Insights into beta cell regeneration for diabetes via integration of molecular landscapes in human insulinomas

Huan Wang1,2,14, Aaron Bender2,3, Peng Wang3, Esra Karakose3, William B. Inabnet4, Steven K. Libutti5, Andrew Arnold6, Luca Lambertini7, Michael Stang8, Herbert Chen9, Yumi Kasai10, Milind Mahajan1, Yayoi Kinoshita11, Gustavo Fernandez-Ranvier4, Thomas C. Becker12, Karen K. Takane3, Laura A. Walker3, Shira Saul3, Rong Chen1,14, Donald K. Scott3, Jorge Ferrer13, Yevgeniy Antipin1,14, Michael Donovan11, Andrew V. Uzilov1,14, Boris Reva1, Eric E. Schadt1,14, Bojan Losic1, Carmen Argmann1 & Andrew F. Stewart3

Although diabetes results in part from a deficiency of normal pancreatic beta cells, inducing human beta cells to regenerate is difficult. Reasoning that insulinomas hold the “genomic recipe” for beta cell expansion, we surveyed 38 human insulinomas to obtain insights into therapeutic pathways for beta cell regeneration. An integrative analysis of whole-exome and RNA-sequencing data was employed to extensively characterize the genomic and molecular landscape of insulinomas relative to normal beta cells. Here, we show at the pathway level that the majority of the insulinomas display mutations, copy number variants and/or dysregulation of epigenetic modifying genes, most prominently in the polycomb and trithorax families. Importantly, these processes are coupled to co-expression network modules associated with cell proliferation, revealing candidates for inducing beta cell regeneration. Validation of key computational predictions supports the concept that understanding the molecular complexity of insulinoma may be a valuable approach to diabetes drug discovery.
Normal physiologic human beta cell replication occurs only transiently in human infancy and early childhood, ceasing irreversibly thereafter. Therapeutically, there is only one class of drugs, still in early development, that reproducibly induces human beta cell replication: the harmine analogue class of small molecules that inhibit the kinase, DYRK1A. Even here, however, the replication rates induced are modest and not beta cell-specific. Accordingly, there is an urgent need to discover additional beta cell mitogenic drugs and regenerative pathways.

Insulinomas are very rare, small (~1–2 cm), slowly proliferating pancreatic beta cell adenomas. They come to medical attention through their overproduction of insulin, causing hypoglycemia, with resultant psychomotor symptoms. They are almost always benign, and are readily treated by laparoscopic removal. Since they are a rare tumor, they are not captured in large cancer genomic surveys such as The Cancer Genome Atlas (TGCA) or the International Cancer Genome Consortium (ICGC).

Here we report whole-exome sequencing (WES) and RNA sequencing (RNA-seq) of thirty-eight human insulinomas. We provide these findings for public access with extensive sets of annotations relating to the DNA variants identified, with the ability to prioritize selection of high-impact mutations in a user-defined way. Our primary intent was to employ an integrative genomics approach to identify mitogenic mechanisms with potential application for human beta cell expansion (Supplementary Fig. 1). This approach entails integrating whole-exome and RNA-sequencing data into network analysis to computationally model insulinoma molecular events relative to normal adult and juvenile human beta cells. We reasoned that although some molecular events in insulinoma are likely relevant to the mechanisms of tumor formation, some may serve to uncover the genetic mechanisms that enforce beta cell quiescence, and are bypassed in such benign tumors. We further validated combinations of lead candidate genes derived from this approach as beta cell mitogenic mediators. Notably, we focused on insulinomas from subjects not known to be members of multiple endocrine neoplasia type 1 (MEN1) kindreds, as the MEN1 gene has been previously reported as one of the most frequently mutated genes in hereditary pancreatic neuroendocrine tumors (PNETs). Although MEN1 mutations are uncommon in sporadic insulinomas. Despite attempting to exclude MEN1 subjects, we nevertheless find widespread abnormalities in genes functionally related to MEN1, revealing a previously unsuspected unifying mechanism underlying insulinoma.

Results

Insulinomas harbor recurrent mutations in epigenetic genes. WES was performed on genomic DNA from insulinomas and patient-matched blood cells (as normal controls) from 22 patients, with a mean usable sequencing depth of 79X and 105X for the blood and tumor samples, respectively. We included WES data from four additional insulinomas from a prior report, yielding a total of 26 insulinomas with paired normal and tumor DNA data (Supplementary Tables 1, Supplementary Data 1). Mutation analysis was performed as described. After manual

---

**Fig. 1** Global molecular characterization of the insulinoma genomic landscape from whole-exome sequencing data. **a** A summary of 26 insulinomas (top box) and a subset of their protein-changing key driver variants, each sample displaying the variant with the highest allelic fraction and, if any, the variant from recurrently mutated genes. All variants are somatic, except for a germline MEN1 variant from sample 5967T (at chr1:64,572,613, G>A, p.R420*, nonsense). **b** A summary of somatic copy number variants from selected model-predicted epigenetic modifiers and CDKNIC. **c** A summary showing which 85% of insulinomas harbor mutations or CNVs in selected epigenetic modifiers. **d** GO pathway analysis (Molecular Function) of insulinoma key driver variants reveals terms associated with chromatin-binding and SMAD signaling as the most prominent pathways. **e** Circos plot of copy number gain (red), loss (blue), or copy-neutral cnLOH (green). Each line within a track represents 20% of the total number of insulinomas. Note that ~20% of insulinomas have CNV loss on chromosome 11 (blue), and ~20–40% have CNV gain on chromosome 7 (red). **f** GISTIC2.0 analysis showing finer mapping of regions of significant chromosomal amplification or deletion throughout the genome. Cytoband labels indicate significant calls (FDR < 0.1)
review of all the variants, a total of 258 somatic (tumor-specific) single-nucleotide variants (SNVs) and 20 non-SNV variants, including indels and multiple-nucleotide variants (MNVs), that alter protein sequence were identified (Supplementary Data 2 and 3), revealing an average of 10.7 such somatic variants per insulinoma exome.

Notably, recurrent variants were rare in insulinomas: only four of 26 insulinomas harbored the previously reported T372R SNV in the YY1 gene. Only two tumors had MENI mutations, one somatic and one germline. We also identified several novel recurrently mutated genes in insulinomas, including H3 histone family 3A (H3F3A; two tumors), lysine-specific demethylase 6A (KDM6A; two tumors), Filamin C gamma (FLNC; two tumors), ATR serine/threonine kinase (ATR; two tumors), and Rho GTPase-activating protein 35 (ARHGAP35; two tumors). Interestingly, the seven recurrently mutated genes identified in 10 of
the 26 insulinomas (five being epigenetic regulators: YY1, MEN1, H3F3A, KDM6A, ATR) are significantly enriched for epigenetic regulators as defined by the curated database, EpiFactors13 (fold enrichment = 16.7, \( p = 2.8 \times 10^{-6} \); Fisher’s exact test). Key driver analysis14, 15 nominated 92 genes, including the seven recurrently mutated genes noted above, as potential disease-causing key driver mutations (Fig. 1a–c, Supplementary Table 2) and revealed similar enrichment for epigenetic regulators (fold enrichment = 2.8, \( p = 0.002 \); Fisher’s exact test). Furthermore, gene ontology (GO) molecular function pathway enrichment analysis of the predicted key drivers revealed similar biological processes of “chromatin-binding” (AdjP = 0.005) among the most significantly enriched terms along with and “SMAD-binding” (AdjP = 1.6 \times 10^{-6} ) (Fig. 1d, Supplementary Data 4). Thus, surprisingly, despite having largely excluded MEN1-associated insulinomas, the strongest recurrent mutational signal nonetheless arose from genes encoding epigenetic modifiers functionally related to MEN1, which also encodes an epigenetic modifying enzyme. Interestingly, “developmental pathways” were also prominent in GO biological processes, likely reflecting the major role of chromatin-modifying enzymes in development and differentiation.

Somatic copy number variants (CNVs) including gain, loss, or copy-neutral loss of heterozygosity (cnLOH) were also investigated. A Circos plot summarizing the recurrence of CNVs, identified by the saasCNV algorithm16, across all 26 insulinomas revealed that chromosome 7 and 11 had the most frequent gain-of-copy and loss-of-copy/cnLOH CNVs, respectively (Fig. 1e). GISTIC2.0 analysis17 confirmed that some of the strongest amplification and deletion signals arose from chromosomes 7 and 11, respectively, and further defined regions of significant focal CNV events throughout the genome (Fig. 1f, Supplementary Data 5 and 6). Interestingly, MEN1 resides in one of the significantly deleted regions of chromosome 11q13.1, in 8 of 26 insulinomas via loss-of-copy/cnLOH events (Fig. 1b, f), consistent with previous MEN1 studies18, 19–21. In addition, frequent loss-of-copy/cnLOH of CDKN1C, which resides on chromosome 11p15 and encodes the cell cycle inhibitor p57kip2, was also observed in 9 of 26 insulinomas (Fig. 1b), as reported previously in diabetic insulinomas18. Pursuing the “epigenetic modifier” gene ontology

---

**Fig. 3** Global epigenetic dysregulation in insulinomas revealed by comparison of ~ 14 K genes from insulinomas and beta cells to histone marks in normal pancreatic cells. a H3K4me3, H3K9ac (active) and H3K27me3 (repressive) marks in normal beta cells and other islet-cell types from three independent data sources27–29. In the bottom 12 panels are compared to DEGs (insulinomas vs. beta cells) in the bottom panel. In the top panel, the numbers in the “Cell Type” column represent the number of genes included in each histone mark signature that were also tested in the differential expression analysis from the bottom panel, and the black vertical lines in the “Insulinoma vs. Beta Cell DEG” column indicate genes with histone marks. In the bottom panel, the red box indicates genes upregulated in insulinomas and blue represents genes downregulated. The white line represents −log10(FDR) of the DE analysis. As can be seen, upregulated genes in insulinomas (red), in contrast to downregulated DEGs (blue) and non-DEGs (gray), were highly significantly over-enriched for the repressive mark, H3K27me3, and under-enriched for the active marks, H3K4me3 and H3K9ac across the three epigenetic data sources. The histone mark signature visualization curves in the top 12 panels represent the average scores for each sliding window, as described in Methods. b Projection of all 12 histone signatures onto the insulinoma co-expression network reveals that the upregulated DEG-enriched modules (red labels) were also significantly over-enriched for the repressive mark, H3K27me3, and under-enriched for the active marks, H3K4me3, H3K9ac, at FDR of 0.05, while the downregulated DEG enriched modules (blue labels) were generally not enriched for histone signatures. Significant enrichment (FDR < 0.05) is shown in red for over-enrichment or blue for under-enrichment.
signature from the earlier mutational analysis, we also surveyed CNVs harboring other epigenetic modifications in addition to MEN1. Among this class, the EZH2 locus on chromosome 7 (7q36.1) displayed a gain of copy CNV in 10 of 26 insulinomas (Fig. 1b). Thus, collectively, at the DNA level, 22 of 26 insulinomas (85%) contain recurrent SNVs/MNVs/indels (YY1, H3F3A, KDM6A, MEN1, ATR) and/or recurrent CNVs (MEN1, EZH2) in genes encoding epigenetic modifiers (Fig. 1c), four of which are members of two key epigenetic modifying complexes, the polycomb group (EZH2, YY1) and the trithorax group (MEN1, KDM6A). These complexes are recruited to specific regions of DNA and direct the post-translational modification of histones (e.g., H3F3A) to either repress (polycomb) or activate (trithorax) gene expression21, 22.

Transcriptomic networks reveal pathways underlying growth. RNAseq analysis was performed on 25 insulinoma samples, 13 of which had matching WES analysis (Supplementary Data 1 and 7). We compared the transcriptomes of the insulinomas to those of beta cells isolated using three independent methodologies from 22 normal human cadaveric islet donors23–25 (Supplementary Data 7). Unsupervised classification analysis, after adjustment for gender and library platform covariates, showed that insulinomas cluster separately from beta cells, regardless of isolation technique (Fig. 2a).

Comparison of insulinoma to beta cell transcriptomes revealed 3709 differentially expressed protein-coding genes (DEGs) (FDR < 0.01) (Fig. 2b, Supplementary Data 8) of which 2125 were upregulated and 1584 were downregulated in insulinomas. Notably, insulin was consistently the most highly expressed transcript across all samples, confirming the identity of the PNETs as insulinomas. We further inferred differential splicing in insulinomas vs. beta cells through analysis of differential exon usage/retention (Fig. 2c, Supplementary Data 9 and 10). For example, exon-level analysis identified the major histocompatibility II transcriptional regulator, CIITA (Simes Test FDR = 4.2 x 10^{-27}), which was also upregulated at the DEG level in insulinomas, as being among the top 10 most differentially spliced genes (DSGs) in insulinomas (Supplementary Data 10).

As part of a data dimensionality reduction approach to identify the key modules of genes that underlie the altered biological
processes occurring in insulinomas relative to beta cells, a signed, weighted gene co-expression network analysis (WGCNA) was performed (Supplementary Data 11, Supplementary Fig. 2). Projection of the DEGs (insulinoma vs. beta cells) onto the insulinoma-only co-expression network (52 co-expressed modules), yielded six modules that were over-enriched for upregulated genes, and two modules were significantly over-enriched for downregulated genes (fold enrichment >2, FDR < 0.01) (Fig. 2d, e, Supplementary Data 12). Two modules (red and grey60) were found to be significantly under-enriched for both upregulated and downregulated genes (fold enrichment <0.6, FDR < 0.01) (Fig. 2d, e). Pathway enrichment analysis based on GO and KEGG databases was used to annotate the biological functions of each insulinoma co-expression module (Fig. 2e, Supplementary Data 13). Key biology highlighted by the DEG-enriched modules included “immune system”, “extracellular matrix”, “vasculature development”, “cell proliferation”, “RNA splicing” and “ubiquitination”. This diverse biology indicates a complex transformation of the insulinoma, despite the conserved function of insulin secretion.

**Insulinoma transcriptomes link to epigenetic dysregulation.** Since the genomic evidence indicated a unifying theme of perturbation of epigenetic modifiers in insulinoma, including modifiers of histone methylation (Fig. 1), we looked for hints of altered chromatin states regulating the insulinoma transcriptome. We focused on histone modifications, in particular histone-H3 post-translational modifications, by testing the enrichment of DEGs (insulinoma vs. beta cell) across prior reports in which three histone mark signatures (H3K27me3, H3K4me3, H3K9Ac) had been characterized in normal human beta cells or other islet cells.27–29 (Fig. 3a, Supplementary Data 14). Remarkably, the upregulated DEGs in insulinomas were significantly over-enriched for the repressive mark (H3K27me3) in normal beta cells (Ferrer beta: fold enrichment = 3.3 and FDR 3.1 × 10−37; and Kaestner beta: fold enrichment = 9.0, FDR = 2.6 × 10−216) and under-enriched for the active mark (H3K4me3) (Kaestner beta: fold enrichment = 0.2, FDR = 4.7 × 10−222). This striking contrast between genes upregulated in insulinnomas vs. normal beta cells bearing the repressive mark was consistent across the three different data sources. This trend also extended to non-beta cells: alpha cells, exocrine cells and pancreatic progenitor cells, as well as whole islets. The reverse scenario (under-enrichment for the active histone mark, H3K4me3, in beta cells) was also evident for upregulated DEGs in insulinomas (Fig. 3a, Supplementary Data 14). Furthermore, the histone mark signature visualization scores revealed that even among the upregulated DEGs, the DEGs with more significant FDRs (the white line in the bottom track in Fig. 3a) were more likely to have a high score for H3K27me3-signatures and low score for H3K4me3 signatures (Fig. 3a). To summarize, precisely the same collection of ~2000 genes that are repressed in normal beta cells (and related pancreatic and endocrine cells) and which bear strong H3K27me3 repressive chromatin marks and weak H3K4me3 open chromatin marks in beta cells, are overexpressed in insulinomas; this is true whether genes are assessed individually (Fig. 3a) or as modules of co-expressed genes (Fig. 3b). These findings provide compelling evidence to support a functional role for the recurrent mutations in polycomb and trithorax genes in the broadly abnormal gene expression in insulinomas.

To explore further the different beta cell histone mark signatures in the context of insulinoma transcriptome, we projected the same histone mark signatures onto the insulinoma co-expression network. Importantly, the same modules over-enriched for upregulated DEGs were also significantly over-enriched for the repressive mark (H3K27me3) and under-enriched for the active marks (H3K4me3, H3K9ac) (FDR < 0.05). The downregulated DEG-enriched modules, in contrast, were generally not enriched for any specific histone signatures (Fig. 3b, Supplementary Data 15). Thus, a significant set of genes that are upregulated in insulinomas relative to beta cells appear to be de-repressed through epigenetic modifications that are recorded by trithorax and polycomb family members.

**Global abnormalities in allele-specific expression.** The epigenetic aberrations described above may have additional consequences, among which may be changes in allele-specific expression (ASE). Lacking parental genotype information, we interrogated the insulinoma and beta cell RNAseq data to identify genes subject to ASE. Across the entire insulinoma genome, we detected 717 loci (586 of which were common SNPs) showing potential allelic imbalance (FDR < 5%) (Supplementary Data 16). A Manhattan plot of these data (Fig. 4a) reveals a strong allelic bias in the insulinomas relative to beta cells, largely involving loci not previously reported in beta cells.

Although the present data set cannot dissociate these ASE results from epigenetic vs. genetic (e.g., cis-eQTL) causation, they nonetheless provide additional insights into insulinoma biology. Strikingly, ASE highlighted the well-known imprinted region of chromosome 11 as preferring mono-allelic expression (Fig. 4a). For example, 8 of 24 insulinomas revealed a significant (FDR < 0.01%) preference for the alternate “T” allele at the SNP, rs689. This SNP is located at the 5’ end of the INS-IGF2 locus, near the differentially methylated regions, DMR0 and DMR1. These DMRs assist the imprinting control region (ICR) located immediately upstream of H19 to enforce ASE, by the parent-of-origin, of maternally expressed H19 and paternally expressed IGF2 in chromosomal region 11p15.5. The rs689 allelic imbalance was also observed in 5 of 22 beta cell samples, albeit at much lower significance (FDR < 3%), consistent with a previous report.24 For further understanding the possible impact of the rs689 SNP in the pathobiology of human insulinoma, we explored the relationship between the A and T alleles and expression of transcripts in this region. There was no impact on expression of H19, IGF2, INS, but usage of the alternate T allele was associated with significantly higher expression of the INS-IGF2 anti-sense transcript, and a trend toward higher expression of the INS-IGF2 read-through transcript (Supplementary Fig. 3, Supplementary Data 17). Analysis of RNAseq profiles in insulinomas expressing the T vs. A alleles at rs689 revealed multiple DEGs, the gene ontology analysis of which suggested that usage of the T allele favors a neural signature in insulinomas (Supplementary Fig. 3).

**Altered DNA methylation patterns at 11p15 in insulinomas.** Because 11p15 is a paradigmatic imprinted region, because 11p15 imprinting abnormalities are associated with beta cell proliferation in the focal variant of congenital hyperinsulinism (FoCHI) and the Beckwith–Wiedemann Syndrome (BWS)31–34, and because insulinomas display marked reductions in the cell cycle inhibitor, p57KIP2 encoded by CDKN1C in the 11p15 imprinted region (see below), we performed deep CpG methlyome sequencing in the 11p15.5-p15.4 region on 10 insulinomas and compared this to the methylation pattern in two sets of FACS-sorted normal human beta cell preparations (Fig. 4b–h). As compared to normal beta cells, insulinomas display broad and widespread hypomethylation, together with a hypermethylation in a few selected areas. These regions contain sequences known to have important regulatory functions in imprinting regulation, possibly affecting ASE, and which may hint at underlying disordered 3-D chromatin structure in this region. Insulnoma also
KDM6A also binds to additive. In contrast, silencing p detectable in approximately half of beta cell (all induce human beta cells to enter cell cycle at rates comparable to those observed in insulinomas, and that CDKN1C in human islets by the KDM6A inhibitor, GSKJ1. Cre recombinase and adenoviruses encoding indicate mean SEM of 10 different human islet preparations. Asterisks indicate p < 0.05 by Student's t-test. The inset shows examples of Ki67 immunolabeling (red) in human beta cells (green) induced by a control adenovirus encoding Cre recombinase and adenoviruses encoding CCND1. Similar effects were seen with EZH2 or shCDKNIC (not shown). Repression of CDKNIC expression in human islets by the KDM6A inhibitor, GSKJ1. Error bars indicate mean ± SEM of five human islet prepss; asterisks indicate p < 0.05 by Student’s t-test. Repression of CDKNIC expression by adenosial silencing of KDM6A in human islets. Error bars indicate mean ± SEM of five human islet prepss; asterisks indicate p < 0.05 by Student’s t-test. Top panel: ChIP analysis of KDM6A binding to the CDKNIC locus in five pairs of FACS-sorted human beta cells and non-beta cells. Error bars indicate mean ± SEM. Bottom panel: A genome browser view of the CDKNIC locus highlighting primer pairs 1-8[8], promoters, enhancers, exons, and transcriptional start site of CDKNIC. In beta cells, KDM6A binds to CDKNIC, predominantly at an upstream enhancer delineated by primer pair #2. KDM6A also binds to CDKNIC in non-beta cells of the islet. Note that the directionality of this panel is opposite that in Fig. 4h
display CNV loss, copy-neutral LOH in the 11p15 region overlapping with key regulatory elements (e.g., transcriptional start sites, enhancers, CTCF-binding sites, ICRs, splice sites) of important transcripts in this region including INSL, INS-IGF2, IGFI, KCNQ1, KCNQ1OT1, H19, and CDKN1C. This is corroborated by marked overexpression of H19 and KCNQ1 and underexpression of CDKN1C in the large majority of insulinomas, as described in more detail below. IGFI2 was upregulated in several insulinomas, but mostly unchanged compared to beta cells (Supplementary Data 8). Notably, despite the widespread CpG abnormalities in this region, INS expression is preserved in insulinomas. The 11p15.4 region also contains the non-imprinted genes, ABCG8 and KCNJ11, which encode the sulfonylurea receptor, SUR1, and the Kir6.2 potassium inward rectifier, respectively. These were not significantly differentially expressed (Supplementary Data 8). Collectively, the widespread and recurrent structural and/or imprinting abnormalities within 11p15—already strongly associated with pathologic beta cell proliferation in FoCHI and the BWS syndrome31—suggest an important role for abnormalities in this region in the molecular pathogenesis of insulinoma. This should be more deeply explored in the future using 3-D structural approaches such as CTCF ChIPseq, 3-C and 5-C chromatin capture studies.

An integrative approach suggests proliferation drivers. Our integrative genomics approach led to the identification of ten co-expression modules (Fig. 2e) that may underlie insulinoma pathogenesis, many of which were linked to potential epigenetic dysregulation (Fig. 3b). Given our interest in beta cell regeneration, and seeking the mechanisms that may permit escape from beta cell quiescence in benign insulinomas, we elected to focus on the bisque4 co-expression module, the module in insulinomas that contained DEGs enriched for cell proliferation. Furthermore, the bisque4 cell cycle module was particularly enriched for beta-cell specific histone mark signature, including genes with bivalent marks, both H3K27me3 and H3K4me3 (Kaestner beta28; fold-enrichment = 5.2, FDR = 4.0 × 10−6) (Fig. 3a,b).

To assess the relevance of the bisque4 module to normal beta cell cell proliferation, we compared the co-expression of the genes in the bisque4 module to those within normal beta cells. Similarly as for insulinomas, WGCNA was performed for human islets from a large publicly available data set of 89 samples30, the majority of which were from healthy adults (Supplementary Fig. 4). We reasoned that an islet-based co-expression network could serve as a proxy for a normal beta cell co-expression network, given beta cells on average make up approximately 50% of the human islet cell population35, 36. Importantly, the insulinoma bisque4 cell cycle module was found to significantly overlap with that of only one islet co-expression module, the lightcyan module (fold enrichment = 23.7, FDR = 9.6 × 10−25, Fig. 5a). This suggests that co-expression of cell cycle genes in normal beta cells is preserved in insulinomas, despite their greater rate of proliferation.

Seeking to model the normal physiologic beta cell expansion seen in childhood, we further projected two independent gene expression signatures derived from proliferating juvenile human37 and mouse38 beta cells onto the insulinoma co-expression network. Among these many modules, only the bisque4 module was over-enriched (human35: fold enrichment = 12.7, FDR = 9.6 × 10−6; mouse38: fold enrichment = 13.4, FDR < 1.0 × 10−16) (Fig. 5b). Altogether, these observations suggest a commonality in the mechanisms that underlie proliferation in insulinomas compared to those in juvenile mouse and human beta cells.

Importantly, among the top 20 hub genes (based on intra-modular connectivity)26 in the bisque4 module were genes directly involved in beta cell cycle progression (Fig. 5c), including E2F1, MKi67 and CDK1. However, since many of the bisque4 module genes were found to be related to late (G2/M) cell cycle stages (Supplementary Data 11), we searched for genes farther upstream of the G0/G1 transition by expanding the bisque4 module to include genes whose expression in insulinomas was significantly correlated with the bisque4 module gene expression profiles, thereby assigning a “bisque4 module membership p-value” (MMP). At a bisque4 MMP of 0.01, an additional ~200 genes were revealed (Supplementary Data 18).

We first surveyed differentially expressed (insulinoma vs. beta cells) cyclins, cyclin-dependent kinases (CDKs) and CDK inhibitors for their MMP status (Fig. 5d). Interestingly, CDKN1C/p57KIP2 was the only CDK inhibitor with significant bisque4 MMP, nominating it as a key candidate driver of the proliferative phenotype in human insulinomas. The consistent reduction of CDKN1C relative to normal beta cells was confirmed both by DEG analysis as well as immunohistochemistry on a human insulinoma tissue microarray (Fig. 5e, f), and supports the putative imprinting abnormalities and CNV loss in the CDKN1C region on chromosome 11p15.5 (Fig. 4b). Moreover, a central role for CDKN1C/p57KIP2 is supported by its well documented loss in FoCHI, BWS, and pediatric insulinomas31, 32, 33, 34, and by the observation that lenti viral silencing of CDKN1C/p57KIP2 in transplanted human islets leads to proliferation35. Notably, among differentially expressed genes encoding G1/S phase cell cycle activators, two of the most highly expressed were CCND1 and CCND3 (Fig. 5d, e). We have reported previously that CCND1 is increased in 40% of insulinomas40, and that overexpression of either CCND1 or CCND3 leads to human beta cell proliferation41.

In addition, we explored the recurrently mutated genes, genes encoding polycomb and trithorax complex members for their bisque4 MMP and DEG status (Fig. 5d). Interestingly, among the polycomb genes, only EZH2 achieved a significant bisque4 MMP (Spearman correlation = 0.60, p = 0.002), concordant with the observation that in mouse models, Ez2h epigenetically represses CDKN2A/p16INK4A in beta cells, and is requisite for beta cell proliferation in juvenile mice42. As noted earlier, EZH2 upregulation in some insulinomas may arise from the increase in copy number at the EZH2 locus on chromosome 7q. The observations that CDKN1C (rather than CDKN2A) was the CDK inhibitor best co-correlated with bisque4, and that EZH2 can transcriptionally repress CDKN1C43 suggest that combined EZH2 and CDKN1C loss/repression may be critical to enable proliferation in insulinoma cells.

Biological validation of candidate proliferation drivers. To validate combinations of lead candidate genes derived from our integrative approach as human beta cell mitogenic mediators, from the list of ~200 bisque4-associated genes, we developed adenoviruses to overexpress EZH2 and YY1 (mutant and wild-type), to silence MEN1 or CDKN1C, alone or in various combinations, and to determine which, if any, candidates might induce replication in human beta cells (Fig. 5g). Surprisingly, neither silencing MEN1 nor overexpressing mutant or wild-type YY1-induced proliferation, perhaps reflecting a requirement for additional mitogenic events, or longer lead time to induce requisite epigenetic changes. In contrast, adenoviral overexpression of CCND1 or EZH2, or silencing of CDKN1C, alone and in several combinations, were all capable of inducing human beta cell replication, at rates mimicking the 1−2% Ki67 labeling index characteristic of benign human insulinomas44. Importantly, overexpression of EZH2 in combination with silencing CDKN1C was more effective than either experimental perturbation alone.
The EZH2-CCND1 combination yielded the highest Ki67 labeling indices of all. As expected, overexpression of CCND1 also induced Ki67 labeling in alpha, delta and PP cells in dispersed islets (Supplementary Fig. 5). The other combinations had no effect on delta or PP cells, whereas EZH2 overexpression in combination with silencing CDKN1C did low level induce Ki67 labeling in alpha cells.

**Trithorax dysfunction leads directly to CDKN1C reduction.** To determine if a direct link exists between trithorax members and cell cycle machinery in human beta cells, we selected the H3K27me3 demethylase, KDM6A, for further study, since it is a canonical trithorax member, since it is one of the rare recurrently mutated genes in insulinoma (Fig. 1), since it’s expression is reduced in insulinomas, (Supplementary Data 8) and since its biology has not previously been explored in human beta cells. We selected CDKN1C as a target because it is the only cell cycle inhibitor in the bisque4 MMP list (Fig. 5d), because it is reproducibly reduced in insulinomas (Fig. 5e, f), because its loss is associated with beta cell proliferation in FoCHI and BWS31–34, because CDKN1C is expressed only in beta cells in human islets23–25, 32, and, because silencing CDKN1C leads to induction of human beta cell proliferation in our hands (Fig. 5g) and others39. Inhibition of KDM6A, whether pharmacologically with GSKJ1 or adenovirally with shKDM6A, led to prompt and substantial reduction of CDKN1C in human islets, an effect not observed by adenoviral silencing the closely related KDM5B or KDM6B (Fig. 5h, i, Supplementary Fig. 6a–d). Chromatin immunoprecipitation (ChIP) analysis demonstrated that KDM6A directly interacts with a CDKN1C upstream enhancer in human beta cells (Fig. 5j). Further, the CDKN1C locus in human beta cells bears the H3K4me3 open chromatin mark, but lacks the H3K27me3 repressive chromatin mark, as anticipated by trithorax occupancy (Supplementary Fig. 6e). Collectively, these results reveal that KDM6A, and by extension, the trithoraxcomplex, directly support CDKN1C expression in human beta cells, and, conversely, that interfering with trithorax function results in CDKN1C loss, with consequent cell cycle entry. Notably, they also suggest yet another pharmacological approach to

---

**Fig. 6** A schema summarizing the broad heterogeneity of insulinomas and the molecular hallmarks that distinguish insulinoma from beta cells. **a** A waterfall plot summarizing cumulative coding mis-sense variants (black), CNV loss (blue), copy-neutral LOH (green), or copy number gain (red) among polycomb, trithorax and related chromatin-modifying genes. In the grid boxes, each vertical column represents the DNA landscape of each of the 26 insulinomas. Collectively, these observations highlight the frequency of multiple genomic and transcriptic abnormalities in polycomb, trithorax, and other epigenetic modifying genes across the majority of human insulinomas, the marked heterogeneity among insulinomas, the frequent mutation of multiple chromatin modifiers in almost all insulinomas, all despite having selected insulinomas independent of known MEN1 kindreds. **b** A cartoon emphasizing the relationship between polycomb, trithorax genes, and H3K27me3 and H3K4me3 marks in human beta cells and insulinomas. **c** An extension of **b** highlighting mutations (red stars), and abnormal CpG methylation. Subsequent panels highlight altered DNA methylation/imprinting abnormalities (d), asymmetrical gene expression from imprinted loci and abnormal chromatin marking patterns (e), all of which may lead to differential expression pattern of cell cycle genes such as CDKN1C (f), all of which may lead to increased beta cell proliferation in insulinomas.
induction of human beta cell proliferation (Supplementary Fig. 6f).

Discussion

Overall, this integrated and systematic analysis of a large human insulinoma cohort, focused on non-MEN1 insulinomas, reveals a near-universal preponderance of mutations in, and differential expression of, epigenetic modifying genes and their targets, together with structural and CpG methylation alterations of the key imprinted region on chromosome 11 (Fig. 6). Genomic mutations were particularly notable in genes related to the polycomb and trithorax complexes. While inactivation or loss of the trithorax complex member, menin, encoded by MEN1, is a well-known cause of insulinoma in mice and humans, the striking prevalence of mutations and/or dysregulation in many other trithorax and polycomb members in non-MEN1 insulinomas was not anticipated. In retrospect, it aligns nicely with prior work37, 38, 45–47, which suggest that the epigenome, through its broad control of cell cycle and function-modulating beta cell genes, represents the key rate limiting step in human beta cell proliferation.

Within this unifying context, the model also illustrates that multiple specific “epigenetic roads” may lead to insulinoma, and conversely, that despite their apparent clinical homogeneity, insulinomas display marked mutational heterogeneity, which, in retrospect, could only have been observed in a large number of insulinomas such as described herein. The simplest of these heterogeneous paths might be loss of the maternal chromosome 11, or chromosome 11 uniparental paternal disomy, as occurs in some cases of FoCHI, BWS, and pediatric insulinoma syndromes18, 31–34. Chromosome 11 errors, suggested by recurrent loss of this region and the unique ASE profile, as well as imprinting errors directly confirmed by deep CpG methylation sequencing, with congruent increases in H19, KCNQ1, and recurrent loss of CDKN1C (Fig. 4), similarly likely contribute to a large subset of insulinomas. Interestingly, the insulinomas with chromosome 11 loss are distinct from those with chromosome 7 gain, suggesting that these may represent two insulinoma subtypes. In the latter, the increase in E2H2 relative to normal beta cells may result from a CNV gain event. Importantly, the average of 3.5 driver mutations and multiple additional CNV gains and losses that characterize insulinomas (Figs. 1 and 6) is congruent with the observation that MEN1-associated human insulinomas occur only after decades5, 6, and that loss of both men1 alleles in mice nonetheless requires months of latency before insulinomas appear48. Thus, it appears inescapable that mutations in single genes such as MEN1 alone cannot cause insulinoma; instead, “hits” in multiple genes are likely required: MEN1 is just one of many core polycomb/trithorax genes (e.g., MEN1, KDM6A, YY1, E2H2, PGC5, KTM1C, CREGBBP etc.), which may contribute to the insulinoma phenotype.

The heterogeneity observed among insulinomas is reminiscent of the heterogeneity among human beta cells that has captured recent attention19–21. To determine whether insulinomas are also heterogeneous with respect to the hallmark markers of human beta cell heterogeneity, we explored CD9 and ST8SIA1 expression in our beta cell and insulinoma transcriptome data sets (Supplementary Table 10). Indeed, both ST8SIA1 and CD9 expression were highly variable among both human beta cells as well as insulinomas (Supplementary Fig. 7).

Since we are interested in inducing benign, rather than malignant growth of beta cells, it was reassuring to observe parallels between cell cycle genes and modules in insulinomas and proliferating juvenile beta cells (Fig. 5a, b), and also to observe that mutations in DNA repair genes such as ATRX, DAXX, TP53, BRCA2, CHEK2, and the PI3 kinase/PTEN/mTor/TSC pathway, characteristic of malignant non-functioning PNETs52, 53, were not observed in the insulinomas. This may reflect an important role for genes such as DAXX, ATRX, CHEK2, BRCA2, and TP53 in the rare malignant transformation of insulinomas; these genes may also provide clues to genes and pathways that must be avoided in attempts at therapeutic expansion of human beta cells.

Human insulinomas do not retain normal control of glucose sensing and insulin secretion. To explore mechanisms underlying this imperfect control, we surveyed families of so-called “disallowed genes”54, 55, glucose transporters and hexokinases, and beta cell-specific transcription factors (Supplementary Figs. 8–10). Remarkably, abnormalities were abundant and tantalizing here, including, alterations in glucose transporter and HK3 profiles, and a suggestion of restitution of expression of certain disallowed genes, although their expression remained at low levels. These observations provide clues to deeper understanding of the misregulation of insulin production in insulinoma. Canonical beta cell transcription factors such as PDX1, NKX6.1, MAFIA, and others were not substantially altered in insulinoma vs. beta cells. We also explored whether expression of mitogenic genes such as CCND1, E2H2, YY1, or silencing CDKN1C or KDM6A might alter expression of these families of genes in human islets (Supplementary Figs. 8–10). Reassuringly, these maneuvers had little effect, with the exception of overexpression of YY1, which altered expression of PDX1, NKX6.1, SLC16A1, and SLC2A2. These observations may suggest that it is best to avoid YY1 activation in therapeutic attempts to induce beta cell expansion.

We observed that several insulinoma co-expression modules were enriched not only for cell cycle genes (the bisque4 module), but also enriched for “immune function”, “vascularization”, “extracellular matrix”, “RNA splicing”, and “ubiquitination” (Fig. 2d, e). While we chose to focus on the bisque4 cell cycle module, the co-expression analysis highlights the importance of the beta cell microenvironment in supporting insulinoma growth and function. The immune-insulinoma connection was supported by the appearance of CIITA among the top DSG in insulinomas (Fig. 2c). Furthermore, many of the CIITA transcriptional targets, the MHC class II genes, were also differentially expressed in insulinoma relative to beta cells. Interestingly, YY1 and JARID’s are DNA-binding partners, which interact with, and regulate E2H2 recruitment to, the inducible promoter IV of CIITA56. While there have been prior reports linking beta cell proliferation to inflammation in type 1 diabetes57, this apparent link between CIITA and insulinoma is unexplained.

Collectively, this study provides a novel and complex lens through which to view insulinoma and its relationship to normal beta cell function, and raises a number of additional questions noted above that merit further study. For example, what drives the near universal E2H2 overexpression in insulinomas? The precise genetic and molecular pathophysiology remain unexplained. Similarly, what drives CCND1 overexpression? Further, these studies do not clearly distinguish between genomic/transcriptomic abnormalities that lead to beta cell proliferation, vs. those that lead to insulin overproduction. Clarification of this issue would be facilitated by a similar comparison of insulinomas to other benign PNETs that do not overproduce insulin. Similarly, it would be informative to compare these non-MEN1 insulinomas to the whole-exomes and transcriptomes of bona fide MEN1-associated insulinomas, to elucidate which features the two different types of tumors share on their passage from normal beta cell to adenoma. Finally, we highlight that previously reported mitogenic targets of the harmine class of beta cell mitogenic drugs24–26—DYRK1A (from the red module, Fig. 2), and MYC (also upregulated in insulinomas, Supplementary Data 8)
and its inhibitor, MNT (Fig. 6)—are also present in this analysis. The observation that pharmacologic or adenoviral KD/M6A interference can downregulate the cell cycle inhibitor, CDKN1C, suggest another entirely novel therapeutic target. Altogether, these observations clearly illustrate that “insulinoma data mining” can yield beta cell regenerative targets. Further studies involving whole-genome sequencing of insulinomas, defining IncRNA and microRNA profiles, whole-genome DNA methylation patterns, imprinting marks, histone marks, and chromatin structure, and comparing malignant to non-malignant insulinomas, will deepen the description of the proliferating human beta cell, and provide additional opportunities for therapeutic drug discovery.

Methods

Insulinomas. Thirty-eight human insulinomas were studied. Thirty-four were from nine individuals who provided informed consent, and were deposited in the Icahn School of Medicine at Mount Sinai Biorepository; WES data from an additional four insulinomas from Cao et al.8 passed our QC filter. The library was derived from distinct DNA molecules, thus avoiding over-counting possibly identical reads. Each insulinoma was sequenced at Mount Sinai, with the other four from Cao et al.8. Of the 38 insulinomas, 25 were subject to RNA sequencing; 13 from the same samples that underwent paired WES; 12 did not undergo WES because of a lack of paired WBC samples. All displayed marked increases in insulin RNA expression by RNAseq. Details of the patients with insulinomas are provided in Supplementary Table S1 and Supplementary Data1. All of the samples were obtained from subjects not known at the time to be in MEN1, although after WES, subject 3967 was found to have a germline MEN1 mutation.

Purified normal human beta cells. RNAseq was performed on purified beta cells from 22 donors of normal human cadaveric islets obtained from three sources. Five (three male, two female age (SEM) = 49.8 ± 7.6) and Mass Index (BMI) = 30.4 ± 2.6) were from islets provided by the NIH/NIDDK-supported Integrated Islet Distribution Program (IIDP) (https://iidp.cooh.org/overview.aspx). For these, beta cells were transduced 72 h before harvesting for fluorescence-activated cytomteric sorting (FACS) with an adenovirus driven by a RIP1-miniCMV construct that included 177 bases of the ICMV IE1 promoter (Clontech, Mountain View, CA)25. The beta cell fraction was confirmed to be >99% pure by immunolabeling of sorted cells with insulin, by quantitative reverse transcription PCR (qRT-PCR) and by RNAseq (see below). For ten of these FACS-sorted normal human cadaveric FACS-sorted islet beta cells were labeled using Newport Green were generously provided by Nica et al.24. For the remaining seven, FASTQ files were obtained from Blodgett et al.23.

DNA sequencing. Shearing of 0.5–1 µg genomic DNA to a mean of 200–300 bp fragments was performed using the Covaris E210 focused acoustic energy system (Covaris, Woburn, MA). Whole-genome libraries were prepared using either the NEBNext DNA Library Prep kit (New England Biolabs, Ipswich, MA) or KAPA Hyper Prep kit (Kapa Biosystems, Wilmington, MA) according to the manufacturer’s protocol. Illumina compatible paired-end adapters were used, and the adapter-ligated DNA fragments were amplified by ligation-mediated PCR (KAPA Biosystems, Wilmington, MA) using a reverse PCR primer containing a six nucleotide barcode that allowed for multiple samples to be pooled and sequenced in the same run. The library was enriched for exonic sequences with the SeqCap EZ Human Exome Kit v3.0 capture system (Roche NimbleGen, Madison, WI). The libraries were then sequenced with a 100 bp paired-end protocol on the Illumina HiSeq 2500 according to standard manufacturer’s protocol (Illumina, San Diego, CA).

Variant calling. Variant calling and filtering were carried out as described.10 For each individual insulinoma and blood sample, FASTQ files from all available WES runs were combined into a patient-specific “cohort” and run through an in-house pipeline to yield BAM and VCF files with germline and somatic variant calls (SNVs and small indels). Briefly, this in-house pipeline implements Genome Analysis Toolkit (GATK)35 version 2.7.2b best practices for alignment, base quality recalibration, variant calling (using HaplotypeCaller), and variant quality score recalibration (VQSQR).35 VQSR was set to 99.5% sensitivity. Read pairs whose S’ coordinates were identical were marked (except for one read pair) as duplicates by the Picard software (http://broadinstitute.github.io/picard/) and were not used for variant calling, per the above best practices, to ensure that evidence for each variant was derived from distinct DNA molecules, thus avoiding over-counting possibly over-amplified or over-sampled DNA. A GATK genomic interval list was created from the design file from the WES hybridization-capture kit manufacturer; sequencing depth (Supplementary Data 1) was computed only within these genomic intervals, whereas variant calling was done within these intervals, padded downstream on both sides. For somatic variant calling, MuTect61 (version 1.1.6-101bn19, HC + PON mode with default settings, using COSMIC22 version 65, dsSNP83 version 137, and using variant calls from patient-matched normal control as the “panel of normals” setting) and Varscan264 (version 2.3.5, with flags --tumor-purity 0.7 and --min-var-freq 0.07) were used. All variant calls were annotated using Snpeff v4.3B (using the Ensemble version 74/GRCh37 resource bundle) and loaded into a custom MySQL (Percona MySQL Server Community Edition 5.6.14-re162.0.483.rhel6) database schema using in-house scripts, where they were filtered as follows: only variants annotated by using the amino-acid sequence (missense, nonsense, affecting a canonical splice site, indel in coding sequence) were retained for interpretation; however, the full set of variants was used for routine post-sequencing QC described below. Somatic calls whose population allele frequency in either ESP5400 (http://evs.gs.washington.edu/ESV/) or 1000Genomes exceeded 2% were discarded on the presumption that they are any combination of: contamination, a variant present but missed in normal sample, a low-level artifact, or could not be pathogenic because it was too common in general population. All SNV and indel calls were manually reviewed in IGV and the UCSC Genome Browser89 to inspect supporting alignment quality in the BAM files and alignability of the genomic region in the hg19 human genome assembly, paying attention to whether a variant call was located in a short tandem repeat or a low-complexity sequence region90, a region with self-homology/duplication in the reference genome, or a region of low alignability according to the GEM track from ENCODE/CGP71. Uncertain calls were manually rejected at this step.

Indel force calling. To further reduce the indel FDR caused by any systematic artifact, we used UnifiedGenotyper (version 2.7.1) to force-call all indel variants that passed manual review across the entire cohort in order to systemically collect variant indel call signals in depth in all blood/tumor samples. Any indel variant with all read evidence from other normal blood samples was also removed from the final indel variant list reported.

Copy number alterations. Somatic copy number alterations were identified using saasCNV16. This algorithm uses heterozygous SNV calls from the normal control sample, and the change in their sequencing depth and allelic fraction in the tumor vs. the normal sample, to identify tumor-specific copy number change events. Joint circular segmentation is performed on the two signals: the log-ratio of depth and the log-ratio of mirrored allelic fractions in tumor vs. normal. Identified segments are then classified into one of these categories: loss, cnLOH, gain, normal (no change), or undetected. Because our GATK variant calling workflow combines normal and tumor as a cohort, heterozygous variants called in the normal sample are implicitly filtered out in the GATK process. For somatic samples, yielding their depth-by-allele in both normal and tumor in the output VCF. Variant calls passing VQSR, having a 0/1 genotype in normal, having a mapping quality > 30, and having non-zero depth in tumor, were input into saasCNV. The saasCNV segments classified as loss, gain, and cnLOH CNVs were input to variant calls with the Circos plot for visualization of recurrent CNV regions. Finally the statistics for significant recurrent CNV regions was analyzed by GISTIC2.0 using the segmentation output of saasCNV with log2ratio values74. GISTIC2.0 analysis was performed using default parameters at q-value threshold of 0.1.

Targeted DNA methylation analysis. Two sets of DNA (300 ng) from four sorted adult beta cell samples was utilized. Three-hundred nanogram DNA was obtained from one sample, and three other sorted beta cell DNA preparations were pooled to achieve the 300 ng minimum. Three-hundred nanogram DNA was prepared from ten insulinomas, of which six were in the original sample (4973, 5329, 5333, 5863,
6066, 6107) and four were new samples (5322, 5596, 5597, 5326). Probes were designed to capture the region of chromosome 11 spanning 1.35 Mb from position 1,850,000 to 3,200,000 within bands p15.5-15.4, (GRCh37/hg19 Assembly Feb 2009) according to the SeqCap Epi Enrichment System (Roche NimbleGen, Roche Sequencing & Life Sciences, Indianapolis, IN, USA) for hybridization-based target enrichment of bisulfite-treated DNA. Pre-capture libraries were prepared using the Kapa Hyper Prep kit, PCR-free version (Roche Sequencing & Life Sciences). The SeqCap Epi was hybridized to the custom SeqCap Epi Choice probe pool at 42 °C for 72 h. Hybridized products were purified by capture with Capture Beads and PCR amplified for 15 cycles to create the final libraries for sequencing. Final yields were quantified in a Qubit 2.0 Fluorometer (Life Technologies, Grand Island, NY, USA), and quality of the library was assessed on a DNA1000 Bioanalyzer chip (Agilent Technologies, Santa Clara, CA, USA). Three post-capture multiplexed libraries, each containing four different indexes, were normalized to 10 ng, pooled at 10 pm on a V2 paired-end read flow cell and sequenced for 150 cycles on an Illumina MiSeq (Illumina, San Diego, CA, USA). Primary processing of sequencing images was done using Illumina’s Real Time Analysis software (RTA) as suggested by Illumina. CASAVA 1.8.2 software was then used to demultiplex samples, generate read counts, and to assess the quality scores. Each sample was trimmed and reads aligned to the bisulfite converted reference human genome (GRCh37/hg19 Assembly Feb 2009—whole-genome alignment approach) and the methylation context for each cytosine determined. Custom scripts were used to compute the percent methylation scores and average conversion rates. Average conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinucleotides assessed by the targeted DNA methylation conversion rates obtained ranged from 99.56 to 99.77%. The percent methylation data across the CpG dinuc
plotted as follows: for each histone mark track, each gene was given a binary value (1 for genes with positive histone mark, and 0 for genes with negative histone mark). A sliding window of 201 genes was defined for each gene, including 100 flanking genes on the left and 100 flanking genes on the right, according to the gene-ranking order (increasing or decreasing expression differences between insulinoma vs. beta cells). This was employed to calculate the average gene set scores for each sliding window. These average scores were used to plot the visualization curve for each track.

One-sided Fisher’s exact test was performed between genes upregulated in insulinoma (when compared with normal beta cells) and each histone mark signature, with all the protein-coding genes tested in the differential expression analysis between insulinoma and normal beta cells described above as background set.

For the enrichment of histone mark signatures in insulinoma co-expression network (Fig. 3b), a two-sided Fisher’s exact test was performed between each histone mark signature and each co-expression module, and FDR value was set at 0.5. FDR values were then transformed into FDR scores using the following definition: \( -\log_2(\text{FDR}) \) for over-enrichment (odds ratio from Fisher exact test > 1) and \( \log_2(\text{FDR}) \) for under-enrichment (odds ratio from Fisher exact test < 1). All insignificant FDR scores (when FDR > 0.05) were converted to a value of 0. The maximum absolute value of FDR score was set at 50. Finally, the FDR score matrix was used to create the enrichment visualization curve for each track.

### Chromatin immunoprecipitation assays
ChiP was performed using the EZ-ChIP Kit (Millipore) according to the manufacturer’s protocol. Beta cells were FACs sorted using Ad.ZsGreen as described earlier. Forty-four thousand ZsGreen+ (beta) and ZsGreen− (non-beta) cells were collected per experiment for each H3K4me3 and H3K27me3 immunoprecipitation, and 130,000 sorted cells were collected for each KDM6A immunoprecipitation. The primer sets were designed in a previous study. Immuneoprecipitated DNA was quantified using ABI 7500 real-time quantitative PCR detection system (Life Technologies). The following antibodies were used: anti-H3K4ac (Millipore #17-622), anti-KDM6A (Abcam #ab84190). Data are presented as ChiP reads normalized relative to input controls, and fold-enrichment divided by respective IgG. Three separate islets preparations were used for each figure shown.

### Gene expression analysis
Gene expression was analyzed by real-time PCR performed using an ABI 7500 system. For RT-PCR, all data are expressed as the mean ± SEM. Results were accepted as statistically significant at \( p < 0.05 \), as determined using two-tailed Mann–Whitney test. Primer sequences used are listed in Table 1.

### Data availability
DNAseq, RNaseq, and Cpg Bisulfite seq raw data files are available through dbGAP (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/about.html) under accession number: phs001422.v1.p1 and The Catalogue of Somatic Mutations in Cancer (COSMIC) via ID number: COSP44132. Processed data are also available on our Insulinoma Genomic Portal at http://insulinoma.genomicportal.org

Received: 27 March 2017 Accepted: 10 August 2017
Published online: 03 October 2017

### References
1. Wang, P. et al. Diabetes mellitus—advances and challenges in human beta-cell proliferation. *Nat. Rev. Endocrinol.* 11, 201–212 (2015).
2. Dirice, E. et al. Inhibition of DYRK1A stimulates human beta-cell proliferation. *Diabetes* 65, 1660–1671 (2016).
3. Shen, W. et al. Inhibition of DYRK1A and GSK3B induces human beta-cell proliferation. *Nat. Commun.* 6, 8372 (2015).
4. Wang, P. et al. A high-throughput chemical screen reveals that harmine-mediated inhibition of DYRK1A increases human pancreatic beta cell replication. *Nat. Med.* 21, 383–388 (2015).
5. Davis S. N., Lamos E. M. & Younk L. M. in *Endocrinology: Adult and Pediatric 7th edn* (eds DeGroot, L. I. & Jameson, J. L.) ch. 47 833–836. (Elsevier, 2016).
Acknowledgements

We wish to thank the following people and organizations for their support: The Foundation for Diabetes Research, the Murray-Helig Fund in Molecular Medicine at the University of Connecticut; the NIDDK Islet Integrated Distribution Program (IIDP); Dr Tapaszy Kim at the University of Edinburgh and Dr Patrick MacDonald at the Alberta Diabetes Institute; The Biorepository and Pathology Core, the Flow Cytometry Center for Research Excellence, and the computational resources and staff expertise provided by the Department of Scientific Computing at the Icahn School of Medicine at Mount Sinai; The Human Islet and Adenoviral Core (HIAC) of the Einstein-Sinai Diabetes Research Center (DRC); The Epigenomics Core of Weill-Cornell Medicine. We thank Drs Ronald Tamler, Erika Villanueva for helping to obtain insulinomas; Drs Yanan Cao and Guang Ning of the Shanghai Key Laboratory for sharing data; Mr Justin Bellizzi for technical assistance; and Drs Alexandra Nica, Philippe Halban, Manolis Dermitzakis, and Drs David Bledgett and David Harlan for sharing their normal beta cell transcriptome data with us. We thank Drs Adolfo Garcia-Ocaña and Rupangi Vasavada for continued discussion of these findings. Dr Jordi Ochando for help with human beta cell flow cytometry, Dr Christopher Newgard for sharing the RIP1-mini-CMV enhancer promoter for construction of the RIP1-ZsGreen adenovirus, and Ms. Aye Moe and Drs Wei-Yi Cheng, Ke Hao and Antonio di Naro for bioinformatic support. Finally, we apologize to the authors whose important work we were unable to cite because of reference number limitations. This work was supported by NIDDK grants U54 DK104211 and P-30 DK 020541.

Additional information

Supplementary Information accompanies this paper at doi:10.1038/s41467-017-00992-9.

Competing interests: The authors declare no competing financial interests.

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

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.