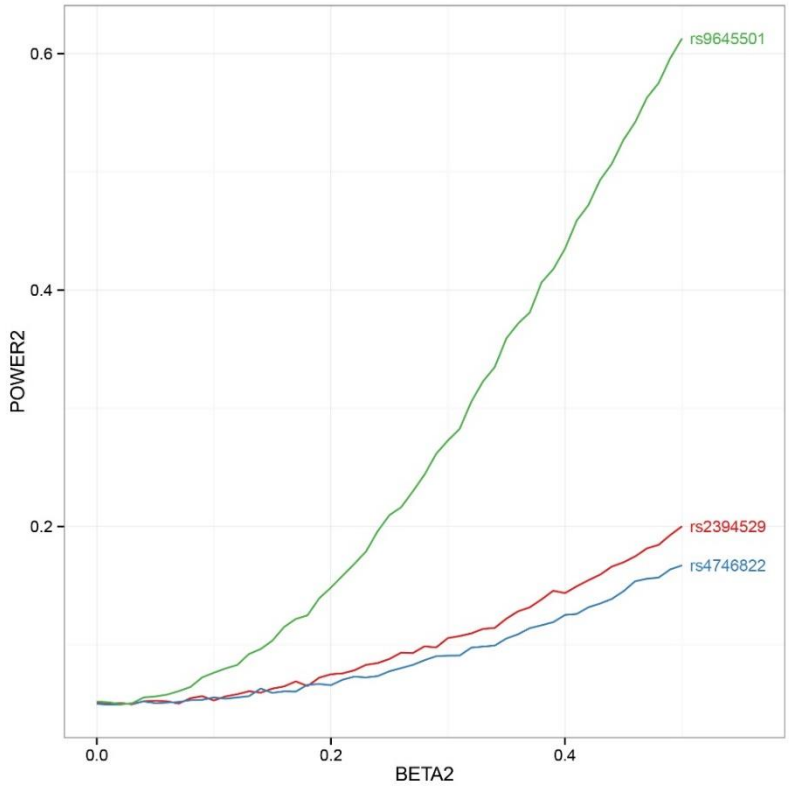

**Supplementary Figure 10:** Expression of other HK mRNAs in INS-1 cells after either GFP or HKDC1 adenovirus transduction. Expression was quantified via qPCR and expression was normalized using the ΔΔ CT method (n = 3). Error bars show s.d.



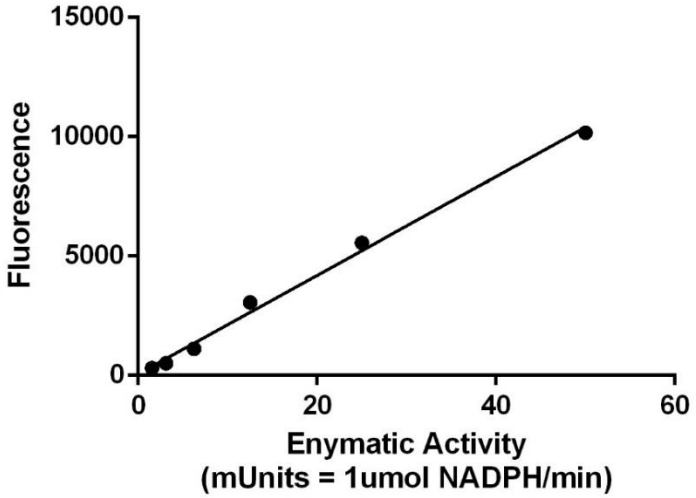

**Supplementary Figure 11:** Cell number is consistent between groups of INS-1 cells treated with adenovirus. Cell number was quantified using the Cell-Titer Glo assay (Promega) (n = 3). Error bars show s.d.




**Supplementary Figure 12:** Replicate of HKDC1 activity dose response curve. This experiment is a duplicate experiment represented in Supplementary Figure 4 (n = 3). Error bars show s.d.




**Supplementary Figure 13:** Power analysis to detect independent additive effects of regulatory variants. The beta of SNP1 (rs10762264) = 0.48 and the MAF = 0.4. $R^2$ values between the 4 SNPs range from 0.8-0.99.




**Supplementary Figure 14:** Standard curve for specific activity versus fluorescence (n = 3).

# Supplementary Tables

**Supplementary Table 1**. Putative regulatory element chromatin marks

| Region | Coordinates | DHS Width (bp) | DNaseI HS | H3Kme1 | H3K27Ac |
|---|---|---|---|---|---|
| I | chr10:70973319-70974076 | 757 | HeLa, K562, HepG2, HRPEpiC | K562 | K562 |
| II | chr10:70975656-70976016 | 360 | HUVEC, HeLa, k562, HMVEC, HRPE | HUVEC, K562, NHEK, NHLF | HUVEC, NHEK, K562 |
| III | chr10:70976807-70977827 | 1020 | HUVEC, HeLa, K562, HMVEC, HRPE, A549, Stellate, Urothel, Myometr, Medullo, HMEC, Osteoblasts, PanIsletD, HPDE6, FibroP | HUVEC, K562, NHEK, NHLF | HSMM, HUVEC, NHEK, NHLF |
| IV | chr10:70979612-70980551 | 939 | HepG2, HRPE, A549, K562 | K562 | H1-hESC |
| V | chr10:70981067-70981290 | 223 | HepG2, A549 | K562, NHEK | GM12878, H1-hESC, NHEK |
| VI | chr10:70982000-70982955 | 955 | HUVEC, HeLa, K562, HepG2, HMVEC, HRPE, A549, Stellate, Urothel, Myometr, Medullo, HMEC, HMEC, Osteoblast, PanIsletD, HPDE6, FibroP | GM12878, H1-hESC, HSMM, HUVEC, K562, NHEK, NHLF | GM12878, H1-hESC, HSMM, HUVEC, K562, NHEK |
| VII | chr10:70984649-70985565 | 916 | HUVEC, HeLa, HepG2, HMVEC, HRPE, Medullo, HMEC, HPDE6, FibroP | GM12878, HSMM, HUVEC, K562, NHEK | GM12878, HUVEC, NHEK |
| VIII | chr10:70986715-70987113 | 398 | HepG2, Medullo | | |
| IX | chr10:70988887-70990294 | 1407 | HUVEC, K562, HepG2, HMVEC, HRPE, A549, Stellate, Urothel, Medullo, HMEC, Osteoblast, PanIslet, FibroP | GM12878, H1-hESC, HSMM, HUVEC, K562, NHEK, NHLF | GM12878, K562 |
| X | chr10:70991933-70992631 | 698 | HUVEC, HeLa,HepG2, HMVEC, HRPE, A549, Stellate, Urothel, Myometr, Medullo, HMEC, HMEC, Osteoblast, PanIsletD, HPDE6, FibroP | H1-hESC, HSMM, HUVEC, NHEK, NHLF | HSMM, NHEK, NHLF |
| XI | chr10:70997194-70998073 | 879 | HUVEC, HeLa, K562, HMVEC, HRPE, A549, Stellate, Urothel, Myometr, Medullo, HMEC, Osteoblasts, PanIsletD, HPDE6, FibroP | GM12878, H1-hESC, HSMM, HUVEC, K562, NHEK, NHLF | |

**Supplementary Table 2**. Haplotype frequencies across the regulatory elements

| Region | Coordinates | SNPs (common) | Common Haplotypes (>1%) | Rare Haplotypes (<1%) | Total Haplotypes |
|---|---|---|---|---|---|
| I | chr10:70973319-70974076 | 9(4) | 5 | 6 | 11 |
| II | chr10: 70975656-70976016 | 4(2) | 2 | 3 | 5 |
| III | chr10:70976807-70977827 | 11(8) | 5 | 6 | 11 |
| IV | chr10:70979612-70980551 | 13(2) | 3 | 12 | 15 |
| V | chr10:70981067-70981290 | 2(1) | 2 | 1 | 3 |
| VI | chr10:70982000-70982955 | 21(9) | 8 | 19 | 27 |
| VII | chr10:70984649-70985565 | 15(4) | 5 | 14 | 19 |
| VIII | chr10:70986715-70987113 | 7(4) | 4 | 4 | 8 |
| IX | chr10:70988887-70990294 | 26(12) | 13 | 55 | 68 |
| X | chr10:70991933-70992631 | 15(7) | 6 | 9 | 15 |
| XI | chr10:70997194-70998073 | 13(7) | 7 | 14 | 21 |
| | **Total** | 136(60) | 60 | 143 | 203 |

**Supplementary Table 3**. 1000 Genomes individual IDs and ancestries

| Catalog ID | Ancestry | Sex |
|---|---|---|
| NA19399 | LUHYA IN WEBUYE, KENYA | Female |
| NA19428 | LUHYA IN WEBUYE, KENYA | Male |
| NA18557 | YORUBA IN IBADAN, NIGERIA | Male |
| NA19007 | JAPANESE IN TOKYO, JAPAN | Male |
| HG01133 | COLOMBIAN IN MEDELLIN, COLOMBIA | Male |
| HG00346 | FINNISH IN FINLAND | Female |
| NA19473 | LUHYA IN WEBUYE, KENYA | Female |
| NA18861 | YORUBA IN IBADAN, NIGERIA | Female |
| NA19346 | LUHYA IN WEBUYE, KENYA | Male |
| NA19457 | LUHYA IN WEBUYE, KENYA | Female |
| HG01437 | COLOMBIAN IN MEDELLIN, COLOMBIA | Male |
| NA19438 | LUHYA IN WEBUYE, KENYA | Female |
| HG00189 | FINNISH IN FINLAND | Male |
| NA19150 | YORUBA IN IBADAN, NIGERIA | Male |
| HG01626 | IBERIAN POPULATIONS IN SPAIN | Female |
| NA19445 | LUHYA IN WEBUYE, KENYA | Female |
| HG01079 | PUERTO RICAN IN PUERTO RICO | Male |
| GM19707 | AFRICAN ANCESTRY IN SOUTHWEST USA | Female |
| GM12144 | CEPH/UTAH | Male |

**Supplementary Table 4**. Pre- and post-mutagenesis rare allele frequencies

| SNPid | Pre-mutagenesis | Post-mutagenesis |
|---|---|---|
| rs7089312 | 0.125 | 0.2857 |
| rs5030941 | 0.125 | 0.2857 |
| rs2394529 | 0.1667 | 0.2727 |
| rs141651118 | 0.2 | 0.3 |
| rs147449838 | 0.2 | 0.4 |
| rs55755107 | 0.1667 | 0.4286 |
| rs4072135 | 0.1667 | 0.35714 |
| rs11813186 | 0.1667 | 0.21429 |
| rs5030945 | 0.125 | 0.38889 |
| rs144643300 | 0.125 | 0.16667 |
| rs12246517 | 0.25 | 0.5 |
| rs4746829 | 0.25 | 0.5 |

**Supplementary Table 5**. Table of significant regulatory variants

| Region | SNP | Allele A | Allele B | MAF | β Luciferase | P-Value |
|---|---|---|---|---|---|---|
| III | rs10762264 | G | A | 0.48 | -3.13 | 5.05E-05 |
| III | rs12241136 | A | T | 0.1 | 2.77 | 4.14E-08 |
| VI | rs78983061 | C | A | 0.06 | 2.21 | 0.003 |
| VI | rs7089277 | T | G | 0.07 | -8.59 | 8.90E-11 |
| VI | rs4746822 | C | T | 0.49 | -6.7 | 8.00E-08 |
| VII | rs4746824 | C | A | 0.22 | 1.18 | -2.00E-06 |
| VII | rs75405157 | T | C | 0.06 | 1.55 | 3.08E-08 |
| VII | rs2394529 | G | C | 0.5 | -1.41 | 1.06E-04 |
| VIII | rs9645501 | G | A | 0.34 | -2.08 | 9.04E-08 |
| VIII | rs147449838 | G | A | <.01 | -1.23 | 4.35E-04 |
| VIII | rs200216341 | G | A | <.01 | -1.31 | 1.71E-04 |
| IX | rs1983128 | G | A | <.01 | -4.45 | 9.49E-09 |
| X | rs5030945 | T | C | 0.45 | -28.14 | 2.19E-10 |
| X | rs874557 | A | G | 0.45 | -31.58 | 1.54E-06 |

**Supplementary Table 6**. Concordant SNP effects in four different cell lines

| Region | SNP | Allele A | Allele B | MAF | P-value | β Luc A549 | β Luc K562 | β Luc Fibroblast | β Luc HepG2 |
|---|---|---|---|---|---|---|---|---|---|
| III | rs10762264 | G | A | 0.48 | 5.05E-05 | -2.368 | N.E. | -8.9581 | -3.13 |
| VI | rs4746822 | C | T | 0.49 | 8.00E-08 | -2.724 | N.E. | -3.86 | -6.7 |
| VII | rs2394529 | G | C | 0.5 | 1.06E-04 | N.E. | N.E. | 3.137 | -1.41 |
| VIII | rs9645501 | G | A | 0.34 | 9.04E-08 | -29.078 | -115.44 | -15.29 | -2.08 |

**Supplementary Table 7**. Discordant SNP effects in four different cell lines

| Region | SNP | Allele A | Allele B | MAF | β Luc A549 | β Luc K562 | β Luc Fibroblast | β Luc HepG2 |
|---|---|---|---|---|---|---|---|---|
| X | rs5030945 | T | C | 0.45 | 5.04 | 66.89 | -3.13 | -28.14 |
| X | rs874557 | A | G | 0.45 | 28.37 | 62.46 | -2.3 | -31.58 |

**Supplementary Table 8**. RPKM for primary hepatocytes and HepG2 cells

| | Hepatocytes (RPKM) | HepG2 (RPKM) |
|---|---|---|
| **HK1** | 6.5 | 0.7 |
| **HK2** | 1.3 | 27.2 |
| **HK3** | 6.6 | 0 |
| **GCK** | 1 | 0.1 |
| **HKDC1** | 1.2 | 28.3 |

**Supplementary Table 9**. Amino acid percent identity matrix

| | GCK | HK3 | HKDC1 | HK1 | HK2 |
|---|---|---|---|---|---|
| GCK | 100 | 52.41 | 51.52 | 53.13 | 54.64 |
| HK3 | 52.41 | 100 | 52.97 | 53.23 | 55.93 |
| HKDC1 | 51.52 | 52.97 | 100 | 70.8 | 67.79 |
| HK1 | 53.13 | 53.23 | 70.8 | 100 | 72.71 |
| HK2 | 54.64 | 55.93 | 67.79 | 72.71 | 100 |

**Supplementary Table 10**. DNA percent identity matrix

| | GCK | HK3 | HK1 | HKDC1 | HK2 |
|---|---|---|---|---|---|
| GCK | 100 | 53.18 | 50.35 | 48.77 | 49.46 |
| HK3 | 53.18 | 100 | 57.31 | 57.86 | 60.38 |
| HK1 | 50.35 | 57.31 | 100 | 63.44 | 63.09 |
| HKDC1 | 48.77 | 57.86 | 63.44 | 100 | 64.12 |
| HK2 | 49.46 | 60.38 | 63.09 | 64.12 | 100 |

**Supplementary Table 11**. Primer sequences for amplifying regulatory elements

| Region | Forward | Reverse |
|---|---|---|
| I | CGGTACCTGAGCTCGAAGGCCACAGGTTTTTAGATCTCCTT | TATCCTCGAGGCTAGTGGTCCTGGCGCATCTTGTA |
| II | CGGTACCTGAGCTCGTCCTGAGCATTCCTTGTCTTCG | TATCCTCGAGGCTAGGGAACAGTTGGGAATTTTATGGAAA |
| III | CGGTACCTGAGCTCGCCAGATGGGTGATACGCTTT | TATCCTCGAGGCTAGCCTAATTCCCCCACTAGTTGATT |
| IV | CGGTACCTGAGCTCGAATGGAAGAATCTGCCCCAACA | TATCCTCGAGGCTAGTGGCCTTTCCGAGCAAACAT |
| V | CGGTACCTGAGCTCGTTGTCACTTTGCTCAGGGAACTTG | TATCCTCGAGGCTAGGACCTTGGCTAAGTCCTCCTGCT |
| VI | CGGTACCTGAGCTCGAGAGGAAGTGTGGCCCCTTACA | TATCCTCGAGGCTAGGCCATGAAAAATACACACTGAAAACC |
| VII | CGGTACCTGAGCTCGCCCTTAGTGCCTGGGACGTG | TATCCTCGAGGCTAGTCCCCAGGTGAGAGGTGAGG |
| VIII | CGGTACCTGAGCTCGTCTGCAGCCCTCTCACCTCA | TATCCTCGAGGCTAGGAATGGCCCTGACGAAGGTG |
| IX | CGGTACCTGAGCTCGTCCCATGTCAGGGTCCCAGT | TATCCTCGAGGCTAGCCCTCCATGAACTGCCTTGG |
| X | CGGTACCTGAGCTCGTCCTCAGATTCTGCCCTTCAGA | TATCCTCGAGGCTAGGAATGGCCCTGACGAAGGTG |
| XI | CGGTACCTGAGCTCGTCCCACCACTAGAATTGTGGGTTT | TATCCTCGAGGCTAGCATGGTTTGCCAGGTCCACA |

**Supplementary Table 12**. Primer sequences for site directed mutagenesis

| SNP ID | Forward | Reverse | Edit |
|---|---|---|---|
| kg0002 | CTTGTCTTCGGTTCA**C**AATAAACCTGTAAAA | TTTTACAGGTTTATTGTGAACCGAAGACAAG | G>C |
| rs5030937 | TACCAGCAGGTCACA**C**TGATTTATGTTATGA | TCATAACATAAATCAGTGTGACCTGCTGGTA | T>C |
| rs5030938 | TTTATGTTATGACGA**C**TTAACTTTGGACTTA | TAAGTCCAAAGTTAAGTCGTCATAACATAAA | T>C |
| rs9645501 | ACCTTCTCAGCCCTC**A**TTCCATTCTGCTTTC | GAAAGCAGAATGGAATGAGGGCTGAGAAGGT | G>A |
| rs147449838 | CACATGCGGCTCTCC**A**ATGACACCCTTTTGG | CCAAAAGGGTGTCATTGGAGAGCCGCATGTG | G>A |
| rs200216341 | AGGCGGTTCCGGGCT**A**AGATGGAGAAGGGCC | GGCCCTTCTCCATCTTAGCCCGGAACCGCCT | G>A |
| kg0001 | TGGCTGCTCCTCGCC**A**CACATCCTGTCTTCC | GGAAGACAGGATGTGTGGCGAGGAGCAGCCA | G>A |
| rs5030945 | CCTTTATCCGAGTTT**T**CCCAGGCTGGGAGCA | TGCTCCCAGCCTGGGAAAACTCGGATAAAGG | C>T |
| rs5030946 | CCAGGCTGGGAGCAC**A**AGCTCTCTCTTCCCA | TGGGAAGAGAGAGCTTGTGCTCCCAGCCTGG | C>A |
| rs144643300 | ACCGCAAGGGCTGTG**T**TCACATTTGCCTTCT | AGAAGGCAAATGTGAACACAGCCCTTGCGGT | C>T |
| rs4746828 | GTGCTCACATTTGCC**T**TCTCCAGTTGACATG | CATGTCAACTGGAGAAGGCAAATGTGAGCAC | C>T |
| rs12241136 | AAAATCCAACTACTT**T**TAAGATTTAAACCTA | TAGGTTTAAATCTTAAAAGTAGTTGGATTTT | A>T |
| rs10998649 | GGCAAGGCTGGTCTC**G**AACTCCTGACCTCAGGCAATCTGCC | GGCAGATTGCCTGAGGTCAGGAGTTCGAGACCAGCCTTGCC | T>G |
| rs7914256 | GGCAAGGCTGGTCTCTAACTCCTGA**T**CTCAGGCAATCTGCC | GGCAGATTGCCTGAGATCAGGAGTTAGAGACCAGCCTTGCC | C>T |

**Supplementary Table 13**. Primer sequences for qPCR

| | Forward | Reverse |
|---|---|---|
| Bactin | GTGGCCATCTCTTGCTGCAAG | GGGAAATCGTGCGTGACATTAAG |
| HKDC1 | GGTCAGGATGCTGCCCACCT | CCCAAGATCCAGGGCGAGAA |
| HK1 | TGAAGTCGGCCTGATCATCG | TCCTCCCCTCGTCTCCTTCC |
| HK2 | GGGTCCTGCTGGTCCGTGTT | TCCTGCGGGATGGCGTAGAT |
| HK3 | GAGGAGACCCTGGCCCCATT | CCTTCCGCATCTGTGCCTGA |
| GCK | TGGATGTGGTGGCAATGGTG | GATCATGCCGACCTCGCACT |
| ratHK1 | TTAACCCGCTTGGGAGTGGA | GGCTGATCGGAAGGAGACGA |
| ratHK2 | AGCGACTTCGCTCCACCATC | GGAGACGCTTGGCAAAATGG |
| ratHK3 | CTCTTCCAGGATGCGCCTGT | TCCCCTCTGTGGATGGTGGT |
| ratGCK | GGCACTGCCGAGATGCTCTT | GAAGCCCAGGGGCAGTTTCT |
| ratBactin | CACTGCCGCATCCTCTTCCT | GGAACCGCTCATTGCCGATA |