**Supplementary Appendix**

## Table of Contents

# Detailed Methods

## Next generation sequencing

DNA from the patients in families 2 (2.1, 2.2), 6 (6.1), and 8 (8.1) underwent commercially available exome capture (Agilent SureSelect V4) and next generation sequencing (Perkin Elmer, Branford, CT, USA). Data analysis was performed with Ingenuity Variant Analysis (Version 4.1.20160602; Qiagen, Redwood City, CA, USA). In these four individuals 715,799 variants were noted. 457,942 variants remained after quality control (call quality ≥ 20). Excluding common variants (allele frequency ≥ 1 % in either of the 1000 Genomes project (phase 3v5b), ExAc (0.3), or NHLBI ESP exomes (ESP6500SI-V2) left 74,187 variants. Of those 24 variants were within the region of linkage (i.e. rs11077126 (NC_000016.10:g.7719191) – rs117691315 (NC_000016.10:g.9925462). Applying a recessive model of inheritance showed three variants in a (compound) heterozygote state in a single gene only (i.e. a heterozygote promoter mutation c.-167G>T in *PMM2* in all patients (2.1, 2. 2, 6.1, 8.1) and another heterozygote mutation in *PMM2* in each patient (c.422G>A (2.1, 2.2, 6.1); c.255+1G>A (8.1)).

Exome sequencing was performed separately also for patients 9 and 10 after exome capture with Agilent SureSelect v4. We aligned reads to hg19 using BWA and called variants using GATK as previously described.[1] Ninety-one percent of target bases were covered at ≥20X. Excluding common variants as above we identified 10 variants in the region of linkage across the two patients. The only promoter or coding variants were c.-167G>T in PMM2 (heterozygous in both patient 9 and 10) and p.F157S (heterozygous in patient 9) and p.F207S

(heterozygous in patient 10). We performed whole genome sequencing on DNA from patient 7 at Complete Genomics as previously described.[2] (Variants were aligned to NCBI build 37. 96.2% of bases were covered at ≥20X and 97.1% of the genome was fully called. A total of 3921750 SNPs and indels were called across the genome. In the linkage region there were a total 115 variants called after filtering out any variant at ≥1% in the 1000 Genomes project or present in 69 publically available genomes from Complete Genomics that were not flagged as clinically associated in dbSNP.[2] Of these, only two variants were coding (p.R141H in *PMM2*) or occurred in the promoter region of a gene (c.-167G>T in *PMM2)*.

### Luciferase promoter studies

The LightSwitch Promoter Reporter construct (S714930, Activemotif, UK) having the promoter region of human PMM2 gene cloned 5' of the Renilla luciferase gene (RenSP) was used to assess promoter activity for different mutations. This construct was generated through cloning of the 960 bp 5' to the *PMM2* startcodon into the 3656 bp vector pLightswitch_Prom. The fragment was cloned between the *MluI* and *BglII* restriction sites. Two different mutations were generated: the patients' mutation, c.-167G>T, and the c.-163_-180del, which deletes the ZNF143 binding site, detailed previously as SBS1.[3] The mutated constructs were generated using the Quickchange II XL Site-directed mutagenesis kit (200522, Agilent Technologies, California, USA) according to the manufacturer's protocol. The following oligos were used for the mutagenesis:

5'-CTGCGTTGCACCCT**T**GGAGTTGCGGTC-3',

5'-GACCGCAACTCCA**A**GGGTGCAACGCAG-3',

5’-GGCGGGCGTGATCTTGCGGTCCAGGA-3’ and

5’-TCCTGGACCGCAAGATCACGCCCGCC-3’.

All reporter constructs were sequence verified. The empty vector (Active Motif, S790005) was used as negative control.

Two different cell lines were transfected with all reporter constructs. The human renal proximal tubular cell line RPTEC/TERT1 (LGC standards, ATCC-CRL-4031, UK) was transfected in Lipofectamine® 2000 (Invitrogen, Thermo Fisher Scientific), 24 hours after seeding (30,000 cells/well in 96 well plates) with 200 ng of each experimental construct (wild type or mutated promoter or empty vector). The human pancreatic Beta-cell line 1.1B4 (Sigma Aldrich/ECACC, 10012801-1VL, UK) was transfected in Lipofectamine® 2000, 24 hours after seeding (7,500 cells/well in 96 well plates) with 50 ng luciferase construct. After 24 hours, cell lysates were prepared for the Renilla luciferase component of the LightSwitch™ Dual Assay Kit (DA010, Active Motif, UK) according to the manufacturer’s protocol and the luminescence was measured on a multimode DTX880 plate reader (Beckman Coulter, UK) for 2s. Each transfection experiment and luciferase assay was done in triplicate. The empty reporter vector was used as negative control and luminescence readouts were plotted relative to the negative control.

### Assessment of allele specific expression of PMM2 in primary renal cells

*Primary renal cells*

Primary renal cell cultures were obtained from the renal cortex of the nephrectomy specimen from patient 6.1, compound heterozygous for the *PMM2* promoter mutation (c.-167G>T) and a missense mutation (c.422G>A). Cortical tissue was separated from medulla and minced into explants <1 $mm^3$. Explants

were plated into pre-wetted flasks (Corning) with multiple explants in each flask and cultured in a minimal volume of 4.5 g/L D-Glucose DMEM GlutaMax™-I (Life technologies, 31966-047) supplemented with 10% fetal calf serum (Sera Laboratories International Ltd), 1% antibiotic-antimycotic (Sigma,A5955-100ML) at 37°C, 5% $CO_2$. Explants were left undisturbed for at least 1 week to maximise attachment and the initiation of cellular outgrowth. Thereafter the medium was changed every 4-5 days taking care not to dislodge the explants from the substrate. At confluence cellular outgrowth and explants were trypsinised (Trypsin/EDTA Life Technologies 253200-062) explants sedimented by gravity and cell suspension plated in new flasks, cells were cultured under standard conditions (37°C, 5% $CO_2$) with the medium changed (every 4-5 days). Cells were passaged at confluence and split 1:3. Cells were further maintained in 4.5 g/L D-Glucose DMEM GlutaMax™-I (Life Technologies, 31966-047) with 10% fetal bovine serum (Life Technologies, 10270-106), 2.5 µg/mL Amphotericin B (PAA laboratories, P11-001) and penicillin/streptomycin (1000 U/mL) (Life Technologies, 15140-122) at 37°C, 5% $CO_2$ and grown until about 80% confluence with a regular medium change every three to four days.

*RNA isolation*

RNA was isolated from primary renal cell cultures at different passages and days (passage 3, 2 extractions; passage 4, 1 extraction). RNA Isolation was performed using TRIzol® (Life Technologies, 15596018), according to the manufacturer's protocol for adherent cells. RNA was dissolved in UltraPure DEPC-treated Water (Sigma, D5758-100ML). The concentration of RNA was determined using the Nanodrop ND1000 spectrophotometer (Thermo Fisher)

and concentrations were in the range of 150 – 200 ng/µL per sample. RNA was stored at -80°C.

*cDNA synthesis*

Each of the three RNA isolates was transcribed into cDNA by a one-step reverse transcription reaction using the iScript™ Reverse Transcription Supermix for RT-qPCR Kit (Bio-Rad, 1708840). A total amount of 1000 ng per RNA sample was set into the cDNA synthesis step. As suggested by the manufacturers' protocol a no-reverse-transcriptase control for each RNA template and a no-template control were used as negative controls. cDNA samples were stored at -20°C.

### Allelic discrimination of PMM2 in patient derived cells

Absolute quantification of allelic discrimination was implemented and analysed using the chip-based QuantStudio™ 3D Digital PCR system (Thermo Fisher) and the QuantStudio™ 3D AnalysisSuite Cloud Software (Thermo Fisher) respectively. A custom allele specific assay, containing a primer/probe mix was designed using the TaqMan® Assay Design Tool (custom TaqMan® SNP Genotyping Assay 40x, Thermo Fisher, 4331349, Assay-ID: AHGJ9ED).

We used the missense mutation c.422G>A to discriminate the two patient alleles, as c.422G is linked with the mutant and c.422A with the wild type promoter.

The probes were differently labelled with a fluorescent dye (FAM™ or VIC®) for discrimination of the expression levels of each *PMM2* allele. Both probes were designed to span the nucleotide position 422 (Ref.-Seq.: NM_000303.2) of *PMM2* within the cDNA of patient and control. Therefore, the FAM™-labelled probe (FAM™-5'-CCAAGAAGAAC**G**CATTGA-3'-MGB) was used to detect the

allele with the mutant promoter and the VIC®-labelled probe (VIC®-5'-CCAAGAAGAAC**A**CATTGA-3'-MGB) was specific for the allele with the wildtype promoter. For amplification, the assay also consisted of the following forward and reverse primers: 5'-GTGTCCCCTATTGGAAGAAGCT-3' and 5'-CTTTTGTCTTATATTTTCTTTTTTATCGAGTTCGTAGA-3', respectively.

*Control constructs (PMM2 wt/mut)*

As positive controls, two linearized plasmid constructs (pcDNA™3.1(-)_PMM2_wt.dna; pcDNA™3.1(-)_PMM2_mut.dna) were used in a 1:1 ratio, mimicking the two *PMM2* alleles of the patient with respect to the c.422G>A mutation. For construct design, the mammalian expression vector pcDNA™3.1(-) (5.4 kb; Invitrogen, V79520) was used and linearised with BAM HI and Xho I for ligation of the insert. The insert for both, wild type and mutant construct, was generated via PCR resulting in a 349 bp fragment of the *PMM2* gene. The primers used, also introduced the restriction sites for BAM HI and Xho I on both ends of the amplicon to make ligation in correct orientation possible. Thereafter, the constructs were transformed into *E.coli* for recombination. After selection (ampicillin resistance) of positive clones the desired plasmid containing the *PMM2* wt or mutant fragment was purified via plasmid preparation and linearized through *BAMHI* and *XhoI* digestion.

### Digital PCR sample preparation

Twenty-five nanogram per patient cDNA sample (based on previous RNA determination) and a total of 200 fg of the two constructs (control) in a 1:1 ratio (100 fg each) were used to ensure an optimal concentration of approximately

200 to 2000 copies/µL of target sequence in the final reaction. In addition, three different negative controls were measured: A no RT patient control (pool of all three RNA isolates set into RT PCR, Bio-Rad, 1708840) with a total RNA amount of 25 ng, a no template control (Bio-Rad, 1708840) and the two constructs individually (200 fg each).

All samples were made to a volume of 6 uL using nuclease free water (Sigma, W1754-1VL). Each (n=3) cDNA sample was measured once and for the control, the 1:1 construct mix was prepared three times independently (n=3) and each was assayed once. For the final reactions, master mix, custom TaqMan® SNP Genotyping Assay (1:2 dilution), nuclease-free water and sample were combined and a total volume of 14.5 µL was loaded on each chip according to the manufacturer's recommendations (Thermo Fisher, MAN0007720). All dPCR reactions were carried out using the GeneAmp PCR system 9700 (Thermo Fisher) and samples were amplified with the following conditions: Initial denaturation at 96°C for 10 mins followed by 39 cycles of 60°C for 2 mins; 98°C for 30 sec and 60°C of 2 mins for extension (see also Thermo Fisher, MAN0007720).

### Data analysis from digital PCR

All chips (20,000 partitions of 755 pL) were read on the QuantStudio™ 3D Digital PCR Instrument (chip reader). The data were then imported into a web-based QuantStudio™ 3D AnalysisSuite™ Cloud Software for further analysis.

Quality threshold was determined by measuring the amount of ROX-signal of the master mix. For each chip, the quality threshold was set to a default value of 0.5 and the target number of copies/µL did never exceed 2,000. The clusters in the scatterplot of the AnalysisSuite software were examined to judge both, adequacy of separation and outliers, and adjusted manually. Only chips that met

the aforementioned criteria were incorporated in the pooled statistical analysis that was subsequently done using Microsoft Office Excel. To assess significance of expression levels of the *PMM2* alleles in the patient a two-tailed and paired t-test was done. For comparison of patient versus control a two-tailed and unpaired t-test was done for each allele separately.

## Electrophoretic Mobility Shift Assay (EMSA)

*Oligonucleotides*

All oligonucleotides were obtained from Integrated DNA Technologies; two 29-mer 5'-biotinalyted double stranded oligonucleotide probe containing either the wild type or mutant PMM2 promoter sequence c.-150_-178:

5'-GATCTGCGTTGCACCCT**G**GGAGTTGCGGT-3' (wild type) and

5'-GATCTGCGTTGCACCCT**T**GGAGTTGCGGT-3' (mutant).

The same two sequences were also purchased unlabeled for the competition experiments.

*ZNF143 expression*

ZNF143 expression was performed using the TNT T7 Quick Coupled Transcription/Translation System (Promega, L1170) and confirmed by Western analysis using an antibody (Origene, 4C5 anti DDK-Ab, TA50011) that detects the DDK tag. Expression of this transcription factor was achieved using a PrecisionShuttle pCMV6-Entry vector consisting of a T7 promotor, c terminal (myc and DDK) tag and the open reading frame (ORF) for *ZNF143* (Origene, RC205850). The same vector without the ORF for *ZNF143* was used as a negative control.

*EMSA*

EMSA was performed essentially as described by Garner and Revzin,[4] but with slight modifications; binding reactions and detection of the 5'-biotinylated double stranded oligonucleotide probes corresponding to the c.-156_-184 fragment of the *PMM2* and TMEM186 bidirectional promotor containing the predicted binding site for ZNF143 were performed using the LightShift Chemiluminescent EMSA Kit (Thermoscientific, 20148) as described by the manufacturer's instructions.

Briefly, binding reactions were performed using 1 X binding buffer (10 mM Tris, 50 mM KCl, 1 mM DTT; pH 7.5), 2.5 % glycerol, 5 mM $MgCl_2$, 50 ng/µL of poly(dI-dC), 0.05% NP40, 2 µL of the respective TnT reaction (vector only or vector containing the ORF of ZNF143) reaction and 2 pmol of biotinylated probes in a total volume of 20 µL for 20 minutes at room temperature. Competition were performed by adding 200-fold excess of unlabeled probes to the binding reaction prior to addition of expressed transcription factor. Five microliter of loading dye was added to each of the samples and all were subjected to gel electrophoresis at 4$^{o}$C using a pre-electrophorese 4 % native polyacrylamide gel in 0.5 X TBE buffer at a constant 100 V. DNA and protein were then electroblotted onto a pre-soaked Biodyne B Nylon Membrane in 0.5 % TBE. Transfer was performed at 4$^{o}$C in the Xcell SureLock Mini-Cell system at a constant 380 mA for 30 minutes using pre-cooled 0.5 % TBE. Once transferred, the DNA was crosslinked to the membrane by exposing it to a transilluminator for 15 minutes. The membrane was immunoblotted with a 1:300 dilution of streptavidin-horseradish peroxidase conjugate antibody in block buffer. Detection of the biotin-labelled DNA was

achieved by chemiluminescence. Blots were imaged using the Syngene Dyversity imaging system and images were analyzed using the Syngene software. EMSA was performed in 5 independent experiments.

*Densitometry*

ImageJ version 1.42q was used to select and determine the mean grey value in a given area of the bands of interest and the same area representing the background noise. The value obtained from the background noise was used to subtract the value obtained from each of the bands. Results from 5 independent experiments were assessed. A two-tailed and unpaired t-Test was performed on these results.

## Insulin secretion studies

Insulin-secreting MIN6 cells[5] were kindly provided by Prof J I Miyazaki (University of Tokyo, Japan) and maintained at 37°C (95% air/ 5% $CO_2$) in DMEM supplemented with 10% fetal calf serum (FCS), 2mM L-glutamine, 100 U/ml penicillin and 0.1 mg/ml streptomycin. For measurements of insulin secretion MIN6 cells were seeded into 96-well plates at a density of 30,000 cells/well and maintained as adherent monolayers under standard tissue culture conditions. To deglycosylate cell surface proteins the cells were incubated (37°C, 120 min) in a physiological salt solution[6] supplemented with 2mM glucose, 2mM $CaCl_2$ and 0.5 mg/ml BSA in the absence or presence of Peptide N-Glycosidase F (PNGase F, 4000U/ml, New England Biolabs) and Endoglycosidase H (Endo H, 4000U/ml, New England Biolabs). Immediately following deglycosylation the cells were washed and incubated (37°C, 60 min, 200µl/well) in the physiological salt solution supplemented with 2mM glucose, or

20mM glucose, or 20mM glucose plus the protein kinase C activator, phorbol-12 myristate-13 acetate (PMA, 500nM). At the end of the incubation period supernatant samples were removed and insulin was measured by radioimmunoassay as described previously.[7] Differences in insulin secretion between treatment groups were assessed by two-way ANOVA.

### Bioinformatic analysis of the promoter mutation

The genomic region on chromosome 16 including surrounding *PMM2* was analysed with the UCSC Genome browser[8] using human genome data (HG19). Transcription factor binding was assessed with the UCSC Genome Browser transcription factor binding site track (TFBS) using ENCODE ChiP-Seq data from 67 cell lines (March 2012 Freeze). Chromatin segmentation was assessed using all available ENCODE data from human cell lines (GM12878, H1-hESC, K562, HeLa-S3, HepG2, HUVEC).

Prediction of chromatin looping in the genomic region on chromosome 16 was performed by analyzing ENCODE CHiP-Seq data for binding of CTCF and markers of Cohesin (RAD21 and SMC3) and assessing of the orientation of the CTCF sites, so that loops would be delimited by CTCF in convergent orientation.[9]

We further assessed evidence for loop formation in this region by analyzing Hi-C data from Rao et al.[10] Data from this project had been deposited in Gene Expression Omnibus (GEO) under the accession number GSE63525. Data on loops involving *PMM2* were available for 3 cell types: K562, IMR90 and GM12878. We further analyzed ENCODE ChIA-PET data from the cell lines with available CTCF data: K562 and MCF-7.

We further assessed whether chromatin formation may be different in tissues affected by HIPKD compared to those that are not. Cell lines with available CHiA-PET or Hi-C data are either derived from various forms of cancer (K562, GM12878, MCF-7) or fetal lung (IMR90) and thus may not be representative of tissues affected by HIPKD. We therefore looked at indirect evidence for altered chromatin architecture by assessing ENCODE data for accessibility to formaldehyde (FAIRE mapping), DNase I or transcription factor binding,

### Bioinformatic analysis of HNF4A binding sites

HNF4A binding sites identified from ChIP-seq (Supplemental Figure 2) were analyzed for HNF4A specificity using a positional weight matrix based on the data from Fang et al[11], as well as their web-based tool (http://nrmotif.ucr.edu/NRBSScan/H4SBM.htm).

# Supplemental Results

## Chromatin loops

Analysis of CTCF and Cohesin binding sites and CTCF orientation identified a potential loop including *PMM2* spanning approximately 220 kb (chr 16: 8754981 to 8974276), as depicted in Supplemental Figure 2. Direct evidence for the existence of this predicted loop was provided from Hi-C data from the K562 and IMR90 cell lines and ChIA-PET data from MCF-7 cells. (Supplemental Figure 2). Analysis of chromatin segmentation, FAIRE mapping, DNase I and transcription factor binding shows a distinct and unique pattern for HepG2 cells compared to other cell types (data not shown, except for chromatin segmentation in Supplemental Figure 2), suggesting most notably increased chromatin accessibility over the *ABAT* gene. HepG2 is a liver-derived cell line and thus likely representative of HIPKD affected tissue. The altered chromatin accessibility coincides with the presence of HNF4A in HepG2 cells, which binds to a number of sites in *ABAT* and expression of *ABAT* in this cell line; moreover, novel peaks of DNase I hypersensitivity in HepG2 cells predominantly occur at HNF4A binding sites, consistent with altered chromatin structure in HNF4A expressing cells, which may interfere with *PMM2* transcription from the mutant promoter.

### Specificity of HNF4A binding sites

HNF4A belongs to a nuclear receptor family including RXR and COUPTF2 that can share a classical DNA binding motif (AGGTCA). However, a similar motif (CAAAGTCCA) is highly specific for HNF4A binding.[11] We analyzed the 13 HNF4A binding sites identified in the PMM2 containing loop (Supplemental Figure 2) and noted that the binding sites at the 5' and 3' end of the loop (assessed as strong binding sites by ENCODE analysis of ChIP-seq data and detailed in Supplemental Figure 3), located in the respective *ABAT* and *CARHSP1* promoter best fit the HNF4A specific binding motif, consistent with the hypothesis of HNF4A being the key transcription factor affecting tissue-specific *PMM2* transcription.

# Supplemental Figure legends

## Supplemental Figure 1: Pedigrees of the 11 Families

Shown are the pedigrees of the 17 patients with HIPKD described here. Dark symbols denote affected, empty symbols unaffected individuals. A small black circle in an empty symbol indicates a parent in whom the carrier state for the respective mutation of the affected child has been confirmed. Note that inheritance is consistent with being autosomal recessive.

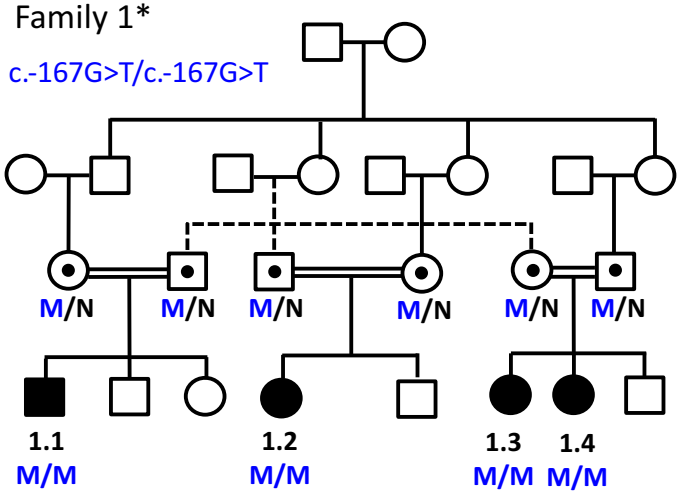
*: Families used for linkage analysis (Figure 1)

M: carrier of the promoter mutation c.-167G>T

M: carrier of a coding mutation in PMM2

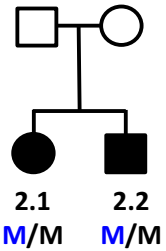
Family 1*

c.-167G>T/c.-167G>T



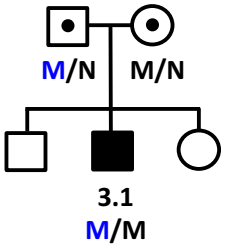

Family 2*

c.-167G>T/p.R141H



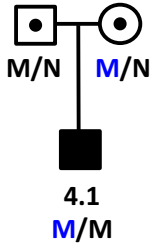

Family 3

c.-167G>T/p.F157S



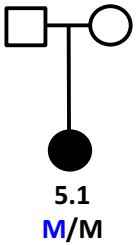

Family 4

c.-167G>T/p.R141H



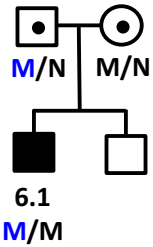

Family 5

c.-167G>T/p.R141H



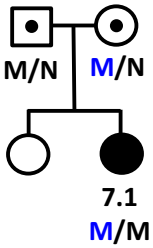

Family 6*

c.-167G>T/p.R141H



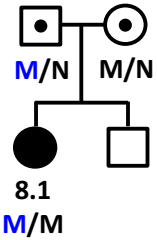

Family 7

c.-167G>T/p.R141H



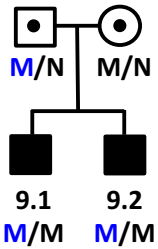

Family 8*

c.-167G>T/c.255+1G>A



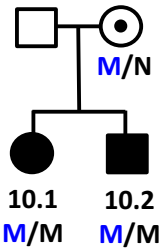

Family 9

c.-167G>T/p.F207S



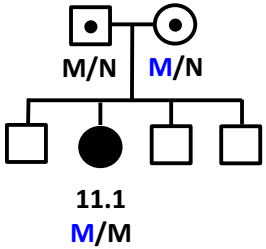

Family 10

c.-167G>T/p.F157S




Family 11*

c.-167G>T/p.F157S



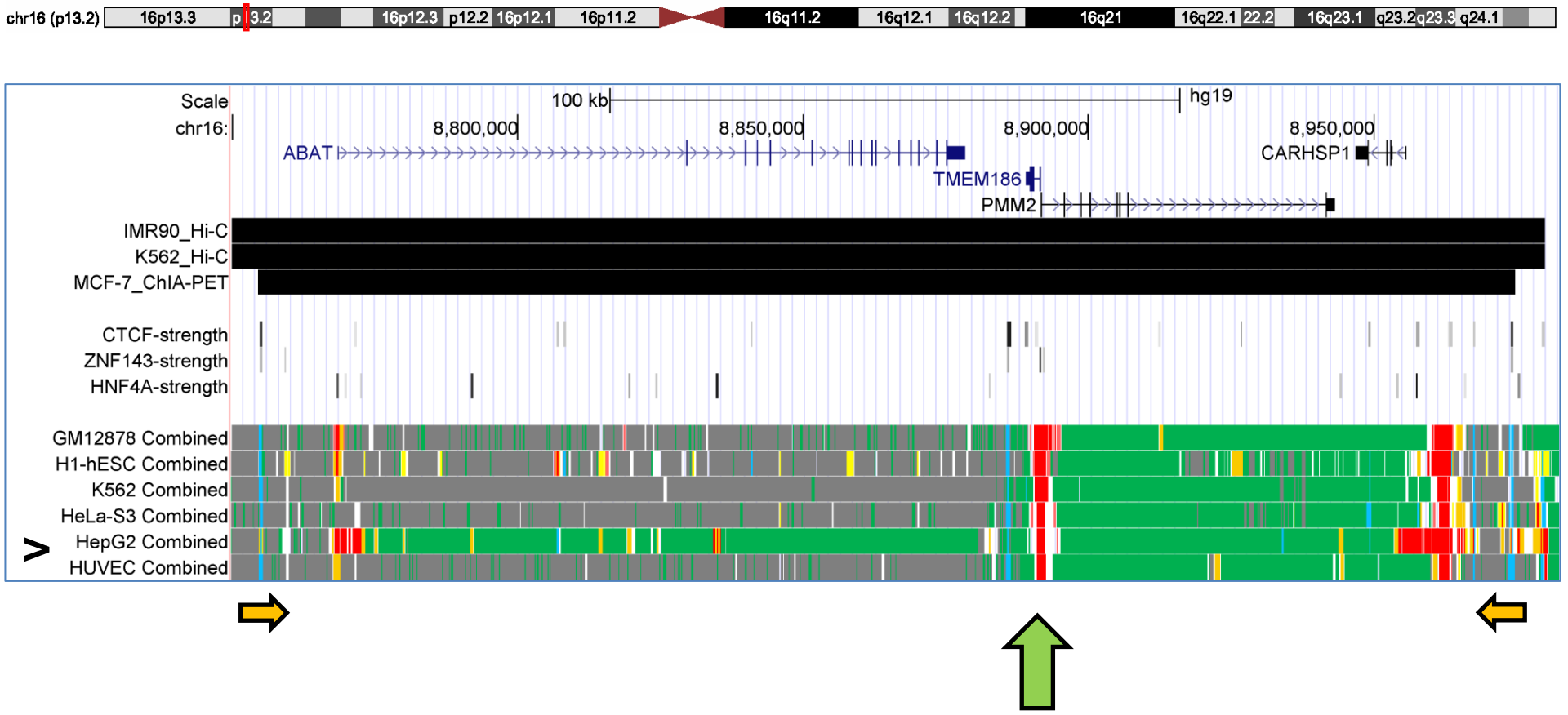
## Supplemental Figure 2: Bioinformatic analysis of the PMM2 genomic region suggests an organ-specific disease mechanism

A) *Bioinformatic analysis*. Shown is the genomic region on chromosome 16p13.2. Chromatin interaction analysis (ChIA-PET, Hi-C) identifies a loop of chromatin (size approximately 220 kb) that includes *PMM2* (HG19). The upper part of the picture shows the genes in the region. Below are three tracks indicating DNA binding sites based on ENCODE analysis of ChiP-seq data, thus reflecting transcription factor binding in experimental cells. Grayscale according to strength.

1) CTCF: Note the binding sites (yellow arrows), which are in convergent orientation, as expected for CTCF sites limiting chromatin loops.
2) ZNF143: Note the strong ZNF143 binding site immediately 5' of *PMM2* (green arrow), where the c.-167G>T mutation is located.
3) HNF4A: Note the clusters of HNF4A binding sites at the flanks of the predicted loop, the closest of which is <2 kb from the CTCF site.

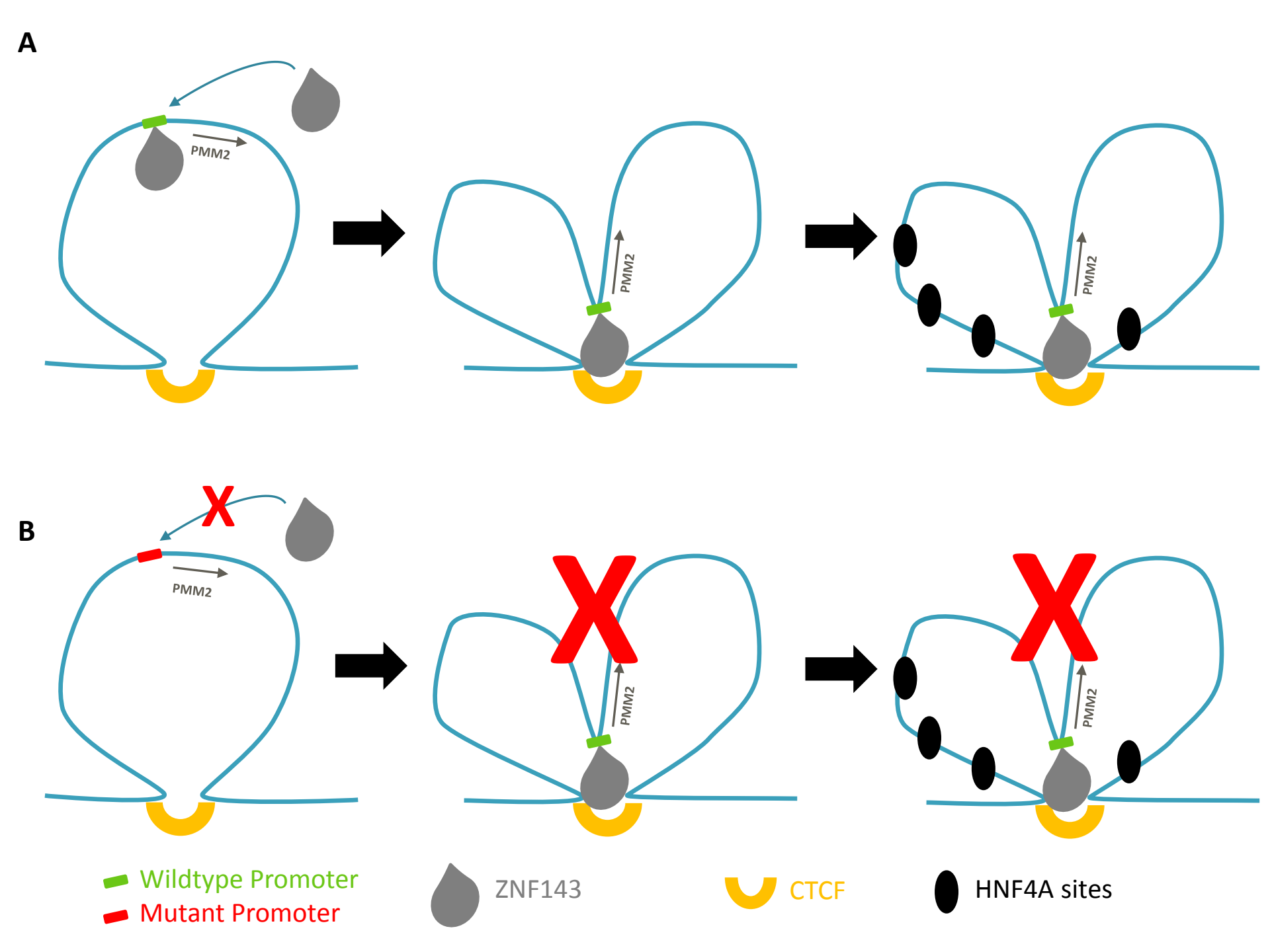
The lower 6 tracks show chromatin segmentation from the six available cell lines (ENCODE). Color-coding as follows: gray: repressed or low activity region, blue: CTCF-enriched elements (insulator regions), red: promoter region, green: transcribed region, orange: enhancer, yellow: weak enhancer.

Note that there is evidence for general expression (all 6 cell lines) on the centromeric side of the PMM2 promoter (to the right of the green arrow) and specific expression (only in the hepatic cell line, left arrowhead) on the telomeric side of the promoter (to the left of the green arrow) consistent with organ-specific chromatin regulation.


chr16 (p13.2)
16p13.3
16p12.3
p12.2
16p12.1
16p11.2
16q11.2
16q12.1
16q12.2
16q21
16q22.1
22.2
16q23.1
q23.2
q23.3
q24.1
Scale
100 kb
hg19
chr16:
8,800,000
8,850,000
8,900,000
8,950,000
ABAT
CARHSP1
TMEM186
PMM2
IMR90_Hi-C
K562_Hi-C
MCF-7_ChIA-PET
CTCF-strength
ZNF143-strength
HNF4A-strength
GM12878 Combined
H1-hESC Combined
K562 Combined
HeLa-S3 Combined
HepG2 Combined
HUVEC Combined

## Supplemental Figure 3: Cartoon of a potential disease mechanism

Simplified cartoon (not to scale) depicting the proposed disease mechanism. A) Wild type promoter: Note the establishment of a chromatin loop delimited by binding sites for CTCF (yellow). Binding of ZNF143 to the wild type *PMM2* promoter (green) and to CTCF changes the 3-dimensional structure of the loop (the functional loop) and thus can allow interaction between the *PMM2* promoter and clusters of HNF4A binding sites (black). B) Mutant promoter: there is impaired binding of ZNF143 to the mutant *PMM2* promoter (red) with altered loop formation in HNF4A expressing cells and thus impaired *PMM2* transcription. HNF4A is expressed predominantly in the organ systems involved in HIPKD: pancreatic beta cells, kidney and liver.


A
PMM2
PMM2
PMM2
B
X
PMM2
X
PMM2
X
PMM2
Wildtype Promoter
Mutant Promoter
ZNF143
CTCF
HNF4A sites

# Supplemental References


1. Flanagan, SE, Haapaniemi, E, Russell, MA, Caswell, R, Lango Allen, H, De Franco, E, McDonald, TJ, Rajala, H, Ramelius, A, Barton, J, Heiskanen, K, Heiskanen-Kosma, T, Kajosaari, M, Murphy, NP, Milenkovic, T, Seppanen, M, Lernmark, A, Mustjoki, S, Otonkoski, T, Kere, J, Morgan, NG, Ellard, S, Hattersley, AT: Activating germline mutations in STAT3 cause early-onset multi-organ autoimmune disease. *Nature genetics,* 46**:** 812-814, 2014.
2. Drmanac, R, Sparks, AB, Callow, MJ, Halpern, AL, Burns, NL, Kermani, BG, Carnevali, P, Nazarenko, I, Nilsen, GB, Yeung, G, Dahl, F, Fernandez, A, Staker, B, Pant, KP, Baccash, J, Borcherding, AP, Brownley, A, Cedeno, R, Chen, L, Chernikoff, D, Cheung, A, Chirita, R, Curson, B, Ebert, JC, Hacker, CR, Hartlage, R, Hauser, B, Huang, S, Jiang, Y, Karpinchyk, V, Koenig, M, Kong, C, Landers, T, Le, C, Liu, J, McBride, CE, Morenzoni, M, Morey, RE, Mutch, K, Perazich, H, Perry, K, Peters, BA, Peterson, J, Pethiyagoda, CL, Pothuraju, K, Richter, C, Rosenbaum, AM, Roy, S, Shafto, J, Sharanhovich, U, Shannon, KW, Sheppy, CG, Sun, M, Thakuria, JV, Tran, A, Vu, D, Zaranek, AW, Wu, X, Drmanac, S, Oliphant, AR, Banyai, WC, Martin, B, Ballinger, DG, Church, GM, Reid, CA: Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science,* 327**:** 78-81, 2010.
3. Anno, YN, Myslinski, E, Ngondo-Mbongo, RP, Krol, A, Poch, O, Lecompte, O, Carbon, P: Genome-wide evidence for an essential role of the human Staf/ZNF143 transcription factor in bidirectional transcription. *Nucleic acids research,* 39**:** 3116-3127, 2011.

4. Garner, MM, Revzin, A: A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. *Nucleic acids research,* 9**:** 3047-3060, 1981.

5. Ishihara, H, Asano, T, Tsukuda, K, Katagiri, H, Inukai, K, Anai, M, Kikuchi, M, Yazaki, Y, Miyazaki, JI, Oka, Y: Pancreatic beta cell line MIN6 exhibits characteristics of glucose metabolism and glucose-stimulated insulin secretion similar to those of normal islets. *Diabetologia,* 36**:** 1139-1145, 1993.

6. Gey, GO, Gey, MK: Maintenance of human normal cells in continuous culture: preliminary report. In Cultivation of mesoblastic tumors and normal cells and notes on methods of cultivation. *American Journal of Cancer,* 27**:** 45-76, 1936.

7. Jones, PM, Salmon, DM, Howell, SL: Protein phosphorylation in electrically permeabilized islets of Langerhans. Effects of Ca2+, cyclic AMP, a phorbol ester and noradrenaline. *The Biochemical journal,* 254**:** 397-403, 1988.

8. Kent, WJ, Sugnet, CW, Furey, TS, Roskin, KM, Pringle, TH, Zahler, AM, Haussler, D: The human genome browser at UCSC. *Genome Res,* 12**:** 996-1006, 2002.

9. Dekker, J, Mirny, L: The 3D Genome as Moderator of Chromosomal Communication. *Cell,* 164**:** 1110-1121, 2016.

10. Rao, SS, Huntley, MH, Durand, NC, Stamenova, EK, Bochkov, ID, Robinson, JT, Sanborn, AL, Machol, I, Omer, AD, Lander, ES, Aiden, EL: A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell,* 159**:** 1665-1680, 2014.

11. Fang, B, Mane-Padros, D, Bolotin, E, Jiang, T, Sladek, FM: Identification of a binding motif specific to HNF4 by comparative analysis of multiple nuclear receptors. *Nucleic acids research,* 40**:** 5343-5356, 2012.