Non-coding variants disrupting a tissue-specific regulatory element in *HK1* cause congenital hyperinsulinism

Matthew N. Wakeling[1]*, Nick D. L. Owens[1]*, Jessica R. Hopkinson[1], Matthew B. Johnson[1], Jayne A.L. Houghton[2], Antonia Dastamani[3], Christine S. Flaxman[1], Rebecca C. Wyatt[1], Thomas I. Hewat[1], Jasmin J. Hopkins[1], Thomas W. Laver[1], Rachel van Heugten[2], Michael N. Weedon[1], Elisa De Franco[1], Kashyap A. Patel[1], Sian Ellard[2], Noel G. Morgan[1], Edmund Cheesman[4], Indraneel Banerjee[5,6], Andrew T. Hattersley[1], Mark J. Dunne[6], International Congenital Hyperinsulinism Consortium[¥], Sarah J. Richardson[1] and, Sarah E. Flanagan[1]

1. Institute of Biomedical and Clinical Science, University of Exeter Medical School, UK
2. Exeter Genomics Laboratory, Royal Devon and Exeter NHS Foundation Trust, Exeter, UK
3. Endocrinology Department, Great Ormond Street Hospital for Children, London, UK
4. Department of Paediatric Pathology, Royal Manchester Children's Hospital, Oxford Road, Manchester, UK
5. Department of Paediatric Endocrinology, Royal Manchester Children's Hospital, Oxford Road, Manchester, UK
6. Faculty of Biology, Medicine and Health, The University of Manchester, Oxford Road, Manchester, UK

* The authors contributed equally to this work.

¥ A list of the authors and their affiliations appears at the end of the paper.

Corresponding author:
Dr Sarah E. Flanagan
Associate Professor in Genomic Medicine
RILD Building
University of Exeter Medical School,
Barrack Road,
Exeter, EX2 5DW
UK
S.Flanagan@exeter.ac.uk

Introductory paragraph

Gene expression is tightly regulated with many genes exhibiting cell-specific silencing when their protein product would disrupt normal cellular function[1]. This silencing is largely controlled by non-coding elements and their disruption might cause human disease[2]. We performed gene-agnostic screening of the non-coding regions to discover new molecular causes of congenital hyperinsulinism. This identified 14 non-coding *de novo* variants affecting a 42bp conserved region encompassed by a regulatory element in intron 2 of *Hexokinase 1* (*HK1*). HK1 is widely expressed across all tissues except for liver and pancreatic beta-cells and is thus termed a "disallowed gene" in these specific tissues. We demonstrated that the variants result in a loss of repression of *HK1* in pancreatic beta-cells, thereby causing insulin secretion and congenital hyperinsulinism. Using epigenomic data accessed from public repositories, we demonstrated that these variants reside within a regulatory region that we determine to be critical for cell-specific silencing. Importantly, this has revealed a disease mechanism for non-coding variants that cause inappropriate expression of a disallowed gene.

Main text

Genetic discovery in Mendelian disease has focused on identifying highly penetrant variants affecting the function of genes expressed in clinically affected tissue(s). Whilst this approach has proven successful the underlying aetiology of over 3000 presumed monogenic diseases remains undefined[2-5], of which many have significant clinical and genetic heterogeneity[6,7]. Congenital Hyperinsulinism (CHI), is characterised by inappropriate insulin secretion during hypoglycaemia. It is a clinically and genetically heterogeneous disease where, despite extensive sequencing efforts, the underlying aetiology is not known in up to 50% of individuals[8,9].

To identify new aetiologies, we performed whole-genome sequencing (WGS) on 135 individuals with biochemically confirmed, persistent CHI without an identified disease-causing variant in a known gene (**Supplementary Table 1**). We initially searched for genes containing new coding *de novo* or biallelic variants in two or more individuals. When no such genes were found, we turned to the non-coding genome and searched for *de novo* copy number variants. This identified a single locus within intron 2 of Hexokinase 1 (*HK1*) containing two heterozygous ~4.5Kb deletions with a 2,787bp overlap in two unrelated individuals (**Extended Data Figure 1**). Digital droplet PCR confirmed the deletions in both probands and an affected twin.

We next searched for *de novo* single nucleotide variants and indels within the minimal deleted region within our whole genome sequencing discovery cohort and in a replication cohort of 27 individuals with CHI who had undergone pancreatectomy as part of routine clinical care (**Supplementary Table 1**). This identified 7 different *de novo* variants in 12 probands. In two cases the variants had been inherited by similarly affected offspring (**Figure 1**). All variants were new[10,11] and within a 42bp conserved region constrained against variation in gnomAD.[12]

Overall, the finding of 9 different *de novo* variants in 14 probands, co-segregating with disease in three additional family members provides overwhelming evidence of disease-causality.

All 17 individuals with a *HK1* non-coding variant had severe early-onset CHI (median age at diagnosis: Birth [IQR: 0-14 d]). These variants were not associated with macrosomia at birth (median birthweight z score: 0.61 [IQR: 0.10-1.84]). This suggests that insulin secretion was not markedly increased *in* utero given its role as a potent growth factor during fetal development. In all cases the hyperinsulinism persisted, the eldest individual still known to be affected was 18 years of age. These findings suggest that the variants act to disrupt glucose-induced insulin secretion throughout post-natal life and into adulthood but have less impact during fetal development.

The *HK1* variants caused a beta-cell specific defect with no common additional features between patients. In five individuals, medical management was ineffective, leading to pancreatic resection. Of these, the only case that had resolution of the CHI developed insulin-dependent diabetes following near-total pancreatectomy. Histopathological analysis of resected pancreatic tissue (n=2) demonstrated a discrete pathology, distinguishable from diffuse disease resulting from *ABCC8* variants, the commonest cause of CHI (**Extended Data Figure 2**). An overview of the clinical features is provided in **Supplementary Tables 2-3**.

HK1 encodes a glycolytic enzyme that is silenced in the pancreas and liver but expressed in all other mature tissues (**Extended Data Figure 3a-c**) where it supports glucose metabolism, ensuring cell survival[13]. Within the pancreatic beta-cell, glucose is phosphorylated to glucose-6-phosphate by Glucokinase (GCK; Hexokinase 4) which, due to a low-binding affinity (Km ~8mM) acts as the pancreatic glucose-sensor coupling insulin release to the prevailing glucose concentration[14]. Hexokinase 1 has a markedly higher affinity for glucose (Km <50µM)[14]. Silencing of *HK1* in favour of *GCK* in the beta-cell therefore ensures appropriate glucose-sensing, minimising insulin release at low glucose levels. Whilst bi-allelic loss-of-function *HK1* variants have been reported to cause non-spherocytic haemolytic anaemia[15], dominant or recessive coding variants have not been described in individuals with defects in glucose homeostasis; in keeping with the absence of HK1 protein in beta-cells.

In resected pancreatic tissue of individuals with non-coding variants, we found that HK1 was expressed and co-localised with insulin in the islets of affected tissue but not in controls (**Figure 2**). HK1 did not co-localise with glucagon (secreted by pancreatic alpha-cells) suggesting that the impact of the variant was beta-cell specific (**Extended Data Figure 4**). These results confirmed that the variants cause HK1 to be inappropriately expressed in the pancreatic beta-cell and explain why there is increased insulin secretion in these individuals during hypoglycaemia.

We identified a putative islet cis-regulatory domain encompassing our critical region[16] bound by a broad set of islet transcription factors, with NKX2-2 and FOXA2 most prominent (**Figure 3**). Single-nuclei ATAC-seq (snATAC-seq) data in human islets[17] revealed a peak of open chromatin in beta-cells and showed that the relevant *HK1* promoter remains

accessible (**Figure 3b**). Human islet Hi-C data[18] confirmed that both our region and the relevant *HK1* promoter are contained within a well-insulated domain (**Extended Data Figure 5**), without evidence of contact to distal loci (**Figure 3d, Extended Data Figure 5b**) further supporting direct regulation by the critical region on the promoter.

During development, HK1 is highly expressed in embryonic stem cells but is progressively downregulated during pancreatic cell-type differentiation becoming extinguished during beta-cell maturation[19,20] (**Figure 3a, Extended Data Figure 3d**). GCK shows increased expression during differentiation as HK1 is downregulated (**Figure 3a**). Despite being absent from beta-cells, HK1 is expressed in mature stellate- and duct-cells suggesting differential regulation across pancreatic cell-types (**Extended Data Figure 3e**).

Chromatin accessibility data shows our critical region remains closed over pancreatic-cell differentiation until pancreatic progenitor stages, when it becomes accessible and bound by a complement of factors in the pancreatic progenitor regulatory network[21,22] (**Figure 3c, Extended Data Figure 6b-c**). This suggests the region is necessary for HK1 repression in late pancreatic development and mediating beta-cell specific control in adult cells but that it does not have a role in the reduction of HK1 expression between embryonic stem cells and early pancreatic progenitors (**Figure 3a**).

Analysis of histone modifications during pancreatic differentiation, in human islets[21,23], cell sorted human pancreatic alpha, beta and exocrine cells[24] and the EndoC-βH1 beta-cell line[25], revealed that repression is actively maintained. The critical region is marked by a bivalent state encompassed by focal peaks of active enhancer marks H3K4me1 and H3K27ac and a broader and dominant polycomb H3K27me3 repressive domain (**Extended Data Figure 6d**). This domain is diminished in exocrine cells supporting that this mode of repression is specific to endocrine cell types (**Extended Data Figure 5d**).

Analysis of transcription factor motif families revealed that the variants disrupt transcription factor binding sites for FOX, NKX2 and NFAT families (**Supplementary Table 5, Figure 3e**), which each have family members expressed in beta-cells (**Extended Data Figure 7**). This is supported by ChIP-seq data revealing that NKX2-2 and FOXA2 were prominently bound at this region in islets (**Figure 3b**). FOXA2 loss-of-function causes CHI[26], and NKX2-2 is able to recruit a large repressive complex that regulates beta-cell specification through DNA methylation[27] and may therefore have a role in maintaining the repressive epigenetic state at the *HK1* locus[21,23-25,28]. Multiple members of the NFAT family are expressed in beta-cells (**Extended Data Figure 7**), with NFATC2 inducing beta-cell proliferation in human islets[29]. Analysis of snATAC-data from human islets revealed endocrine cells present in hormone-high and hormone-low states[17], with the former associated with increased promoter accessibility over secreted endocrine genes and the latter over cell-cycle genes. Our critical region is accessible in the hormone-high state of alpha, beta, and delta-cells, and when compared between hormone-low state cells, only remains accessible in beta cells (**Extended Data Figure 6a**). This suggests that HK1 repression is actively maintained during beta-cell proliferation and not during alpha or delta-cell proliferation, which given NFATC2's capacity to induce beta-cell proliferation, supports this factor in maintaining HK1 repression.

Discovering non-coding regulatory variants controlling *HK1* and access to a broad range of publicly available epigenomic data[16-25,30,31] enabled us to identify a regulatory element critical for the selective silencing of an otherwise ubiquitously expressed gene within a single cell-type. It is interesting to note that interruption of transcription factor binding is implicated in each of the 9 distinct variants and the locus is decorated by both active and repressive epigenetic marks suggesting that continual transcription factor binding is necessary for maintenance of repression in beta-cells. Future work should prioritise regulatory regions surrounding disallowed genes with similar epigenetic marks and to explore whether different regulatory mechanisms exist to silence disallowed genes within the beta-cell.

Over 60 beta-cell disallowed genes have been described, although the mechanism(s) controlling cell-specific silencing in humans have not been fully determined[1,32]. Linkage analysis of *HK1* to CHI[33] and HK1 expression in beta-cells[34] have been reported in several patients, but the genetic aetiology was not established. Promoter variants in the beta-cell disallowed gene, *SLC16A1* have also been reported in two families with exercise–induced hyperinsulinism. However, these variants increased transcription across cell types leading to the hypothesis that differences in the post-transcriptional regulation of mRNA across tissues could explain the beta-cell-specific phenotype[35].

The identification of non-coding *HK1* variants in individuals with CHI represents a rare example of regulatory non-coding variants affecting a gene in which coding variation does not cause the same phenotype[36]. These findings highlight a role for undiscovered regulatory variants causing disease through inappropriate expression of a normally functioning protein in a specific cell-type.

Non-coding variants affecting the regulation of *HK1* cause CHI through its aberrant expression in beta-cells. These findings are important for future efforts to discover non-coding regulatory variants as they establish a critical role for disallowed genes in Mendelian disease.

This manuscript is dedicated to the memory of our colleague and friend, Professor Mark Dunne, who will be greatly missed. Mark was an inspirational scientist whose pioneering research served to improve the lives of children and families affected by Congenital Hyperinsulinism.

Author Contributions Statement

S.E.F. designed the study. A.D, S.E., I.B., S.E.F. and the International Congenital Hyperinsulinism Consortium recruited patients to the study. Ma.N.W*, M.B.J., J.A.L.H., R.V-H., S.E. and S.E.F. performed the molecular genetic analysis and the interpretation of resulting data. Ma.N.W. and Mi.N.W. performed the bioinformatic analysis, T.I.H., J.H., I.B., A.T.H. and S.E.F. analysed the clinical data. E.C. and M.J.D. performed the pathological analysis of the pancreatic tissue, J.R.H, C.S.F., R.W., N.G.M. and S.J.R. designed and performed the immunohistochemistry studies, N.D.L.O. designed and performed the epigenomic data analysis, Ma.N.W., N.D.L.O., J.R.H, M.B.J., T.W.L., E.D-F., K.A.P., A.T.H., M.J.D., S.J.R and S.E.F. prepared the draft manuscript. All authors contributed to the discussion of the results and to manuscript preparation.

Competing Interests Statement

The authors declare no competing interests.

Figure Legends

Figure 1: Schematic representation of the *HK1* gene and the identified variants. a)
Schematic representation of the *HK1* gene (GRCh37/hg19 chr10:71,029,756-71,161,637).
The full-length testis-specific isoform (ENST00000448642) and the shorter ubiquitously
expressed isoform (ENST00000359426) are depicted. The positions of the two ~4.5Kb
deletions within intron 2 of the shorter isoform and the 7 different heterozygous single
nucleotide variants and indels within the minimal deleted region identified in 14 probands
are shown. The number of probands with each variant is provided. An asterisk (*) denotes a
single proband with two *de novo* variants (see **Extended Data Figure 8**). The variant
positions are given according to the GRCh37/hg19 genomic coordinates. **b)** Partial pedigrees
showing inherited *HK1* variants in 3 families. Filled symbols represent individuals with
congenital hyperinsulinism. (‡) Haplotype analysis confirmed that the *HK1* variant had
arisen *de novo* in patient 3.2. M, *HK1* variant; N, no variant; NA, DNA not available.
Pedigrees for the 11 probands with *de novo HK1* variants are not shown.

**Figure 2: HK1 expression is present in the beta-cells of donors with *HK1*-variant CHI but
not in non-hyperinsulinism (Non-HI) controls**. **a)** Staining of HK1 (red); insulin (cyan) and
DAPI (dark blue) in pancreatic tissue resected from patient 6 (**Supplementary Table 2**) with
a *HK1* variant. Confocal imaging of sections demonstrates that HK1 co-localises with insulin
whereas, no HK1 expression is observed in the non-HI control donor. Scale bar –
100μm. **b)** HK1 expression (Median fluorescence intensity (MFI)) is significantly increased in
the *HK1* HI donors (Patient 1 and 6; n=2 donors; 17015 beta-cells), when compared to Non-
HI control donors (n=2 donors; 21408 beta-cells). Two-tailed Mann-Whitney
(U=590830; ****p<0.0001). All cell data within a violin plot with the median (line) and IQR
(dashed lines) presented. Source data provided in Supplementary Information.

**Figure 3: HK1 variants disrupt transcription factor binding sites within a cis-regulatory
region a)** Expression of HK1 and GCK for comparison over the course of pancreatic
differentiation from embryonic stem cells to maturing beta-cells, expression data over a
beta-cell differentiation pseudotime[19], Gaussian process regression median (line) and 95%
confidence interval (shaded regions)I shown. DE – definitive endoderm, GT – gut tube, FG –
foregut, PE – pancreatic endoderm, EP – endocrine precursors, IB – immature beta-cells. See
Extended Data Figure 3 for further HK1 expression. **b)** Transcription factor binding in human
islets[16] and chromatin accessibility from snATAC-seq in the cluster of cells assigned to beta-
cells[17] over *HK1* locus, the region containing the variants is marked as grey box. Chromatin
accessibility in other endocrine cell types is provided in **Extended Data Figure 6c)** Chromatin
accessibility during pancreatic differentiation over critical region, stages: ES – embryonic
stem cell, DE – definitive endoderm, GT – gut tube, Early/Late PP – pancreatic progenitor,
describe ATAC-seq from in vitro differentiation[21]cellscells give accessibility in snATAC-seq[17]
HiC data in human islets**d)**[18] is contained within a well-insulated domain, heatmap gives log
scale contact frequencies, white triangle and circles mark chromatin loops called in same

study that span HK1, additional loops marked in **Extended Data Figure 5**mark chromatin loops called in same study that span HK1, additional loops marked in . Region ± 850kb of HK1. Region ± 850kb of HK1. **e)** Transcription factor motif families disrupted by the variants, shown are those motif families maximally disrupted by each variant (red font). Notable secondary motif families given in **Extended Data Figure 8 and Supplementary Table 6**. Left, scatterplot of normalised motif scores for the reference versus mutated score (motif score normalised by maximal motif score), black line gives 1-1 line of equal motif. Number gives final two digits of chromatin position, e.g. 45 – 71108645. Right gives sequence logo for the three maximally disrupted families. Members of each family are expressed in beta-cells **Extended Data Figure 7**.

METHODS

Subjects

Congenital hyperinsulinism (CHI) was defined as an inappropriately high level of plasma insulin at the time of hypoglycaemia associated with inappropriately supressed ketones and free fatty acids presenting within the first 12 months of life. The definition of hypoglycaemia was based on the recommendations of the Pediatric Endocrine Society (blood glucose <2.8mmol/L in the presence of detectable insulin)[37]. Clinical details of the subjects are provided in **Supplementary Table 1**. Subjects with CHI were recruited by their clinicians for molecular genetic analysis to The Exeter Genomics Laboratory. Disease-causing variants in the known CHI genes had been excluded by targeted-next generation sequencing in all cases[38].

Ethical considerations

This study was approved by the North Wales Research Ethics Committee (17/WA/0327, IRAS project ID: 231760) and was conducted in accordance with the Declaration of Helsinki, with all subjects or their parents providing informed consent for genetic testing. All tissue samples were studied with full ethics approval (West of Scotland Research Ethics Committee, reference: 20/WS/0074 IRAS project ID: 283620 or nPOD)[39]. No participant received compensation for entering this study.

Sample collection and DNA Extraction

Peripheral whole blood samples (1-5mls) were taken in EDTA blood collection tubes. Automated DNA extracted was undertaken on a Chemagic STAR (Hamilton Bonaduz AG, Switzerland) using the Chemagic STAR DNA blood extraction kit (CMG-1756, [Perkin Elmer, Waltham,USA]).

Whole-genome sequencing

Whole-genome sequencing of DNA extracted peripheral blood leukocytes was completed on 135 probands with CHI (n=3 with Illumina HiSeq 2500, n=69 with Illumina HiSeq X10, n=63 with BGISeq-500). The mean read depth across the whole genome was 36.9 (stddev 4.9). An additional 191 family members were also sequenced. The sequence data was aligned using BWA MEM version 0.7.15, and processed using a pipeline based on the GATK best practices (Picard version 2.7.1, GATK version 3.7). Variants were annotated using Alamut batch standalone version 1.11 (Rouen, France).

CNVs were called across the genome by SavvyCNV[40] using a bin size of 2Kbp. We also developed a new tool (FindLargeInsertSizes, available at https://github.com/rdemolgen/SavvySuite) to detect small (≥1Kbp) deletions, insertions, inversions, and translocations using read pair information. We used this tool to screen all whole genome sequenced samples for structural variants within *HK1* intron 2.

When a variant was identified, Sanger sequencing (for single nucleotide variants/indels) or digital droplet PCR (for deletions) was performed on samples from each available family member to confirm genome sequencing results.

Digital droplet PCR (ddPCR): was performed on leukocyte DNA from patients 1 and 2.1 and their unaffected parents to confirm the results of the whole-genome sequencing deletion analysis. ddPCR was also performed on the affected twin brother of patient 2.1 and the 27 individuals with CHI of unknown cause who had undergone pancreatectomy.

Relevant primers (**Supplementary Table 4**) were used in reactions for droplet generation, PCR and detection using the Bio-Rad QX200 ddPCR EvaGreen system (Hercules, CA, USA) to search for copy number changes. Reactions were performed as per the manufacturer's recommendations with an annealing/extension temperature during PCR of 59°C. All data were analysed using Bio-Rad QuantaSoft software version v1.6.6.0320.

Sanger sequencing of the *HK1* regulatory element

Sanger sequencing was performed on DNA extracted from peripheral blood leukocytes from 27 individuals with CHI of unknown cause who had undergone pancreatectomy. Briefly, a 397bp genomic region (GRCh37/hg19, Chr10:71,108,536-71,108,932) within intron 2 of *HK1* was amplified by PCR (primer sequences are listed in **Supplementary Table 4**). PCR products were sequenced on an ABI3730 capillary machine (Applied Biosystems, Warrington, UK) and analysed using Mutation Surveyor v3.24 software (SoftGenetics, State College, PA, USA). When a variant was identified, samples from family members were tested to investigate co-segregation and microsatellite analysis using the PowerPlex kit (Promega, Southampton, UK) was performed to confirm family relationships.

Histopathology

Formalin-fixed paraffin-embedded pancreatic tissue was available for immediate analysis from two individuals with *HK1* variants (patients 1 and 6, Supplementary Table 2) age-matched non-HI controls (n=2) and age-matched tissues from individuals with HI due to an *ABCC8* variant (HI controls n=2). Immunohistochemistry was performed on 5-μm-thick sections of tissue[41]. For high-content quantification of tissue, sections were first digitized using the 3DHistech Pannoramic 250 Flash II slide scanner (3DHISTECH, Budapest, Hungary) and quantified using QuPath (available at https://qupath.github.io)[42].

Immunofluorescence

After dewaxing and rehydration, pancreatic samples from HK1 patients 1 and 6 (**Supplementary Table 2**) and 4 age-matched control tissues (HI controls (individuals with *ABCC8* variants (n=2); non-hyperinsulinism donors (n=2)) were subjected to heat–induced epitope retrieval (HIER) in 10mM citrate pH6 buffer for 20 minutes. After blocking, the sections were probed in a sequential manner with rabbit monoclonal anti-hexokinase 1 (Abcam ab150423; 1/100 overnight), followed by detection with goat-anti-rabbit Alexa Fluor™ 488 (Invitrogen, Paisley, U.K). Following washes, sections were then probed with mouse monoclonal anti-glucagon (Abcam ab10988; 1/2000 for 1h) followed by detection with goat-anti-mouse Alexa Fluor™ 555. Finally, sections were probed with guinea-pig anti-insulin (Agilent; C#IR002; 1/5 for 1h), followed by detection with goat-anti-guinea-pig Alexa Fluor™ 647 in the presence of DAPI (1μg/ml) to identify cell nuclei. Sections were not stripped between successive antibody incubations. After mounting, the sections were imaged via a Leica DMi8 confocal microscope (Leica Microsystems UK, Milton Keynes, UK) (using a focal plane of 0.4μm) and the distribution of HK1, insulin and glucagon examined in multiple islets. In addition, sections were scanned at 40x magnification using an Akoya Biosciences Vectra® Polaris™ Automated Quantitative Pathology Imaging System for quantification analyses. The quantification of the HK1 expression in islets was determined using the Random Forest Classifier Module (Version 3.2.1851.354), DenseNet AI V2 and HighPlex FL v4.04 modules included in Indica Labs HALO Image analysis platform (Version 3.2.1851.354). The Random Forest Classifier was used to identify the islets, then the DenseNet AI V2 module was used to identify beta and alpha cells within the islets. This enabled exclusion of HK1 positive non-islet cells such as red blood cells, stellate cells and endothelial cells which are frequently found in proximity to the islet. The median fluorescence intensity (MFI) of HK1 expression in beta and alpha cells were then calculated using the HighPlex FL v4.04 module. This pipeline was applied to all corresponding tissue sections. GraphPad Prism V9.2.0 (332) was used to demonstrate the median and IQR of beta and alpha cells MFI in each of the donors assessed.

Epigenomic analysis

All published ChIP-seq, ATAC-seq, RNA-seq and scRNA-seq datasets used were obtained from accessions provided in **Supplementary Table 5**. All code and source data required for all analysis ,

including to determine disrupted motifs and generate figures is provided in https://github.com/owensnick/HK1FigureNotebook.jl and DOI:10.5281/zenodo.6815326.

Public Bulk ChIP and ATAC-seq

ChIP-seq and ATAC-seq reads aligned with Bowtie2[43] to the GRCh37/hg19 genome with Bowtie2 v2.3.5.1 with default parameters for single end reads and with additional options "-I 0 -X 1000 --no-discordant --no-mixed" for paired-end reads. Alignments were filtered for those with mapping quality > 30 and then reads with identical aligning coordinates were treated as duplicates and collapsed to a single alignment. For islet H3K27me3 data from roadmap epigenomics project, aligned reads were downloaded in BED/TagAlign format[23]. All data is visualised as reads per million using https://github.com/owensnick/GenomeFragments.jl. For human islet single-nuclei ATAC-seq data[17] (GSE160472), reads from the authors' combinatorial barcode approach were aligned as above and separated into cell types using author provided cluster labels: https://github.com/kjgaulton/pipelines/tree/master/islet_snATAC_pipeline.

Human islet Hi-C data was obtained from experiment accession TSTSR043623 and file accession DFF064KIG (.hic file) and TSTFF938730 (bedpe file)[18]. The .hic file from experiment was downloaded and converted to .cool format using hic2cool https://github.com/4dn-dcic/hic2cool with default parameters. Contact matrices were obtained using Cooler[44] at 5kb resolution using KR balance and log-transformed for visualisation. To quantify histogram of contact strengths for all islet loops, loop bases were extended by two 5kb HiC bins up- and downstream and the mean of the KR balance normalised matrix of this region quantified per loop.

scRNA-seq Expression Datasets

scRNA-seq data collected over a time course of pancreatic differentiation[19] projected onto a differentiation pseudotime was obtained from the reference's supplementary table. We identity consistent temporal trends using Gaussian Process (GP) regression, following the approach we have previously applied[45]. Briefly, to stabilise variance we transform data $\log(\alpha y + \beta)$ where y is the gene expression for a given gene and $\alpha = 100$, $\beta = 1$, then perform GP regression using Matern52 kernel and invert the transformation to report GP median and 95% confidence intervals. All GP regression was performed with GaussianProcesses.jl (https://github.com/STOR-i/GaussianProcesses.jl; https://arxiv.org/abs/1812.09064).

For scRNA-seq data in human islets[30] accession GSE101207 we used gene counts per cell for the size healthy donors and normalised by depth per cell. In **Extended Data Figures 3, 7 and 8** we show boxplots of normalised counts for each gene over healthy donors.

Finally, we obtained data on gene expression during beta-cell maturation from the Broad Single Cell portal[20]. In **Extended Data Figures 3, 7 and 8** we show mean normalised counts for each gene over annotated cell types.

Motif analysis

To assess motifs maximally disrupted by the variants, we took all motifs in the non-redundant JASPAR database[46] and found the maximal scoring match that spanned the position of each variant for both reference and mutated sequences using https://github.com/exeter-tfs/MotifScanner.jl. Motif scores are likelihood-ratio scores of motif PWM over a background of the A,C,G,T frequencies in the hg19 genome. We normalised all motif scores, by the maximal score for each PWM and to both primary and secondary candidate motifs that may be disrupted by the variants. We considered two tiers; Tier 1 having normalised motif score ≥ 0.6, and Tier 2 having 0.45 ≤ normalised motif score < 0.45. For each tier we calculated a disruption score by subtracting the mutated sequence score from the reference sequence score and ranked these across all motifs. To report a single disrupted motif family, we grouped multiple motifs with overlapping alignments by family by removing any

trailing numbers from the gene symbol (e.g. NKX2-2 → NKX2, HIC2 → HIC) and all FOX factors we grouped in the FOX family. Tier 1 disrupted motif families include NFAT, NKX2 and FOX (**Figure 3e**). Patient 11 has two de novo variants, the leftmost is shared with other patients and is a candidate for NFAT disruption, and the rightmost interrupts a HIC family motif (**Extended Data Figure 8a**). HIC family members include HIC1 and HIC2, with HIC2 the most greatly expressed in beta-cells (**Extended Data Figure 8b**). Interestingly, HIC2 is a transactivator of SIRT1[47] and the loss of SIRT1 impairs glucose sensing in beta-cells in mice[48]. Tier 2 disrupted motif families include TEAD and SMAD (**Extended Data Figure 8c**). The TEAD family motif shares TTCA consensus with the NFAT family and is an alternative candidate to NFAT, TEAD1 is expressed in beta-cells (**Extended Data Figure 8d**) and plays a critical role in pancreatic progenitors[22], however it should be noted that TEAD1 does not bind the critical region in pancreatic progenitors when the region is bound by FOXA2 (**Figure 3b**). Multiple members of the SMAD family are expressed in beta-cells (**Extended Data Figure 8e**), SMAD factors are signal transducers of TGF-beta signalling and play an important role in beta-cell development, function, and proliferation[49].

Statistics and Reproducibility

No statistical method was used to predetermine sample size as this study involved genetic testing in individuals with a rare monogenic disease.

Immunofluorescence studies were performed on pancreatic samples from 6 donors. This included two individuals with *HK1*-hyperinsulinism (patients 1 and 6, **Supplementary Table 2**), two *ABCC8*-HI controls and two non-HI controls. Each experiment was performed once due to limited sample availability. Representative data from these imaging studies are shown for one of each donor type in **Figure 2a** and **Extended Data Figure 4a-b**.

No data were excluded from any of the experiments described.

Data Availability

All non-clinical data analysed during this study are included in this published article (and its supplementary information files). Clinical and genotype data can be used to identify individuals and is therefore available only through collaboration to experienced teams working on approved studies examining the mechanisms, cause, diagnosis and treatment of diabetes and other beta-cell disorders. Requests for collaboration will be considered by a steering committee following an application to the Genetic Beta Cell Research Bank (https://www.diabetesgenes.org/current-research/genetic-beta-cell-research-bank/). Contact by email should be directed to Prof Sarah Flanagan (S.Flanagan@exeter.ac.uk). All requests for access to data will be responded to within 14 days.

Accession codes and DOI numbers for all ChIP-seq, ATAC-seq, RNA-seq and scRNA-seq datasets are provided in **Supplementary Table 5**. We used the Genome Reference Consortium Human Build 37 (GRCh37) to annotate genetic data. Accession number: GCF_000001405.13. Details of this assembly are provided at: https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/.

Source data

Source data for Figure 2b and Extended Figure 4c are provided in the Additional Supplementary Files. For Figure 3 and Extended Data Figures 3, 5, 6, 7, 8 all source data is contained in https://github.com/owensnick/HK1FigureNotebook.jl and DOI:10.5281/zenodo.6815326.

Code Availability

The code used in this study can be freely downloaded from:

https://github.com/rdemolgen/SavvySuite

https://github.com/owensnick/HK1FigureNotebook.jl

https://github.com/owensnick/GenomeFragments.jl

https://github.com/kjgaulton/pipelines/tree/master/islet_snATAC_pipeline

https://github.com/4dn-dcic/hic2cool

https://github.com/STOR-i/GaussianProcesses.jl

https://github.com/exeter-tfs/MotifScanner.jl

https://qupath.github.io

Extended Data Figure 1: Integrative Genomics Viewer (IGV) screenshot of the ~4.5Kb deleted regions encompassing the regulatory element in intron 2 of *HK1* in patients 1 and 2.1. Both deletions are heterozygous, so reads are present in the deleted regions from the other haplotype, although the read depth is reduced. Reads highlighted in red have a long insert size. Both features are consistent with a deletion. The deleted region in patient 2.1 has detectable break points, as shown by the discontinuity in read depth and the reads clipped at the deletion boundary. In patient 1, the breakpoints do not result in read depth discontinuity or clipped reads as the deletion occurs between two matching repeating sequences.

Extended Data Figure 2: Histological appearance of islets from *HK1*-CHI and *ABCC8*-HI tissue. Panels **A** and **B** illustrate islets from two *HK1* CHI patients (patients 1 and 6 **Supplemental Table 2**) in comparison to age-matched tissue from *ABCC8* CHI patients with diffuse pancreatic disease **(C and D).** The yellow dashed line delineates a single islet in each sample. Nuclear enlargement (B, arrowhead) and nucleomegaly (C, D arrows) was virtually undetected in control islets (n=12/72,145 islet cells) but found in 0.8% of HK1 islets (n=374/46,885 cells) and 4.9% of *ABCC8* islet cells (n=70,929). Inserts illustrate the corresponding haematoxylin and eosin images. Scale bar 50 μm

Extended Data Figure 3: HK1 expression in GTEx a) Boxplots describing gene level HK1 expression in Genotype-Tissue Expression GTEx project ordered by expression with least expression in liver and pancreas. Data shows sum of GTEx v8 isoform level data calculated with RSEM. The GTEx Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on 08/26/21. **b)** Isoform level HK1 expression in GTEX tissues as **a)**, top 3 most expressed isoforms are given. Long HK1 isoform only expressed in testis, short isoform ubiquitously expressed across all tissues. **c)** Transcript models of top three HK1 isoforms given in b). **d)** Expression of HK1 and GCK for comparison in scRNA-seq across alpha and beta-cell maturation, all data quantified by (GSE167880)[20], which quantified their own data over maturation and adult endocrine cells (GSE114297)[31]. **e)** HK1 expression in human islet cell types by scRNA-seq (GSE101207)[30]. Data points from independent human donors, the central lines in the boxplots correspond to the median, boxes span from the first to the third quartiles, and whiskers extend to the furthest data point within 1.5xIQR from the boxes. HK1 expression only in duct and pancreatic stellate cells, expression absent in endocrine cell types.

Extended Data Figure 4: HK1 expression is present in the beta-cells of donors with *HK1*-CHI, but not *ABCC8*-Hyperinsulinism or Non-HI controls. a) Confocal imaging of HK1 (red); glucagon (yellow); insulin (cyan) and DAPI (dark blue) in pancreatic tissue resected from patient 6 (Supplementary Table 2) with a *HK1* variant. HK1 co-localises with insulin, but not with glucagon. Staining in tissue resected from individuals with *ABCC8*-Hyperinsulinism (HI-control) and non-HI control, showed no expression of HK1 in either alpha (glucagon-positive) or beta (insulin-positive cells). Scale bar – 100 μm. **b)** Enlarged image of a portion

of the islet shown in (a) from an individual with *HK1*-HI (Patient 6). This image confirms the expression of HK1 specifically in beta cells (white arrows) and absence of HK1 in alpha cells (orange arrows). Scale bar – 20 μm **c)** HK1 expression (Median fluorescence intensity (MFI)) is higher in the *HK1*-HI donor beta-cells within HK1 positive islets, when compared to alpha cells (Patient 1 - beta-cells (n=1580); alpha-cells (n=161); and Patient 6 - beta-cells (n=2284); alpha-cells (n=9321)). Two-tailed Mann-Whitney (****p<0.0001). This suggests the impact of the *HK1* variant is beta-cell specific. Violin plots shown with the median (dashed line) and IQR (dotted lines) presented. Source data provided in Supplementary Information.

Extended Data Figure 5: Human Islet chromatin conformation at loci spanning HK1 and secreted endocrine genes. a) Histogram of contact strengths of loops in human islet HiC data[18], data shown for all loops called in that study. Vertical line marks the strength of strongest and smallest loop spanning HK1 gene (marked with white arrow in **c)**). **b)** Cumulative frequency plot of contacts shown in **a)**, vertical line marks HK1 loop (white arrow in **c)**), the HK1 loop is in the 86th percentile of all islet HiC loop strengths. **c)** Heatmap of HiC data shown in **a)** for region ± 850 kb of HK1 as **Figure 3d,** now with additional loops within region marked. White arrow gives principal loop spanning HK1 gene marked in **a, b)**. **d,e,f)** HiC data and loops for endocrine cell secreted genes **d)** INS, **e)** SST, and **f)** GCG, shown at same resolution and scale as HK1 locus **c)** for context.

Extended Data Figure 6: Chromatin accessibility, Transcription Factor binding and Epigenetic modifications at the *HK1* locus over pancreatic differentiation a) snATAC-seq data in human islets showing mean chromatin accessibility over nuclei assigned to alpha, beta and delta cell clusters in hormone high and hormone low conditions (GSE160472)[17]. Critical region containing variants encompassed within grey box. Alpha, beta and delta cells each show accessibility in hormone high state, whereas only beta-cells retain accessibility in hormone low cells. Hormone high and low states characterised by level of accessibility over endocrine cell secreted hormone promoter, alpha – GCG, beta – INS, delta – SST. Data show in reads per million (RPM). **b)** Chromatin accessibility of pancreatic differentiation assessed by ATAC-seq (GSE149148)[21], wider locus view of same data shown in **Figure 3c**. Critical region containing variants encompassed within grey box. Stages: ES – embryonic stem cell, DE – definitive endoderm, Early/Late PP – pancreatic progenitor. **c)** Transcription factor binding in pancreatic progenitors (GSE149148)[21] and in Carnegie Stage 16-18 liver bud for FOXA2 (E-MTAB-3061)[22], data shown as a) and b). Liver bud data reveals that whilst FOXA2 binds critical region in pancreatic progenitors and human islets (**Figure 3b**), it does not bind in liver bud, suggesting that critical region encodes pancreas specific regulation of *HK1*. **d)** Epigenetic modifications over *HK1* locus, shown are active marks H3K27ac and H3K4me1 and repressive mark H3K27me3 over in vitro pancreatic cell differentiation (GSE149148)[21] stages as in **b)**; in vitro beta-cell line EndoC-BH1 (GSE118588)[25] and human islets H3K27ac and H3K4me1 (E-MTAB-1919)[16] and H3K27me3 (E087 – roadmap epienomics[23]); cell-sorted pancreatic alpha, beta and exocrine cells (GSE50386)[24]. Grey box encompasses critical region containing variants. Data reveals that

H3K27ac broadly marks short isoform promoter (**Supplemental Figure 3**) at ES cells stage is reduced by DE stage, whilst focal mark over regulatory region is present in EndoC-BH1 and human islets. Similarly, H3K4me1 shows string focal enrichment over regulatory region in EndoC-BH1 and islets. Finally, repressive mark H3K27me3 is absent over the locus over the course of pancreatic cell differentiation and is present in a broad domain in adult islets, EndoC-BH1 cells, beta, cells and to reduced extent in exocrine cells.

Extended Data Figure 7: Expression of family members of disrupted transcription factor binding motifs in critical region: NFAT family, NKX2-2 and FOXA2: **a)** In single-cell RNA-seq in human islets (GSE101207)[30], boxplots of mean normalised counts from independent islet donors in cells assigned to islet cell type clusters (data shown as in **Supplemental Figure 3e**). Data points from independent human donors, the central lines correspond to the median, boxes span from the first to the third quartiles, and whiskers extend to the furthest data point within 1.5xIQR from the boxes. **b)** Over the course of pancreatic differentiation from embryonic stem cells to maturing beta-cells, expression data over a beta-cell differentiation psuedotime[19], Gaussian process regression median and 95% CI shown. ES – embryonic stem cell, DE – definitive endoderm, GT – gut tube, FG – foregut, PE – pancreatic endoderm, EP – endocrine precursors, IB – immature beta-cells, MB – maturing beta-cells. **c)** scRNA-seq across alpha and beta-cell maturation, all data quantified by (GSE167880)[20], which quantified their own data over maturation and adult endocrine cells (GSE114297)[31].

Extended Data Figure 8: Additional transcription factor motifs disrupted by variants
a-b) Sanger sequencing identified 2 *in cis de novo* variants 12 base pairs apart (g.71,108,648C>G and g.71,108,660G>A) in patient 11. g.71,108,648 is affected by 3 different *de novo* substitutions in 6 individuals and resides within a predicted NFATC binding motif. Two further *de novo* variants identified in 3 patients are also predicted to impact on NFATC binding (**Figure 3**). In contrast, further variants affecting g.71,108,660 have not been identified, this suggests that the g.71,108,648C>G is causative of the disease in this patient. Nevertheless, it is of note that g.71,108,660G>A is predicted to affect a HIC family motif **a)** and human islet scRNA-seq (GSE101207)[30] determines HIC2 is expressed in beta-cells **b)** Interestingly, HIC2 is a transactivator of SIRT1[47] and the loss of SIRT1 impairs glucose sensing in beta-cells in mice[48]. **c-e)** Tier 2, secondary candidate motifs for disruption have lower normalised motif scores than Tier 1 matches. **c)** Two additional motif families are implicated TEAD and SMAD. TEAD shares partial consensus with NFAT (TTCCA) and is alternative candidate to NFAT. TEAD1 but not TEAD2 is expressed in beta-cells **e)** and TEAD1 plays a critical role in pancreatic progenitors[22], however it should be noted that TEAD1 does not bind the critical region in pancreatic progenitors when the region is bound by FOXA2 (**Figure 3b**). The SMAD family are signal transducers of TGF-beta signalling and play an important role in beta-cell development, function, and proliferation[49]. For all boxplots, in panels (b, d, e) central lines correspond to the median, boxes span from the first to the third quartiles, and whiskers extend to the furthest data point within 1.5xIQR from the boxes.

References

1. Pullen, T.J. *et al.* Identification of genes selectively disallowed in the pancreatic islet. *Islets* **2**, 89-95 (2010).
2. Spielmann, M. & Mundlos, S. Looking beyond the genes: the role of non-coding variants in human disease. *Hum Mol Genet* **25**, R157-R165 (2016).
3. Ferreira, C.R. The burden of rare diseases. *Am J Med Genet A* **179**, 885-892 (2019).
4. Online Mendelian Inheritance in Man, OMIM. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), Accessed 18/11/2021. World Wide Web URL: https://omim.org/.
5. Turro, E. *et al.* Whole-genome sequencing of patients with rare diseases in a national health system. *Nature* **583**, 96-102 (2020).
6. Shashi, V. *et al.* The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders. *Genet Med* **16**, 176-82 (2014).
7. Sawyer, S.L. *et al.* Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: time to address gaps in care. *Clin Genet* **89**, 275-84 (2016).
8. Kapoor, R.R. *et al.* Clinical and molecular characterisation of 300 patients with congenital hyperinsulinism. *Eur J Endocrinol* **168**, 557-64 (2013).
9. Snider, K.E. *et al.* Genotype and phenotype correlations in 417 children with congenital hyperinsulinism. *J Clin Endocrinol Metab* **98**, E355-63 (2013).
10. Halldorsson, B.V. *et al.* The sequences of 150,119 genomes in the UK biobank. *bioRxiv*, 2021.11.16.468246 (2021).
11. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-91 (2016).
12. Karczewski, K.J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443 (2020).
13. Robey, R.B. & Hay, N. Mitochondrial hexokinases, novel mediators of the antiapoptotic effects of growth factors and Akt. *Oncogene* **25**, 4683-96 (2006).
14. Becker, T.C., BeltrandelRio, H., Noel, R.J., Johnson, J.H. & Newgard, C.B. 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. *J Biol Chem* **269**, 21234-8 (1994).
15. Bianchi, M. & Magnani, M. Hexokinase mutations that produce nonspherocytic hemolytic anemia. *Blood Cells Mol Dis* **21**, 2-8 (1995).
16. Pasquali, L. *et al.* Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat Genet* **46**, 136-143 (2014).
17. Chiou, J. *et al.* Single-cell chromatin accessibility identifies pancreatic islet cell type- and state-specific regulatory programs of diabetes risk. *Nat Genet* **53**, 455-466 (2021).
18. Greenwald, W.W. *et al.* Pancreatic islet chromatin accessibility and conformation reveals distal enhancer networks of type 2 diabetes risk. *Nat Commun* **10**, 2078 (2019).
19. Weng, C. *et al.* Single-cell lineage analysis reveals extensive multimodal transcriptional control during directed beta-cell differentiation. *Nat Metab* **2**, 1443-1458 (2020).
20. Balboa, D. *et al.* Functional, metabolic and transcriptional maturation of human pancreatic islets derived from stem cells. *Nat Biotechnol* (2022).
21. Geusz, R.J. *et al.* Pancreatic progenitor epigenome maps prioritize type 2 diabetes risk genes with roles in development. *Elife* **10**(2021).
22. Cebola, I. *et al.* TEAD and YAP regulate the enhancer network of human embryonic pancreatic progenitors. *Nat Cell Biol* **17**, 615-626 (2015).
23. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-30 (2015).

24. Bramswig, N.C. *et al.* Epigenomic plasticity enables human pancreatic alpha to beta cell reprogramming. *J Clin Invest* **123**, 1275-84 (2013).

25. Lawlor, N. *et al.* Multiomic Profiling Identifies cis-Regulatory Networks Underlying Human Pancreatic beta Cell Identity and Function. *Cell Rep* **26**, 788-801 e6 (2019).

26. Giri, D. *et al.* Novel FOXA2 mutation causes Hyperinsulinism, Hypopituitarism with Craniofacial and Endoderm-derived organ abnormalities. *Hum Mol Genet* **26**, 4315-4326 (2017).

27. Papizan, J.B. *et al.* Nkx2.2 repressor complex regulates islet beta-cell specification and prevents beta-to-alpha-cell reprogramming. *Genes Dev* **25**, 2291-305 (2011).

28. Dhawan, S. *et al.* DNA methylation directs functional maturation of pancreatic beta cells. *J Clin Invest* **125**, 2851-60 (2015).

29. Keller, M.P. *et al.* The Transcription Factor Nfatc2 Regulates beta-Cell Proliferation and Genes Associated with Type 2 Diabetes in Mouse and Human Islets. *PLoS Genet* **12**, e1006466 (2016).

30. Fang, Z. *et al.* Single-Cell Heterogeneity Analysis and CRISPR Screen Identify Key beta-Cell-Specific Disease Genes. *Cell Rep* **26**, 3132-3144 e7 (2019).

31. Xin, Y. *et al.* Pseudotime Ordering of Single Human beta-Cells Reveals States of Insulin Production and Unfolded Protein Response. *Diabetes* **67**, 1783-1794 (2018).

32. Lemaire, K., Thorrez, L. & Schuit, F. Disallowed and Allowed Gene Expression: Two Faces of Mature Islet Beta Cells. *Annu Rev Nutr* **36**, 45-71 (2016).

33. Pinney, S.E. *et al.* Dominant form of congenital hyperinsulinism maps to HK1 region on 10q. *Horm Res Paediatr* **80**, 18-27 (2013).

34. Henquin, J.C. *et al.* Congenital hyperinsulinism caused by hexokinase I expression or glucokinase-activating mutation in a subset of beta-cells. *Diabetes* **62**, 1689-96 (2013).

35. Otonkoski, T. *et al.* Physical exercise-induced hypoglycemia caused by failed silencing of monocarboxylate transporter 1 in pancreatic beta cells. *Am J Hum Genet* **81**, 467-74 (2007).

36. Almeida, A.M. *et al.* Hypomorphic promoter mutation in PIGM causes inherited glycosylphosphatidylinositol deficiency. *Nat Med* **12**, 846-51 (2006).

37. Thornton, P.S. *et al.* Recommendations from the Pediatric Endocrine Society for Evaluation and Management of Persistent Hypoglycemia in Neonates, Infants, and Children. *J Pediatr* **167**, 238-45 (2015).

38. Ellard, S. *et al.* Improved genetic testing for monogenic diabetes using targeted next-generation sequencing. *Diabetologia* **56**, 1958-63 (2013).

39. Campbell-Thompson, M. *et al.* Network for Pancreatic Organ Donors with Diabetes (nPOD): developing a tissue biobank for type 1 diabetes. *Diabetes Metab Res Rev* **28**, 608-17 (2012).

40. Laver, T.W. *et al.* SavvyCNV: Genome-wide CNV calling from off-target reads. *PLoS Comput Biol* **18**, e1009940 (2022).

41. Han, B. *et al.* Enhanced Islet Cell Nucleomegaly Defines Diffuse Congenital Hyperinsulinism in Infancy but Not Other Forms of the Disease. *Am J Clin Pathol* **145**, 757-68 (2016).

42. Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Sci Rep* **7**, 16878 (2017).

43. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-9 (2012).

44. Abdennur, N. & Mirny, L.A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**, 311-316 (2020).

45. Owens, N.D.L. *et al.* Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development. *Cell Rep* **14**, 632-647 (2016).

46. Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **48**, D87-D92 (2020).

47. Song, J.Y. *et al.* HIC2, a new transcription activator of SIRT1. *FEBS Lett* **593**, 1763-1776 (2019).

48. Luu, L. *et al.* The loss of Sirt1 in mouse pancreatic beta cells impairs insulin secretion by disrupting glucose sensing. *Diabetologia* **56**, 2010-20 (2013).
49. Lee, J.H., Lee, J.H. & Rane, S.G. TGF-beta Signaling in Pancreatic Islet beta Cell Development and Function. *Endocrinology* **162**(2021).

Extended data Figure 1

Extended data Figure 2

a GTEx HK1 Expression

b Top 3 most expressed HK1 Isoforms

c Top Expressed HK1 isoforms

ENST00000448642

ENST00000359426

ENST00000450646

chr10:71025800-71165595

d scRNA-seq endocrine cell maturation

HK1

GCK

Maturation
GSE167880

GSE114297

S5
S6
S7
M1
M3
M6
Adult

e scRNA-seq human islets

HK1

GSE101207

Extended data Figure 4

| | Hexokinase1 | Glucagon | Insulin | Hexokinase1/Glucagon | Merge |

a) *HK1* Patient

HI Patient

Non-HI Control

b) HK1 | HK1 Insulin | HK1 Glucagon | HK1 Insulin Glucagon

c) Patient 1 | Patient 6

HK1 MFI — HK1 HI α, HK1 HI β ****

HK1 MFI — HK1 HI α, HK1 HI β ****

a Histogram of contact strengths
All Loops Human Islets

b Cumulative frequency of contact strengths
All Loops Human Islets

c Islets HiC - HK1 Locus

chr10:70257650-70257651

d Islets HiC - INS Locus

chr11:1330974-1330975

e Islets HiC - SST Locus

chr3:186536699-186536700

f Islets HiC - GCG Locus

chr2:162149392-162149393

a

snATAC-seq
Hormone High
GSE160472

snATAC-seq
Hormone Low
GSE160472

chr10:71025740-71165640

chr10:71025740-71165640

b

Pancreatic Differentiation ATAC-seq
GSE149148

chr10:71025740-71165640

c

Pancreatic Progenitor ChIP-seq
GSE149148/E-MTAB-3061

chr10:71025740-71165640

d

H3K27ac
GSE149148/GSE118588/E-MTAB-1919

chr10:70907650-71310242

H3K4me1
GSE149148/GSE118588/E-MTAB-1919

chr10:70907650-71310242

H3K27me3
GSE149148/GSE118588/E087/GSE50386

chr10:70907650-71310242

Extended data Figure 7

a GSE101207 — NFAT Family

Panels (Norm. Counts vs Islet Cell Type): NFATC1, NFATC2, NFATC3, NFATC4, NFAT5, FOXA2, NKX2-2. Cell types: Acinar, Alpha, Beta, Delta, Duct, Epsilon, Pancreatic polypeptide, Pancreatic stellate.

b Differentiation Pseudotime

Panels (Norm counts): NFATC1, NFATC2, NFATC3, NFATC4, NFAT5, FOXA2, NKX2.2. Stages: ES, DE, GT, FG, PE, EP, IB, MB.

c Panels (Norm counts vs Cell Type: Alpha, Beta): NFATC1, NFATC2, NFATC3, NFATC4, NFAT5, FOXA2, NKX2-2.

Legend — Maturation GSE167880: S5, S6, S7, M1, M3, M6; Adult GSE114297.

a Two denovo mutations in patient 11, rightmost mutation 71108660:G>A is unique to patient 11 and disrupts HIC motif

b GSE101207 HIC1 / GSE101207 HIC2

c Secondary candidate disrupted motifs (Tier 2)

d GSE101207 TEAD1 / GSE101207 TEAD2

e GSE101207 SMAD1 / GSE101207 SMAD2 / GSE101207 SMAD4 / GSE101207 SMAD5 / GSE101207 SMAD7 / GSE101207 SMAD9

Fig.1

a)

HK1

Intron 2

g.71,105,619-71,110,170del (n=1)
g.71,107,383-71,111,940del (n=1)

C T G T A T T C C A G C A C A G G C T G G T A C T C G A G A C A C A T C T G T C C T G T G T T T

g.71,108,645T>C
(n=1)

g.71,108,648C>T (n=4)
g.71,108,648C>A (n=1)
g.71,108,648C>G (n=1*)

g.71,108,647C>T
(n=2)

g.71,108,665C>G
(n=2)

g.71,108,683delTGTT
(n=1)

b)

Family: 2
g.71,105,619-
71,110,170del

N/N   N/N

N/M   N/M
2.1   2.2

Family: 3
g.71,108,645T>C

N/N   N/N

N/M ‡
3.2

N/M
3.1

Family: 10
g.71,108,648C>A

N/N   N/N

N/M   NA
10.2

N/M
10.1

Fig 2

Hexokinase1   **Insulin**   **Merge**

a)

HK1 Patient

Non-HI Control

b)

HK1 MFI in beta cells

****

Non-HI Controls   HK1 HI patients

Fig 3

a scRNA-seq expression over differentiation

b Regulatory region resides in HK1 intron

c

d Islets HiC

e Maximally Disrupted Motifs

Supplementary Information

Non-coding variants disrupting a tissue-specific regulatory element in *HK1* cause congenital hyperinsulinism

Authors

Matthew N. Wakeling[1]*, Nick D. L. Owens[1]*, Jessica R. Hopkinson[1], Matthew B. Johnson[1], Jayne A.L. Houghton[2], Antonia Dastamani[3], Christine S. Flaxman[1], Rebecca C. Wyatt[1], Thomas I. Hewat[1], Jasmin J. Hopkins[1], Thomas W. Laver[1], Rachel van Heugten[2], Michael N. Weedon[1], Elisa De Franco[1], Kashyap A. Patel[1], Sian Ellard[2], Noel G. Morgan[1], Edmund Cheesman[4], Indraneel Banerjee[5,6], Andrew T. Hattersley[1], Mark J. Dunne[6], International Congenital Hyperinsulinism Consortium[¥], Sarah J. Richardson[1] and, Sarah E. Flanagan[1]

Affiliations

1. Institute of Biomedical and Clinical Science, University of Exeter Medical School, UK
2. Exeter Genomics Laboratory, Royal Devon and Exeter NHS Foundation Trust, Exeter, UK
3. Endocrinology Department, Great Ormond Street Hospital for Children, London, UK
4. Department of Paediatric Pathology, Royal Manchester Children's Hospital, Oxford Road, Manchester, UK
5. Department of Paediatric Endocrinology, Royal Manchester Children's Hospital, Oxford Road, Manchester, UK
6. Faculty of Biology, Medicine and Health, The University of Manchester, Oxford Road, Manchester, UK

* The authors contributed equally to this work.

¥ A list of the authors and their affiliations appears at the end of the paper.

Supplementary Table 1: Summarised clinical characteristics for 162 probands diagnosed with congenital hyperinsulinism screened for variants in the *HK1* regulatory region by whole-genome sequencing (n=135) or Sanger sequencing and ddPCR (n=27). Median values (IQR) provided.

	Summary data
Female sex	80 (49.1%)
Birth weight SDS	0.10 (-1.16 – 1.14)
Age at onset (days)	56 (7 – 168)
Blood glucose at presentation (mmol/l)	1.6 (1.2 – 2.1)
Insulin at presentation (pmol/l)	48 (25 – 100)
Pancreatectomy performed	34 (20.8%)

Supplementary Table 2: Clinical characteristics of 17 individuals with a *HK1* variant. The variant positions are given according to the GRCh37/hg19 genomic coordinates. The level of mosaicism in leukocyte DNA is estimated by looking at the proportion of reads with the variant at that position.

	Patient 1	Patient 2.1	Patient 2.2	Patient 3.1	Patient 3.2	Patient 4	Patient 5	Patient 6	Patient 7
HK1 variant (inheritance)	g.71,107,383 - 71,111,940del (*de novo*)	g.71,105,619- 71,110,170del (*de novo*)	g.71,105,619- 71,110,170del (*de novo*)	g.71,108,645T>C (dominant)	g.71,108,645T>C (*de novo*)	g.71,108,647C>T (*de novo*)	g.71,108,647 C>T (*de novo*)	g.71,108,648C>T (*de novo*)	g.71,108,648 C>T (*de novo*)
Relationship to proband	-	-	Twin of patient 2.1	-	Mother of patient 3.1	-	-	-	-
Gender	Female	Male	Male	Male	Female	Male	Female	Female	Female
Ethnicity	Caucasian	NZ Māori	NZ Māori	Caucasian	Caucasian	Caucasian	Caucasian	Caucasian	Unknown
Current Age (yrs)	14	12	12	6	42	11	4	9	4
Birth weight (SDS)	1.80	1.95	0.86	0.78	-0.12	2.22	0.18	2.03	2.55
Age at onset of hyperinsulinism	Birth	Birth	Birth	22 weeks	13 weeks	Birth	Birth	2 weeks	0.42 weeks
Blood glucose at diagnosis (mmol/L) [paired insulin (pmol/L)]	1.2 (77)	1.1 (1771)	0.8 (>452)	0.8 (13)	2.3 (11)	1.4 (192)	0.8 (315)	1.7 (1240)	2.2 (151)
Current treatment (dose)	Diazoxide (6.5mg/kg/d)	Diazoxide (10.1mg/kg/ d) Cornstarch	Diazoxide (11.4mg/kg/ d) Cornstarch	Diazoxide (5mg/kg/d)	None Diazoxide (10mg/kg/d) until 7 years	Diazoxide (11.5mg/kg/d)	Diazoxide (11mg/kg/d) Lanreotide (dose unknown)	Insulin	Diazoxide (15mg/kg/d)
Pancreatic resection (age)	Yes (23 months)	Yes (2 yrs 8 months)	Yes (2 yrs 8 months)	No	No	Yes (2.1 yrs and 2.8 yrs)	No	Yes (6 months)	No
Additional Features	Active duodenitis, exocrine insufficiency	None	None	None	None	None	None	None	None

	Patient 8	Patient 9	Patient 10.1	Patient 10.2	Patient 11	Patient 12	Patient 13	Patient 14
HK1 variant (inheritance)	g.71,108,648C>T (*de novo*)	g.71,108,648C>T (*de novo*; mosaic 17%)	g.71,108,648C>A (dominant)	g.71,108,648C>A (*de novo*)	g.71,108,648 C>G and g.71,108,660 G>A (both *de novo*)	g.71,108,665C>G (*de novo*)	g.71,108,665C>G (*de novo*)	g.71,108,683delTGTT (*de novo;* mosaic 30%)
Relationship to proband	-	-	-	Father of patient 9.1	-	-	-	-
Gender	Female	Female	Female	Male	Female	Male	Male	Male
Ethnicity	Sephardic Jew	Caucasian	Caucasian	Caucasian	Black African	South Asian (Punjabi)	Ashkenazi Jew	Pacific Islander
Current Age (yrs)	9	4	9	Unknown	Deceased aged 2.5 months (sepsis)	5	18	6
Birth weight (SDS)	-1.27	0.31	1.38	Unknown	0.09	-4.16	0.10	0.43
Age at onset of hyperinsulinism	Birth	Birth	Birth	Unknown	2 weeks	12 weeks	0.28 wks (resolved) 30 wks (relapse)	Birth
Blood glucose [mmol/L] and (paired insulin [pmol/L]) at presentation	2.7 (118)	2.1 (336)	Unknown	Unknown	1.6 (236)	1.6 (60)	1.4 (112)	2.0 (163)
Current treatment (dose)	Diazoxide (2.6 mg/kg/d) Octreotide (16.2 mcg/kg/d)	Diazoxide (10.32mg/Kg/d) Octreotide LAR (20mcg once monthly)	Unknown	Unknown	-	Diazoxide (15mg/Kg/d)	Octreotide LAR	Diazoxide (13 mg/kg/d) Octreotide LAR (30mg 3 weekly)
Pancreatic surgery (age)	No	No	No	No	Yes (2 months)	No	No	No
Additional Features	Horse-shoe kidney, Asthma, Aberrant left subclavian artery-operated	None	None	None	Microcephaly, hepatoslenomegaly	None	None	None

Supplementary Table 3: Summarised clinical characteristics for 17 individuals from 14 families with *HK1* variants. Median values (IQR) provided.

	Summary data
Sex (female)	9 (53%)
Birth weight (SDS)	0.61 (0.10-1.84)
Consanguinity	3 (18%)
Age at onset	Birth (0-14 days)
Blood glucose at presentation (mmol/l)	1.6 (1.2-2.1)
Insulin during hypoglycaemia (pmol/l)	163 (95-326)
Pancreatic resection performed	5/17 (29%)

Supplementary Table 4: Primer sequences for **a)** ddPCR to confirm the deletions identified by whole-genome sequencing in patients 1, 2.1 and 2.2 **b)** ddPCR to screen for deletions in the replication cohort within the 151bp region harbouring *de novo* single nucleotide variants **c)** Sanger sequencing of the *HK1* element

a)

Primer	Sequence (5' – 3')	Region amplified (Hg19)
HK1_A_Forward	tccctgcagtaatcagaatctgg	Chr10:71105908-71106051
HK1_A_Reverse	agcgccaatttatccctccg	
HK1_B_Forward	agggtctgcgtggtctttatg	Chr10:71107602-71107732
HK1_B_Reverse	ggaaacccaaaatacttcacccc	
HK1_C_Forward	ctcttacagtggcagggacg	Chr10:71109774-71109906
HK1_C_Reverse	aagccaatcacagccctctc	
HK1_D_Forward	gtgattggtacagaaagactcagtg	Chr10:71111625-71111726
HK1_D_Reverse	cagtatgaagtagctcctcgcag	

b)

Primer	Sequence (5'-3')	Region amplified (Hg19)
GCK_exon_1_Foward	TCCACTTCAGAAGCCTACTG	Chr7:44228423-44228617
GCK_exon_1_Reverse	TCAGATTCTGAGGCTCAAAC	
HK1_E_Forward	caggctggtactcgagacac	Chr10:71108653-71108835
HK1_E_Reverse	tgcctgtgacctttcctcac	
HK1_F_Forward	aggcagagtttttgcttgcc	Chr10:71108620-71108804
HK1_F_Reverse	ccactgaagctgggagaacc	

c)

	Sequence (5'-3')
Forward	AGCCTGGGCAACAGAAAC
Reverse	GCTACAAGCTCAGCCTCTTTC

Supplementary Table 5: Table describing all public genomic datasets used in this study.

Accession	DOI	Description	Cell	Target
GSE50386	doi.org/10.1172/JCI66514	ChIP-seq histone modifications in FACS sorted alpha, beta and exocrine cells	Alpha, Beta, Exocrine	H3K27me3, H3K4me3
GSE149148	doi.org/10.7554/eLife.59067	ATAC-seq, ChIP-seq for TFs and Histone modifcations over pancreatic differentiation	ES, DE, GT, PP1, PP2	CTCF, FOXA1, FOXA2, GATA4, GATA6, H3K27ac, H3K27me3, H3K4me1, H3K4me3, HNF6, PDX1, SOX9, ATAC
GSE118588	doi.org/10.1016/j.celrep.2018.12.083	ATAC-seq, ChIP-seq for TFs and Histone modifications in EndoC-BH1 cells	EndoC-BH1	ATAC, CTCF, H3K27ac, H3K27me3, H3K36me3, H3K4me1, H3K4me3, H3K79me2,H3K9me2,H3K9me3
E-MTAB-1919	doi.org/10.1038/ng.2870	ChIP-seq for TFs and Histone modifications in islets	Islets	CTCF, FOXA2, H2AZ, H3K27ac, H3K4me1, MAFB, NKX2_2, NKX6_1, PDX1
E-MTAB-1990, E-MTAB-3061	doi.org/10.1038/ncb3160	ChIP-seq for TFs and Histone modifications in pancreatic progenitors and pancreatic and liver buds	LiverBud, PP, PancreaticBud	FOXA2, GATA6, H3K27ac, H3K4me1, HNF1B, ONECUT1, PDX1, TEAD1
GSE160472	doi.org/10.1038/s41588-021-00823-0	Single nuclei ATAC-seq in islets	Islets	ATAC
E087	doi.org/10.1038/nature14248	ChIP-seq for H3K27me3 in islets	Islets	H3K27me3
GSE101207	doi.org/10.1016/j.celrep.2019.02.043	Single cell RNA-seq in islets	Islets	
GSE167880	doi.org/10.1038/s41587-022-01219-z	Single cell RNA-seq in sc-islets (data obtained from https://singlecell.broadinstitute.org/single_cell/study/SCP1526)	sc-Islets	
GSE114297	doi.org/10.2337/db18-0365	Single cell RNA-seq in islets (data processed in above publication https://doi.org/10.1038/s41587-022-01219-z, and obtained https://singlecell.broadinstitute.org/single_cell/study/SCP1526)	Islets	
GSE143783	doi.org/10.1038/s42255-020-00314-2	Single cell RNA-seq of beta-like cell differentiation, abundances projected on pseuodotime available in ref's Supplementary Table 4	ES cell to beta-like cell differentiation	

Supplementary Note

The International Congenital Hyperinsulinism Consortium

Ivo Barić[1], Liat de Vries[2,3], Samar S. Hassan[4], Khadija Nuzhat Humayun[5], Floris Levy-Khademi[6], Catarina Limbert[7], Birgit Rami-Merhar[8], Verónica Mericq[9], Kristen Neville[10], Yasmine Ouarezki[11], Ana Tangari[12], Charles Verge[10], Esko Wiltshire[13]

1. Division for Genetics and Metabolic Diseases, Department of Pediatrics University Hospital Center Zagreb and University of Zagreb, School of Medicine, Kišpatićeva 12, 10000 Zagreb, Croatia
2. The Jesse Z and Sara Lea Shafer Institute for Endocrinology and Diabetes, National Center for Childhood Diabetes, Schneider Children's Medical Center of Israel, Petach Tikva, Israel
3. Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel
4. Department of Pediatric Endocrinology, Gaafar Ibn Auf Pediatric Tertiary Hospital, Khartoum, Sudan.
5. Department of Paediatrics and Child Health, Aga Khan University, Karachi, Pakistan
6. Division of Pediatric Endocrinology, Shaare Zedek Medical Center, Faculty of medicine, Hebrew University, Jerusalem, Israel
7. Unit for Paediatric Endocrinology and Diabetes, Centro Hospitalar e Universitário de Lisboa Central, Hospital Dona Estefania, Lisbon, Portugal
8. Medical University of Vienna, Department of Pediatric and Adolescent Medicine, Vienna, Austria
9. Department of Pediatrics, Institute of Maternal and Child Research, Clinica Las Condes, Santiago, Chile
10. Endocrinology, Sydney Children's Hospital, Randwick and School of Women's and Children's Health, UNSW, Australia
11. EPH Hassan Badi, Algiers, Algeria
12. Saredo, Sanatorio Güemes, Pediatric Endocrinology Unit, Buenos Aires, Argentina
13. Department of Paediatrics and Child Health, University of Otago Wellington, Wellington, New Zealand

Additional Supplementary Table 6

Mutation	Tier	MotifFamily	MotifName	MotifID	chrom	RefSeqMotifStart	RefSeqMotifStop	RefSeqStrand	RefSeqMaxScore	MutSeqMotifStart	MutSeqMotifStop	MutSeqStrand	MutSeqMaxScore	RefSeqNormMaxScore	AltSeqNormMaxScore	LR_RefMut	Norm_RefMut	Rank	Group
g.71108645T>C	1	NFATC	NFATC1	MA0624.1	chr10	71108642	71108651	+	7.186677348	71108642	71108651	+	3.316936422	0.660146059	0.304683569	3.869740927	0.355462489	1	3
g.71108645T>C	1	RORC	RORC	MA1151.1	chr10	71108636	71108647	-	9.40792432	71108636	71108647	-	4.572355674	0.641235997	0.311647814	4.835568646	0.329588183	2	5
g.71108645T>C	1	NFATC	NFATC2	MA0152.1	chr10	71108643	71108649	+	7.629886216	71108643	71108649	+	3.891472559	0.651763144	0.332418901	3.738413658	0.319344243	3	3
g.71108647C>T	1	NFATC	NFATC2	MA0152.1	chr10	71108643	71108649	+	7.629886216	71108643	71108649	+	0.499016756	0.651763144	0.042627206	7.13086946	0.609135938	1	3
g.71108647C>T	1	NFATC	NFATC1	MA0624.1	chr10	71108642	71108651	+	7.186677348	71108643	71108652	+	3.051542135	0.660146059	0.280305267	4.135135213	0.379840792	2	3
g.71108647C>T	1	RORC	RORC	MA1151.1	chr10	71108636	71108647	-	9.40792432	71108636	71108647	-	9.106219217	0.641235997	0.620672038	0.301705103	0.020563959	3	5
g.71108648C>A	1	NFATC	NFATC2	MA0152.1	chr10	71108643	71108649	+	7.629886216	71108643	71108649	+	0.443004479	0.651763144	0.037842503	7.186881737	0.613920641	1	3
g.71108648C>A	1	NFATC	NFATC1	MA0624.1	chr10	71108642	71108651	+	7.186677348	71108642	71108651	+	0.972975109	0.660146059	0.089374498	6.213702239	0.57077156	2	3
g.71108648C>G	1	NFATC	NFATC2	MA0152.1	chr10	71108643	71108649	+	7.629886216	71108643	71108649	+	0.971674733	0.651763144	0.083002782	6.658211483	0.568760362	1	3
g.71108648C>G	1	NFATC	NFATC1	MA0624.1	chr10	71108642	71108651	+	7.186677348	71108642	71108651	+	1.160405841	0.660146059	0.106591309	6.026271507	0.55355475	2	3
g.71108648C>T	1	NFATC	NFATC2	MA0152.1	chr10	71108643	71108649	+	7.629886216	71108643	71108649	+	0.443004479	0.651763144	0.037842503	7.186881737	0.613920641	1	3
g.71108648C>T	1	NFATC	NFATC1	MA0624.1	chr10	71108642	71108651	+	7.186677348	71108642	71108651	+	1.53264669	0.660146059	0.140784207	5.654030658	0.519361852	2	3
g.71108660G>A	1	HIC	Hic1	MA0739.1	chr10	71108655	71108663	-	9.353118115	71108655	71108663	-	2.188255253	0.662526513	0.155004685	7.164862862	0.507521829	1	2
g.71108660G>A	1	NKX2	NKX2-8	MA0673.1	chr10	71108660	71108668	+	9.023473568	71108660	71108668	+	7.964371793	0.713110232	0.629411165	1.059101776	0.083699066	2	4
g.71108665C>G	1	NKX2	NKX2-8	MA0673.1	chr10	71108660	71108668	+	9.023473568	71108660	71108668	+	4.2091994	0.713110232	0.332646085	4.814274169	0.380464146	1	4
g.71108683delTGTT	1	FOX	FOXL1	MA0033.2	chr10	71108683	71108689	-	7.755237502	71108681	71108687	-	-0.965238859	0.676405287	-0.084187321	8.720476361	0.760592607	1	1
g.71108683delTGTT	1	FOX	FOXP3	MA0850.1	chr10	71108683	71108689	-	7.997699596	71108681	71108687	-	0.186093931	0.738394962	0.017181293	7.811605665	0.721213669	2	1
g.71108683delTGTT	1	SRY	SRY	MA0084.1	chr10	71108685	71108693	-	9.488019364	71108681	71108689	-	3.410416612	0.819722913	0.294644913	6.077602753	0.525078	3	6
g.71108683delTGTT	1	FOX	FOXN3	MA1489.1	chr10	71108686	71108693	-	8.792445625	71108682	71108689	-	3.629444054	0.607040481	0.25058096	5.163001572	0.356459521	4	1
g.71108683delTGTT	1	FOX	FOXC2	MA0846.1	chr10	71108681	71108692	-	9.070042541	71108681	71108692	-	4.976485307	0.608671797	0.333961637	4.093557234	0.27471016	5	1
g.71108683delTGTT	1	FOX	Foxj2	MA0614.1	chr10	71108686	71108693	-	8.67569685	71108682	71108689	-	7.045085564	0.686568851	0.557527123	1.630611286	0.129041728	6	1
g.71108683delTGTT	1	FOX	FOXP2	MA0593.1	chr10	71108685	71108695	-	9.229159905	71108681	71108691	-	8.843592817	0.61072522	0.585210921	0.385567089	0.025514299	7	1
g.71108645T>C	2	TEAD	TEAD3	MA0808.1	chr10	71108642	71108649	+	5.892565234	71108642	71108649	+	-0.236975994	0.457793963	-0.018410688	6.129541229	0.476204651	1	6
g.71108645T>C	2	NFATC	NFATC3	MA0625.1	chr10	71108642	71108651	+	6.889002554	71108642	71108651	+	1.801491577	0.573650768	0.150011125	5.087510977	0.423639642	2	2
g.71108645T>C	2	TEAD	TEAD4	MA0809.2	chr10	71108640	71108651	+	7.967197069	71108640	71108651	+	3.186200984	0.477224423	0.190849167	4.780996084	0.286375256	3	6
g.71108645T>C	2	TEAD	TEAD1	MA0090.3	chr10	71108640	71108652	+	8.893688563	71108640	71108652	+	3.79874374	0.485229365	0.207255066	5.094944823	0.277974299	4	6
g.71108645T>C	2	RORB	RORB	MA1150.1	chr10	71108635	71108645	-	7.607795391	71108635	71108645	-	3.787056814	0.536838139	0.267230706	3.820738577	0.269607433	5	2
g.71108645T>C	2	ZBTB	ZBTB12	MA1649.1	chr10	71108642	71108652	+	7.293883365	71108645	71108655	+	4.277111354	0.481176634	0.282160536	3.016772011	0.199016097	6	8
g.71108645T>C	2	HAND1::TCF	Hand1::Tcf3	MA0092.1	chr10	71108644	71108653	-	7.421589184	71108644	71108653	-	5.642432323	0.506940079	0.385412749	1.779156861	0.12152733	7	2
g.71108645T>C	2	THAP	THAP1	MA0597.1	chr10	71108644	71108652	+	5.910267741	71108644	71108652	+	5.064782004	0.468835231	0.401766634	0.845485444	0.067068597	8	2
g.71108645T>C	2	PPARG::RXRA	Pparg::Rxra	MA0065.2	chr10	71108636	71108650	-	10.10624135	71108636	71108650	-	12.02854321	0.476271879	0.566863255	-1.922301863	-0.090591377	9	2
g.71108645T>C	2	ZBTB	ZBTB26	MA1579.1	chr10	71108643	71108657	+	8.702236569	71108643	71108657	+	10.47782568	0.521930769	0.62842461	-1.775589109	-0.10649384	10	8
g.71108647C>T	2	TEAD	TEAD3	MA0808.1	chr10	71108642	71108649	+	5.892565234	71108640	71108647	-	-0.267466622	0.457793963	-0.020779508	6.160031857	0.47857347	1	6
g.71108647C>T	2	HAND1::TCF	Hand1::Tcf3	MA0092.1	chr10	71108644	71108653	-	7.421589184	71108644	71108653	-	0.446548603	0.506940079	0.03050201	6.975040581	0.476438069	2	2
g.71108647C>T	2	ZBTB	ZBTB26	MA1579.1	chr10	71108643	71108657	+	8.702236569	71108643	71108657	+	1.790557067	0.521930769	0.107391568	6.911679502	0.414539202	3	8
g.71108647C>T	2	ZBTB	ZBTB12	MA1649.1	chr10	71108642	71108652	+	7.293883365	71108642	71108652	+	1.683829338	0.481176634	0.111082025	5.610054027	0.370094609	4	8
g.71108647C>T	2	THAP	THAP1	MA0597.1	chr10	71108644	71108652	+	5.910267741	71108644	71108652	+	1.441873694	0.468835231	0.114377422	4.468394047	0.354457809	5	8
g.71108647C>T	2	NFATC	NFATC3	MA0625.1	chr10	71108642	71108651	+	6.889002554	71108643	71108652	+	2.832743391	0.573650768	0.235883992	4.056259164	0.337766776	6	2
g.71108647C>T	2	TEAD	TEAD1	MA0090.3	chr10	71108640	71108652	+	8.893688563	71108640	71108652	+	3.310574822	0.485229365	0.180621135	5.583113741	0.304608231	7	6
g.71108647C>T	2	TEAD	TEAD4	MA0809.2	chr10	71108640	71108651	+	7.967197069	71108640	71108651	+	2.993055066	0.477224423	0.179279986	4.974142003	0.297944437	8	6
g.71108647C>T	2	PPARG::RXRA	Pparg::Rxra	MA0065.2	chr10	71108636	71108650	-	10.10624135	71108636	71108650	-	6.932020955	0.476271879	0.326681951	3.174220393	0.149589927	9	2
g.71108648C>A	2	TEAD	TEAD3	MA0808.1	chr10	71108642	71108649	+	5.892565234	71108642	71108649	+	-1.294050083	0.457793963	-0.100534876	7.186615317	0.558328849	1	6
g.71108648C>A	2	NFATC	NFATC3	MA0625.1	chr10	71108642	71108651	+	6.889002554	71108642	71108651	+	0.609904446	0.573650768	0.050787055	6.279098108	0.522863712	2	2
g.71108648C>A	2	HAND1::TCF	Hand1::Tcf3	MA0092.1	chr10	71108644	71108653	-	7.421589184	71108644	71108653	-	0.234707447	0.506940079	0.016031959	7.186881737	0.49090812	3	2
g.71108648C>A	2	THAP	THAP1	MA0597.1	chr10	71108644	71108652	+	5.910267741	71108644	71108652	+	0.078658719	0.468835231	0.006239646	5.831609023	0.462595585	4	8
g.71108648C>A	2	ZBTB	ZBTB26	MA1579.1	chr10	71108643	71108657	+	8.702236569	71108643	71108657	+	2.015151874	0.521930769	0.120862006	6.687084695	0.401068763	5	8
g.71108648C>A	2	TEAD	TEAD4	MA0809.2	chr10	71108640	71108651	+	7.967197069	71108640	71108651	+	2.260363096	0.477224423	0.135392719	5.706833973	0.341831704	6	6
g.71108648C>A	2	TEAD	TEAD1	MA0090.3	chr10	71108640	71108652	+	8.893688563	71108640	71108652	+	2.761957571	0.485229365	0.150689212	6.131730992	0.334540154	7	6
g.71108648C>A	2	ZBTB	ZBTB12	MA1649.1	chr10	71108642	71108652	-	7.293883365	71108645	71108655	+	2.462548459	0.481176634	0.162454034	4.831334906	0.318722599	8	8
g.71108648C>A	2	PPARG::RXRA	Pparg::Rxra	MA0065.2	chr10	71108636	71108650	-	10.10624135	71108636	71108650	-	7.473540869	0.476271879	0.352201894	2.632700479	0.124069984	9	2
g.71108648C>G	2	TEAD	TEAD3	MA0808.1	chr10	71108642	71108649	+	5.892565234	71108642	71108649	+	-0.765379829	0.457793963	-0.059462433	6.657945063	0.517256396	1	6
g.71108648C>G	2	HAND1::TCF	Hand1::Tcf3	MA0092.1	chr10	71108644	71108653	-	7.421589184	71108644	71108653	-	0.763377701	0.506940079	0.052143381	6.658211483	0.454796698	2	2
g.71108648C>G	2	NFATC	NFATC3	MA0625.1	chr10	71108642	71108651	+	6.889002554	71108642	71108651	+	2.013043818	0.573650768	0.167627189	4.875958736	0.406023579	3	2
g.71108648C>G	2	TEAD	TEAD4	MA0809.2	chr10	71108640	71108651	+	7.967197069	71108640	71108651	+	1.711841726	0.477224423	0.102537024	6.255355343	0.374687399	4	6
g.71108648C>G	2	ZBTB	ZBTB26	MA1579.1	chr10	71108643	71108657	+	8.702236569	71108643	71108657	+	3.004424841	0.521930769	0.180195259	5.697811728	0.341735511	5	8
g.71108648C>G	2	TEAD	TEAD1	MA0090.3	chr10	71108640	71108652	+	8.893688563	71108640	71108652	+	2.907382165	0.485229365	0.158623409	5.986306398	0.326605956	6	6
g.71108648C>G	2	THAP	THAP1	MA0597.1	chr10	71108644	71108652	+	5.910267741	71108645	71108653	-	2.604807294	0.468835231	0.206627767	3.305460448	0.262207463	7	8
g.71108648C>G	2	ZBTB	ZBTB12	MA1649.1	chr10	71108642	71108652	-	7.293883365	71108645	71108655	+	3.410978424	0.481176634	0.225021848	3.882904941	0.256154785	8	8

g.71108648C>G	2	PPARG::RXRA	Pparg::Rxra	MA0065.2	chr10	71108636	71108650 -	10.10624135	71108636	71108650 -	6.531204812	0.476271879	0.307792885	3.575036536	0.168478993	9	2
g.71108648C>T	2	TEAD	TEAD3	MA0808.1	chr10	71108642	71108649 +	5.892565234	71108642	71108649 +	-1.267391988	0.457793963	-0.098463806	7.159957222	0.556257769	1	6
g.71108648C>T	2	NFATC	NFATC3	MA0625.1	chr10	71108642	71108651 +	6.889002554	71108642	71108651 +	0.76190754	0.573650768	0.063444431	6.127095015	0.510206337	2	2
g.71108648C>T	2	THAP	THAP1	MA0597.1	chr10	71108644	71108652 +	5.910267741	71108645	71108653 -	0.874367256	0.468835231	0.069359662	5.035900485	0.399475568	3	8
g.71108648C>T	2	HAND1::TCF	Hand1::Tcf3	MA0092.1	chr10	71108644	71108653 -	7.421589184	71108644	71108653 -	1.903751901	0.506940079	0.130037936	5.517837283	0.376902143	4	2
g.71108648C>T	2	TEAD	TEAD4	MA0809.2	chr10	71108640	71108651 +	7.967197069	71108640	71108651 +	2.379287026	0.477224423	0.142516103	5.587910043	0.33470832	5	6
g.71108648C>T	2	TEAD	TEAD1	MA0090.3	chr10	71108640	71108652 +	8.893688563	71108640	71108652 +	3.527983561	0.485229365	0.192482704	5.365705002	0.292746661	6	6
g.71108648C>T	2	ZBTB	ZBTB26	MA1579.1	chr10	71108643	71108657 +	8.702236569	71108643	71108657 +	4.538330406	0.521930769	0.272193736	4.163906163	0.249737034	7	8
g.71108648C>T	2	PPARG::RXRA	Pparg::Rxra	MA0065.2	chr10	71108636	71108650 -	10.10624135	71108636	71108650 -	9.659479804	0.476271879	0.455217566	0.446761544	0.021054312	8	2
g.71108648C>T	2	ZBTB	ZBTB12	MA1649.1	chr10	71108642	71108652 -	7.293883365	71108645	71108655 +	7.052438867	0.481176634	0.465248568	0.241444497	0.015928065	9	8
g.71108660G>A	2	HIC	HIC2	MA0738.1	chr10	71108655	71108663 -	7.538489562	71108655	71108663 -	0.362372563	0.564611257	0.027140666	7.176117	0.537470591	1	2
g.71108660G>A	2	GCM	GCM1	MA0646.1	chr10	71108653	71108663 +	10.01362339	71108653	71108663 +	2.828605975	0.540349678	0.152635691	7.185017417	0.387713987	2	2
g.71108660G>A	2	ZBTB	ZBTB12	MA1649.1	chr10	71108655	71108665 +	7.844904275	71108655	71108665 +	2.234850248	0.517527419	0.14743281	5.610054027	0.370094609	3	8
g.71108660G>A	2	RFX	RFX7	MA1554.1	chr10	71108655	71108663 +	6.34449971	71108655	71108663 +	2.974213675	0.471785466	0.221166498	3.370286035	0.250618967	4	2
g.71108660G>A	2	GRHL	GRHL2	MA1105.2	chr10	71108649	71108660 +	7.485658916	71108649	71108660 +	6.718678513	0.4830694	0.433574122	0.766980402	0.049495277	5	2
g.71108665C>G	2	SMAD	Smad4	MA1153.1	chr10	71108664	71108671 -	7.416467828	71108664	71108671 -	0.758256345	0.579046827	0.059201488	6.658211483	0.519845339	1	4
g.71108665C>G	2	ZBTB	ZBTB12	MA1649.1	chr10	71108655	71108665 +	7.844904275	71108655	71108665 +	7.288752952	0.517527419	0.480838181	0.556151323	0.036689238	2	8
g.71108683delTGTT	2	FOX	FOXE1	MA1487.1	chr10	71108681	71108695 -	11.00199637	71108677	71108691 -	-3.438784089	0.547155627	-0.171018968	14.44078046	0.718174594	1	1
g.71108683delTGTT	2	FOX	FOXD2	MA0847.2	chr10	71108682	71108694 -	8.639455652	71108682	71108694 -	-1.565103484	0.523145815	-0.094771867	10.20455914	0.617917682	2	1
g.71108683delTGTT	2	FOX	FOXO3	MA0157.2	chr10	71108686	71108693 -	7.06031742	71108682	71108689 -	3.408296242	0.515952404	0.249070762	3.652021178	0.266881642	3	1
g.71108683delTGTT	2	FOX	Foxj3	MA0851.1	chr10	71108682	71108698 -	10.89036379	71108678	71108694 -	6.882545243	0.563091393	0.355865246	4.007818551	0.207226147	4	1
g.71108683delTGTT	2	FOX	Foxo1	MA0480.1	chr10	71108684	71108694 +	9.315149763	71108680	71108690 +	7.310433194	0.575081785	0.451318237	2.004716569	0.123763548	5	1
g.71108683delTGTT	2	FOX	FOXD1	MA0031.1	chr10	71108686	71108693 -	6.044961173	71108682	71108689 -	4.863138044	0.468936089	0.377256506	1.181823129	0.091679582	6	1
g.71108683delTGTT	2	FOX	FOXF2	MA0030.1	chr10	71108685	71108698 -	8.302393399	71108681	71108694 -	6.901157377	0.46265136	0.3845674	1.401236022	0.07808396	7	1
g.71108683delTGTT	2	FOX	FOXG1	MA0613.1	chr10	71108686	71108693 -	6.591210371	71108682	71108689 -	5.598801489	0.502853073	0.427140749	0.992408883	0.075712324	8	1
g.71108683delTGTT	2	SOX	SOX15	MA1152.1	chr10	71108683	71108692 +	6.394447021	71108679	71108688 +	6.065508391	0.520519541	0.493743342	0.32893863	0.026776199	9	5
g.71108683delTGTT	2	SOX	SOX4	MA0867.2	chr10	71108678	71108687 -	6.595966693	71108678	71108687 -	6.595966693	0.454834365	0.454834365	0	0	10	5
g.71108683delTGTT	2	TFEB	TFEB	MA0692.1	chr10	71108677	71108686 -	7.071691337	71108677	71108686 -	7.071691337	0.450623038	0.450623038	0	0	11	7
g.71108683delTGTT	2	FOX	FOXK1	MA0852.2	chr10	71108683	71108696 -	8.354652379	71108679	71108692 -	8.442728327	0.508765953	0.514129437	-0.088075948	-0.005363484	12	1
g.71108683delTGTT	2	FOX	FOXA3	MA1683.1	chr10	71108685	71108695 -	5.923626713	71108681	71108691 -	6.039445286	0.454255838	0.463137435	-0.115818573	-0.008881597	13	1
g.71108683delTGTT	2	SOX	SOX10	MA0442.2	chr10	71108678	71108688 -	7.917174035	71108678	71108688 -	8.141207772	0.579218969	0.595609235	-0.224033738	-0.016390266	14	5
g.71108683delTGTT	2	FOX	FOXK2	MA1103.2	chr10	71108685	71108695 -	6.59500357	71108681	71108691 -	6.816649008	0.492286513	0.50883132	-0.221645438	-0.016544807	15	1
g.71108683delTGTT	2	FOX	FOXP1	MA0481.3	chr10	71108685	71108695 -	6.325144162	71108681	71108691 -	6.632937747	0.491439886	0.515354288	-0.307793584	-0.023914403	16	1
g.71108683delTGTT	2	SOX	Sox3	MA0514.1	chr10	71108679	71108688 +	7.926432641	71108679	71108688 +	8.723976331	0.530837282	0.584249195	-0.79754369	-0.053411912	17	5
g.71108683delTGTT	2	RUNX	RUNX1	MA0002.2	chr10	71108678	71108688 +	8.352026303	71108678	71108688 +	9.619327382	0.54505785	0.627762618	-1.267301079	-0.082704768	18	3