

## **A novel disease mechanism leading to the expression of a disallowed gene in the pancreatic beta-cell identified by non-coding, regulatory mutations controlling *HK1***

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

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

This research was funded in whole, or in part, by Wellcome [105636/Z/14/Z]. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author accepted Manuscript version arising from this submission

### **Abstract**

Gene expression is tightly regulated with many genes exhibiting cell-specific silencing when their protein product would disrupt normal cellular function. This silencing is largely controlled by non-coding elements and their disruption might cause human disease. 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* mutations affecting a 42bp conserved region encompassed by a regulatory element in intron 2 of *Hexokinase 1 (HK1)*, a pancreatic beta-cell disallowed gene. We demonstrated that these mutations resulted in expression of *HK1* in the pancreatic beta-cells causing inappropriate insulin secretion and congenital hyperinsulinism. These mutations identify a regulatory region critical for cell-specific silencing. Importantly, this has revealed a new disease mechanism for non-coding mutations that cause inappropriate expression of a disallowed gene.

### **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](mailto:S.Flanagan@exeter.ac.uk)

Genetic discovery in Mendelian disease has focused on identifying highly penetrant mutations 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<sup>1-4</sup>, of which many have significant clinical and genetic heterogeneity<sup>5,6</sup>. 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<sup>7,8</sup>.

To identify new aetiologies, we performed whole-genome sequencing (WGS) on 135 individuals with biochemically confirmed, persistent CHI without an identified mutation in a known disease causing gene (**Supplemental Table 1**). We initially searched for genes containing novel 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 (**Supplemental Figure 1**). Digital droplet PCR confirmed the deletions in both probands and an affected twin, in keeping with a germline mosaic mutation.

We next searched for *de novo* single nucleotide variants and indels within the minimal deleted region within our WGS discovery cohort and in a replication cohort of 27 individuals with CHI who had undergone pancreatectomy as part of routine clinical care (**Supplemental Table 1**). This identified 7 different *de novo* mutations in 12 probands. In two cases the mutation had been inherited by similarly affected offspring (**Figure 1**). All mutations were novel<sup>9,10</sup> and within a 42bp conserved region constrained against variation in gnomAD.<sup>11</sup>

Overall, the finding of 9 different *de novo* mutations 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 mutation had severe early-onset CHI (median age at diagnosis: Birth [IQR: 0-14 days]). These mutations were not associated with macrosomia at birth (median birthweight z score: 0.61 [IQR: 0.10-1.84]) suggesting that insulin secretion was not markedly increased *in utero*. In all cases the hyperinsulinism persisted, with the eldest individual still known to be affected being 18 years of age. These findings suggest that the mutations act to disrupt glucose-induced insulin secretion throughout post-natal life and into adulthood but have less impact during fetal development.

The *HK1* mutations 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* mutations, the commonest cause of CHI (**Supplemental Figure 2**). An overview of the clinical features is provided in **Supplemental Table 2**.

*HK1* encodes a glycolytic enzyme that is silenced in the pancreas and liver but expressed in all other mature tissues (**Supplemental Figure 3a-c**) where it supports glucose metabolism, ensuring cell survival<sup>12</sup>. Within the pancreatic beta-cell, glucose is phosphorylated to glucose-6-phosphate by Glucokinase (GCK; Hexokinase 4) which, due to a low-binding affinity ( $K_m \sim 8\text{mM}$ ) acts as the pancreatic glucose-sensor coupling insulin release to the prevailing glucose concentration<sup>13</sup>. Hexokinase 1 has a markedly higher affinity for glucose ( $K_m < 50\mu\text{M}$ )<sup>13</sup>. 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* mutations have been reported to cause non-spherocytic haemolytic anaemia<sup>14</sup>, dominant or recessive coding mutations 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 mutations, we found that HK1 was expressed and co-localised with insulin in the islets of affected tissue but not in unaffected controls (**Figure 2**). HK1 did not co-localise with glucagon (secreted by pancreatic alpha-cells) suggesting that the impact of the mutations was beta-cell specific (**Supplemental Figure 4**). These results confirmed that the mutations 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 an islet enhancer encompassing our critical region<sup>15</sup> 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<sup>16</sup> 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<sup>17</sup> confirmed that both our region and the relevant *HK1* promoter are contained within a well-insulated domain, without evidence of contact to distal loci (**Figure 3d**) 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<sup>18,19</sup> (**Figure 3a, Supplemental 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 that it becomes de-repressed in other pancreatic cell-types (**Supplemental 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<sup>20,21</sup> (**Figure 3c, Supplemental Figure 5b-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 is not required for HK1 repression during early differentiation.

Analysis of histone modifications during pancreatic differentiation, in human islets<sup>20</sup> and the EndoC- $\beta$ H1 beta-cell line<sup>22</sup>, 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 polycomb H3K27me3 domain (**Supplemental Figure 5d**).

Analysis of transcription factor motif families revealed that mutations disrupt transcription factor binding sites for FOX, NKX2 and NFAT families (**Supplemental Table 4, Figure 3c**), which each have family members expressed in beta-cells (**Supplemental Figure 6**). 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<sup>23</sup>, and NKX2-2 is part of a large repressor complex that regulates beta-cell specification<sup>24</sup> and could therefore have a role in maintaining the repressive epigenetic state at the HK1 locus. Multiple members of the NFAT family are expressed in beta-cells (**Supplemental Figure 6**), with NFATC2 inducing beta-cell proliferation in human islets<sup>25</sup>. Whilst NFATC2 is expressed at low levels in adult beta-cells it has greater expression in pancreatic progenitors (**Supplemental Figure 6**), supporting HK1 repression as mediated by loss of NFAT binding in late stages of beta-cell development and proliferation. Analysis of snATAC-data from human islets revealed endocrine cells present in hormone-high and hormone-low states<sup>16</sup>, 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 only remains accessible in beta-cells in the hormone low state (**Supplemental Figure 5a**). Therefore, accessibility may be maintained at the critical region during cell proliferation, by factors including NFATC2, thereby ensuring HK1 is repressed in beta-cells.

Discovering non-coding regulatory mutations in *HK1* enabled us to identify the first 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 mutations and the locus is decorated by both active and repressive epigenetic marks suggesting that continual transcription factor binding is necessary to maintain 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<sup>26,27</sup>. Linkage analysis of *HK1* to CHI<sup>28</sup> and increased HK1 expression<sup>29</sup> have been reported in several patients, but disease causality and genetic aetiology were not established. Novel 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<sup>30</sup>.

The identification of non-coding *HK1* mutations in patients with CHI represents one of the first discoveries of regulatory non-coding mutations affecting a gene in which coding mutations do not cause the same phenotype<sup>31</sup>. These findings highlight a role for

undiscovered regulatory mutations causing disease through inappropriate expression of a normally functioning protein in a specific cell-type.

Mutations 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 mutations as they establish a critical role for disallowed genes in Mendelian disease.

### **Acknowledgements**

SEF has a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (105636/Z/14/Z). MaNW and MBJ are the recipients of an Independent Fellowship and TWL and NO a Lectureship from the Exeter Diabetes Centre of Excellence funded by Research England's Expanding Excellence in England (E3) fund. EDF is the recipient of a Diabetes UK RD Lawrence fellowship (19/0005971). KAP has a Career Development fellowship funded by the Wellcome Trust (219606/Z/19/Z). IB and MJD were supported by the Northern Congenital Hyperinsulinism (NORCHI) charitable fund, by the Manchester Academic Health Sciences Centre and by The University of Manchester MRC Confidence in Concept (CiC) Award (MC\_PC\_18056). SJR and NGM are grateful to JDRF for funding (2-SRA-2018-474-S-B). Ivo Barić is member of the European Reference Network for Rare Hereditary Metabolic Disorders (MetabERN), Project ID No 739543. This research was performed with the support of the Network for Pancreatic Organ donors with Diabetes (nPOD; RRID: SCR\_014641), a collaborative type 1 diabetes research project supported by JDRF (nPOD: 5-SRA-2018-557-Q-R) and The Leona M. & Harry B. Helmsley Charitable Trust (Grant#2018PG-T1D053, G-2108-04793). The content and views expressed are the responsibility of the authors and do not necessarily reflect the official view of nPOD. Organ Procurement Organizations (OPO) partnering with nPOD to provide research resources are listed at <http://www.jdrfnpod.org/for-partners/npod-partners/>. This research has been conducted using the UK Biobank Resource. This work was carried out under UK Biobank project number 9055 and 9072. Targeted next-generation sequencing to exclude known disease-causing mutations was funded by Congenital Hyperinsulinism International (a 501(c)3 organisation) for 20 individuals within the discovery cohort.

### **Competing Financial Interests**

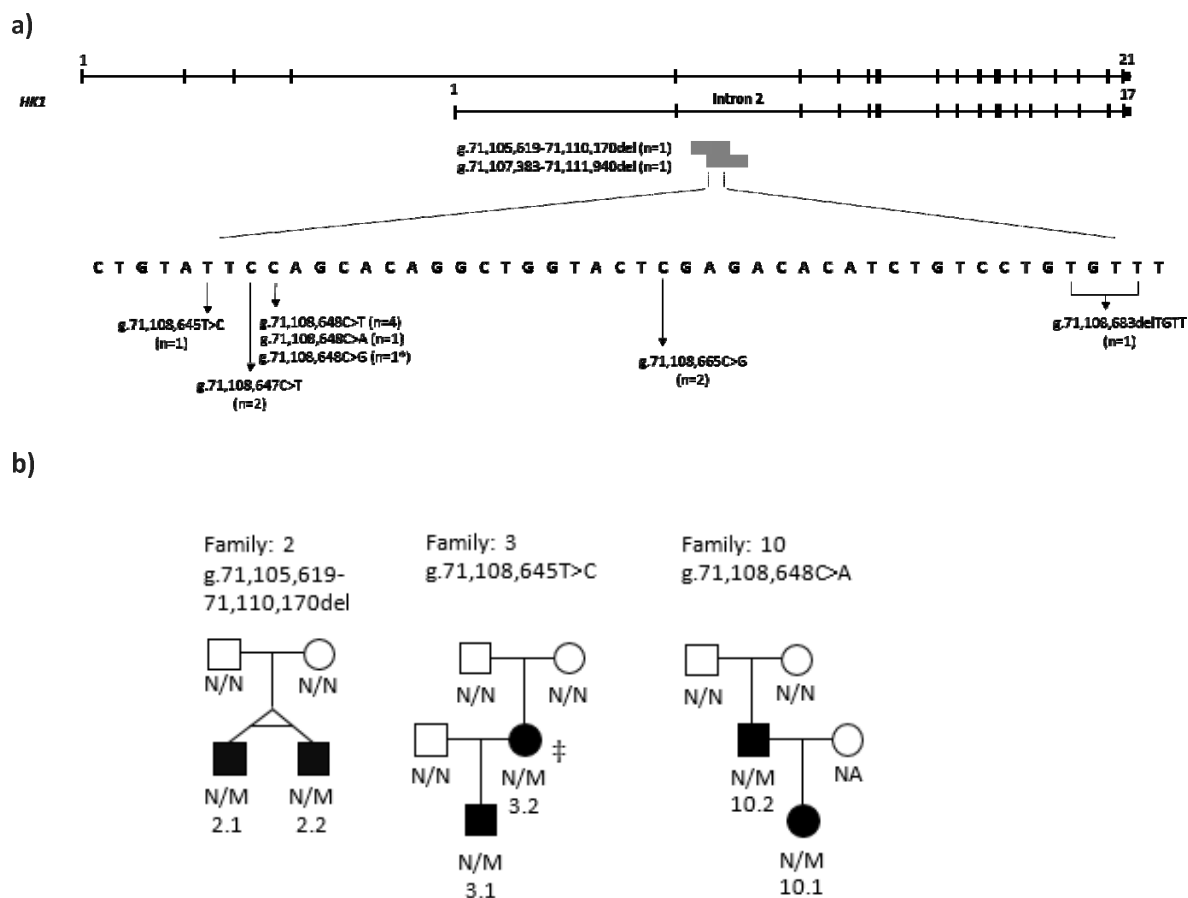
The authors declare no competing financial interests.

### **Author Contributions**

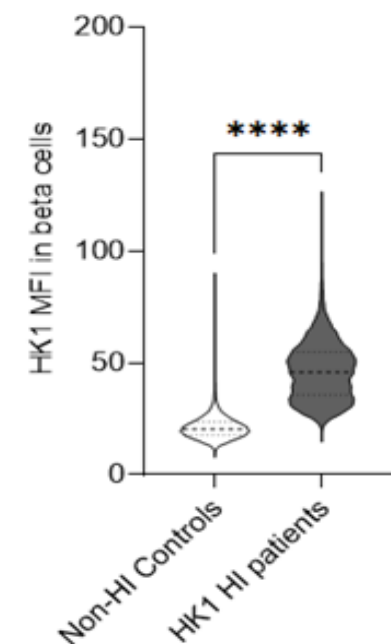
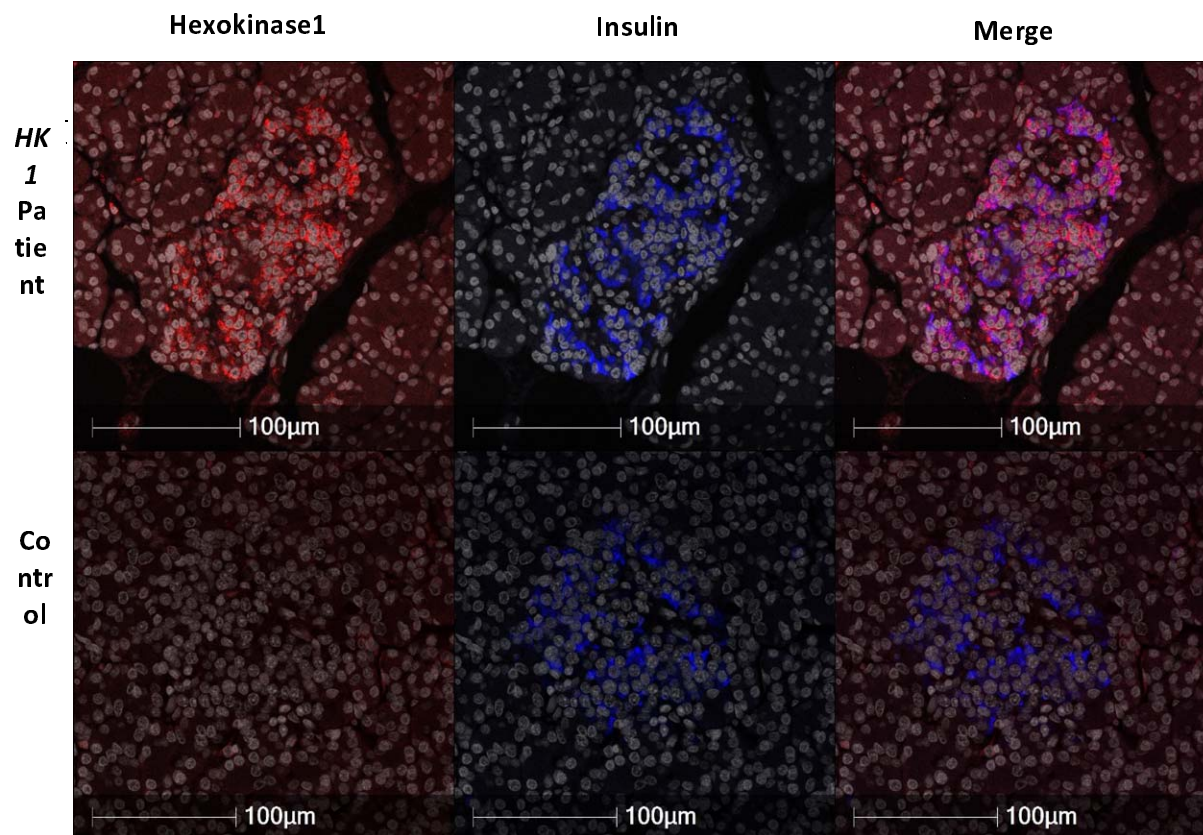
S.E.F. designed the study. A.D, S.E., I.B., S.E.F. and the International Congenital Hyperinsulinism Consortium recruited patients to the study. Ma.N.W\*, M.B.J., J.A.L.H., R.V-H., S.E. and S.E.F. performed the molecular genetic analysis and the interpretation of resulting data. Ma.N.W. and Mi.N.W. performed 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.

## Figures

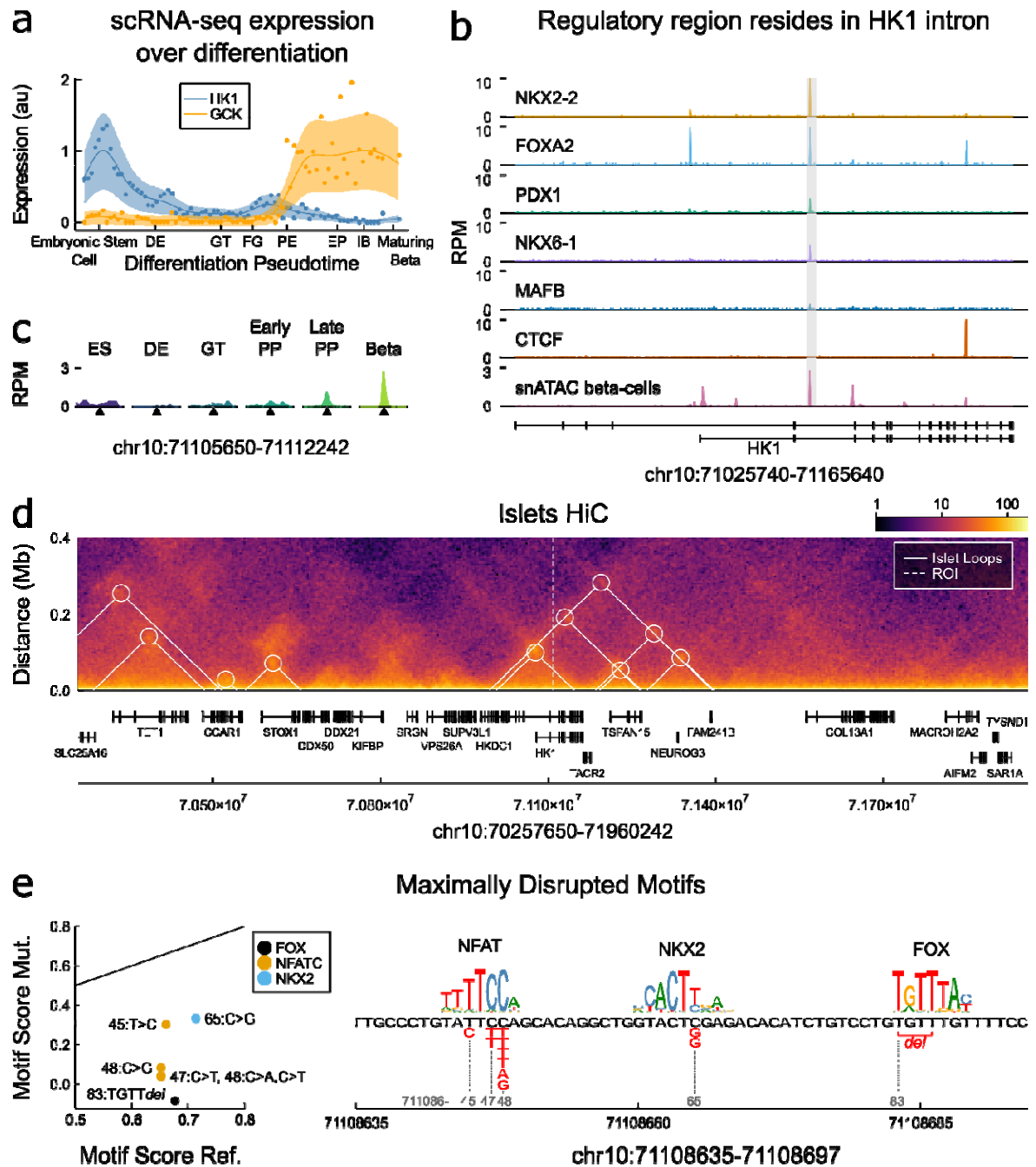


**Figure 1 a)** Schematic representation of the *HK1* gene (GRCh37/hg19 chr10:71,029,756-71,161,637). The full-length testis-specific isoform (ENST00000448642) and the shorter ubiquitously-expressed isoform (ENST00000359426) are depicted. The positions of the two ~4.5Kb deletions within intron 2 of the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



**Figure 2: HK1 expression is present in the beta-cells of donors with *HK1*-CHI, but not controls.** a) Staining of HK1 (red); insulin (blue) and DAPI (white) in pancreatic tissue resected from one patient with a *HK1* mutation. Within this section HK1co-localises with insulin (purple staining) whereas, no HK1 expression is observed in the non-hyperinsulinism (Non-HI) control donor. Scale bar – 100µm. b) HK1 expression (Median fluorescence intensity (MFI)) is significantly increased in the *HK1* HI donors (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).





**Figure 3:** 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<sup>18</sup>, Gaussian process regression median and 95% CI shown. DE – definitive endoderm, GT – gut tube, FG – foregut, PE – pancreatic endoderm, EP – endocrine precursors, IB – immature beta-cells. See **Supplemental Figure 3** for further HK1 expression. b) Transcription factor binding in human islets<sup>15</sup> and chromatin accessibility from snATAC-seq in the cluster of cells assigned to beta-cells<sup>16</sup> over *HK1* locus, the region containing the mutations is marked as grey box. Chromatin accessibility in other endocrine

cell types is provided in **Supplemental Figure 5.** **c)** 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. Beta – gives accessibility in snATAC-seq<sup>16</sup> (as shown in B). **d)** HiC data in human islets<sup>17</sup> reveals *HK1* is contained within a well-insulated domain, heatmap gives log scale contact frequencies, white triangle and circles mark chromatin loops called in same study. **e)** Transcription factor motif families disrupted by the mutations, shown are those motif families maximally disrupted by each mutation (red font). Notable secondary motif families given in **Supplemental Figure 7 and Supplemental Table 4.** Left shows scatter 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 **Supplemental Figure 6.**

## ONLINE METHODS

### Subjects

Congenital hyperinsulinism (CHI) was defined as an inappropriately high level of plasma insulin at the time of hypoglycaemia associated with inappropriately suppressed 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)<sup>32</sup>. Clinical details of the subjects are provided in **Supplemental Table 1**. Subjects with CHI were recruited by their clinicians for molecular genetic analysis to The Exeter Genomics Laboratory. Disease-causing mutations in the known CHI genes had been excluded by targeted-next generation sequencing in all cases<sup>33</sup>.

### Ethical considerations

This study was approved by the North Wales Research Ethics Committee (517/WA/0327) 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)<sup>34</sup>.

### Whole-genome sequencing

Whole-genome sequencing of DNA extracted from 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 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<sup>35</sup> using a bin size of 2Kbp. We also developed a new tool (FindLargeInsertSizes, available at <https://github.com/rdemolgen/SavvySuite>) to detect small ( $\geq 1$ Kbp) 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 mutation was identified, Sanger sequencing (for point mutations/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 (**Supplemental Table 3**) 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.

### **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* (NM\_033497) was amplified by PCR (primer sequences are listed in **Supplemental Table 3**). 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 mutation 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* mutations age-matched controls (n=4) and age-matched tissues from patients with *ABCC8* mutations (n=2). Immunohistochemistry was performed as described previously on 5-mm-thick sections of tissue<sup>36</sup>. For high-content quantification of tissue, sections were first digitized using the 3DHitech Panoramic 250 Flash II slide scanner (3DHISTECH, Budapest, Hungary) and quantified using QuPath<sup>37</sup>.

### **Immunofluorescence**

After dewaxing and rehydration, pancreatic samples from two *HK1* patients and 4 age-matched control tissues (patients with *ABCC8* mutations (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); mouse monoclonal anti-glucagon (Abcam ab10988; 1/2000 for 1h) and guinea-pig anti-insulin (Agilent; C#IR002; 1/5 for 1h). The relevant antigen-antibody complexes were detected using secondary antibodies conjugated with fluorescent dyes (Alexa Fluor™ anti-mouse 555, anti-rabbit 488, anti-guinea pig 647) (Invitrogen, Paisley, U.K). Cell nuclei were stained with DAPI. After mounting, the sections were imaged via a Leica DMI8 confocal microscope (Leica Microsystems UK, Milton Keynes, UK) 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. 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

### Public Bulk ChIP and ATAC-seq

For ChIP-seq and ATAC-seq datasets reads were downloaded from accessions provided in (**Supplemental Table 4**), and then aligned with Bowtie2<sup>38</sup> to the GRC37/hg19 genome with Bowtie2 v2.3.5.1 with default parameters for single end reads and with additional options “-l 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. All data is visualised as reads per million using <https://github.com/owensnick/GenomeFragments.jl>. For human islet single-nuclei ATAC-seq data<sup>16</sup> (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](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)<sup>17</sup>. 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<sup>39</sup> at 5kb resolution using KR balance and log-transformed for visualisation.

### scRNA-seq Expression Datasets

scRNA-seq data collected over a time course of pancreatic differentiation<sup>18</sup> projected onto a differentiation pseudotime was obtained from the reference’s supplementary table. We identify consistent temporal trends using Gaussian Process (GP) regression, following the approach we have previously applied<sup>40</sup>. 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<sup>41</sup> accession GSE101207 we gene counts per cell for the size healthy donors and normalised by depth per cell. In **Supplemental Figures 3, 6 and 7** 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<sup>19</sup>. In **Supplemental Figures 3, 6 and 7** we show mean normalised counts for each gene over annotated cell types.

### Motif analysis

To assess motifs maximally disrupted by the mutations, we took all motifs in the non-redundant JASPAR database<sup>42</sup> and found the maximal scoring match that spanned the position of each mutation 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 mutations. We considered two tiers; Tier 1 having normalised motif score  $\geq 0.6$ , and Tier 2 having  $0.45 \leq$  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  $\rightarrow$  NKX2, HIC2  $\rightarrow$  HIC) and all FOX factors we grouped in the FOX family. Tier 1 disrupted motif families include NFAT, NKX2 and FOX (**Figure 3e**). One has two *de novo* mutations, the leftmost is shared with other patients and is a candidate for NFAT disruption, and the rightmost interrupts a HIC family motif (**Supplemental Figure 7a**). HIC family members include HIC1 and HIC2, with HIC2 the most greatly expressed in beta-cells (**Supplemental Figure 7b**). Interestingly, HIC2 is a transactivator of SIRT1<sup>43</sup> and the loss of SIRT1 impairs glucose sensing in beta-cells in mice<sup>44</sup>. Tier 2 disrupted motif families include TEAD and SMAD (**Supplemental Figure 7c**). The TEAD family motif shares TTCA consensus with the NFAT family and is an alternative candidate to NFAT, TEAD1 is expressed in beta-cells (**Supplemental Figure 7d**) and plays a critical role in pancreatic progenitors<sup>21</sup>, 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 (**Supplemental Figure 7e**), SMAD factors are signal transducers of TGF-beta signalling and play an important role in beta-cell development, function, and proliferation<sup>45</sup>.

## References

1. Ferreira, C.R. The burden of rare diseases. *Am J Med Genet A* **179**, 885-892 (2019).
2. 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/>.
3. Turro, E. *et al.* Whole-genome sequencing of patients with rare diseases in a national health system. *Nature* **583**, 96-102 (2020).
4. Spielmann, M. & Mundlos, S. Looking beyond the genes: the role of non-coding variants in human disease. *Hum Mol Genet* **25**, R157-R165 (2016).
5. 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).
6. 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).
7. Kapoor, R.R. *et al.* Clinical and molecular characterisation of 300 patients with congenital hyperinsulinism. *Eur J Endocrinol* **168**, 557-64 (2013).
8. Snider, K.E. *et al.* Genotype and phenotype correlations in 417 children with congenital hyperinsulinism. *J Clin Endocrinol Metab* **98**, E355-63 (2013).
9. Halldorsson, B.V. *et al.* The sequences of 150,119 genomes in the UK biobank. *bioRxiv*, 2021.11.16.468246 (2021).
10. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-91 (2016).
11. Karczewski, K.J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443 (2020).
12. Robey, R.B. & Hay, N. Mitochondrial hexokinases, novel mediators of the antiapoptotic effects of growth factors and Akt. *Oncogene* **25**, 4683-96 (2006).
13. 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).
14. Bianchi, M. & Magnani, M. Hexokinase mutations that produce nonspherocytic hemolytic anemia. *Blood Cells Mol Dis* **21**, 2-8 (1995).
15. Pasquali, L. *et al.* Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat Genet* **46**, 136-143 (2014).
16. 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).
17. 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).
18. Weng, C. *et al.* Single-cell lineage analysis reveals extensive multimodal transcriptional control during directed beta-cell differentiation. *Nat Metab* **2**, 1443-1458 (2020).
19. Balboa, D. *et al.* Functional, metabolic and transcriptional maturation of stem cell derived beta cells. *bioRxiv*, 2021.03.31.437748 (2021).
20. Geusz, R.J. *et al.* Pancreatic progenitor epigenome maps prioritize type 2 diabetes risk genes with roles in development. *Elife* **10**(2021).
21. Cebola, I. *et al.* TEAD and YAP regulate the enhancer network of human embryonic pancreatic progenitors. *Nat Cell Biol* **17**, 615-626 (2015).
22. Dhawan, S. *et al.* DNA methylation directs functional maturation of pancreatic beta cells. *J Clin Invest* **125**, 2851-60 (2015).

23. Giri, D. *et al.* Novel FOXA2 mutation causes Hyperinsulinism, Hypopituitarism with Craniofacial and Endoderm-derived organ abnormalities. *Hum Mol Genet* **26**, 4315-4326 (2017).
24. 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).
25. 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).
26. Pullen, T.J. *et al.* Identification of genes selectively disallowed in the pancreatic islet. *Islets* **2**, 89-95 (2010).
27. 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).
28. Pinney, S.E. *et al.* Dominant form of congenital hyperinsulinism maps to HK1 region on 10q. *Horm Res Paediatr* **80**, 18-27 (2013).
29. 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).
30. 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).
31. Almeida, A.M. *et al.* Hypomorphic promoter mutation in PIGM causes inherited glycosylphosphatidylinositol deficiency. *Nat Med* **12**, 846-51 (2006).
32. 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).
33. Ellard, S. *et al.* Improved genetic testing for monogenic diabetes using targeted next-generation sequencing. *Diabetologia* **56**, 1958-63 (2013).
34. 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).
35. Laver, T.W. *et al.* SavvyCNV: genome-wide CNV calling from off-target reads. *bioRxiv*, 617605 (2019).
36. 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).
37. Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Sci Rep* **7**, 16878 (2017).
38. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-9 (2012).
39. Abdennur, N. & Mirny, L.A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**, 311-316 (2020).
40. 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).
41. 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).
42. Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **48**, D87-D92 (2020).
43. Song, J.Y. *et al.* HIC2, a new transcription activator of SIRT1. *FEBS Lett* **593**, 1763-1776 (2019).
44. 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).
45. Lee, J.H., Lee, J.H. & Rane, S.G. TGF-beta Signaling in Pancreatic Islet beta Cell Development and Function. *Endocrinology* **162**(2021).