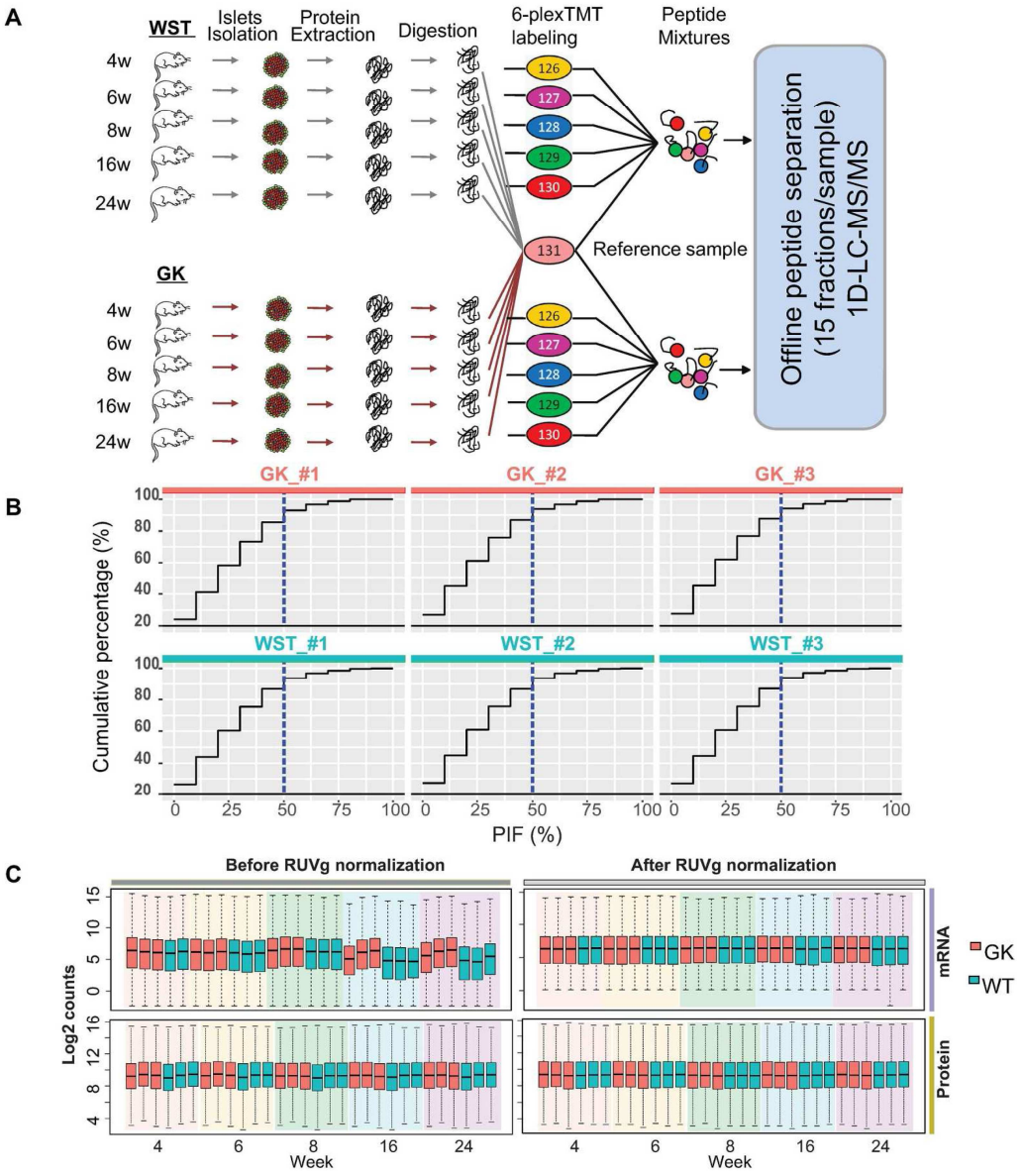
SUPPLEMENTARY DATA

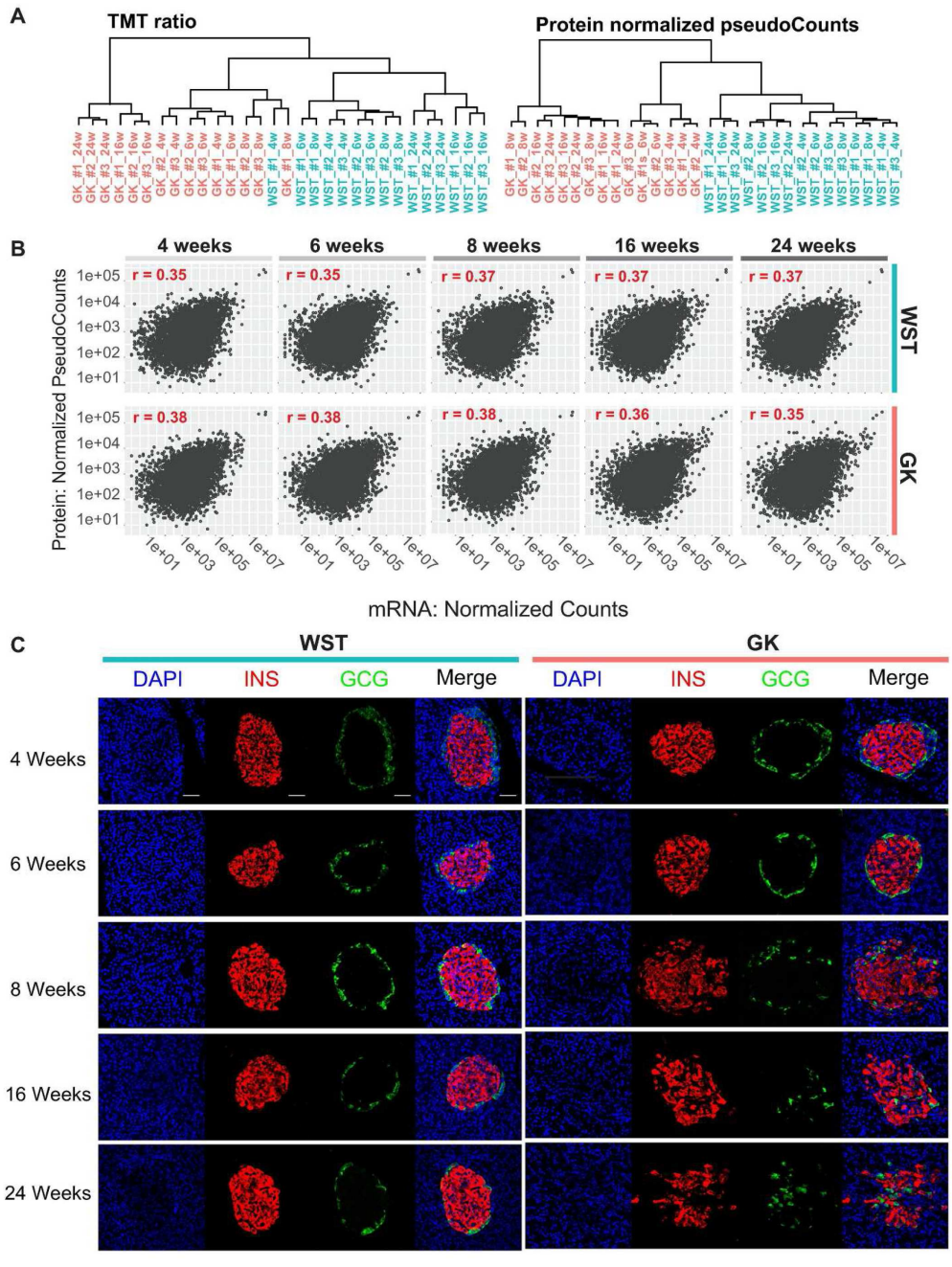
Supplementary Figure 1. Experimental design of proteomic analysis and data processing. (**A**) The experimental design of the proteomic workflow, including protein extraction from isolated GK and WST rat islets, tryptic digestion, 6-plex TMT peptide labeling, offline peptide separation via high-pH reversed-phase chromatography, and repeated 1D-LC-MS/MS analysis of each fraction. (**B**) The accumulation curve for quantifiable proteins with different PIF cut-off values. The vertical blue broken line indicates 50% PIF. (**C**) Box-plotting comparisons of raw mRNA counts and protein pseudocounts before and after RUV normalization. The y-axis indicates the log2 transformation of counts.



Experimental design of proteomic analysis and data processing.

SUPPLEMENTARY DATA

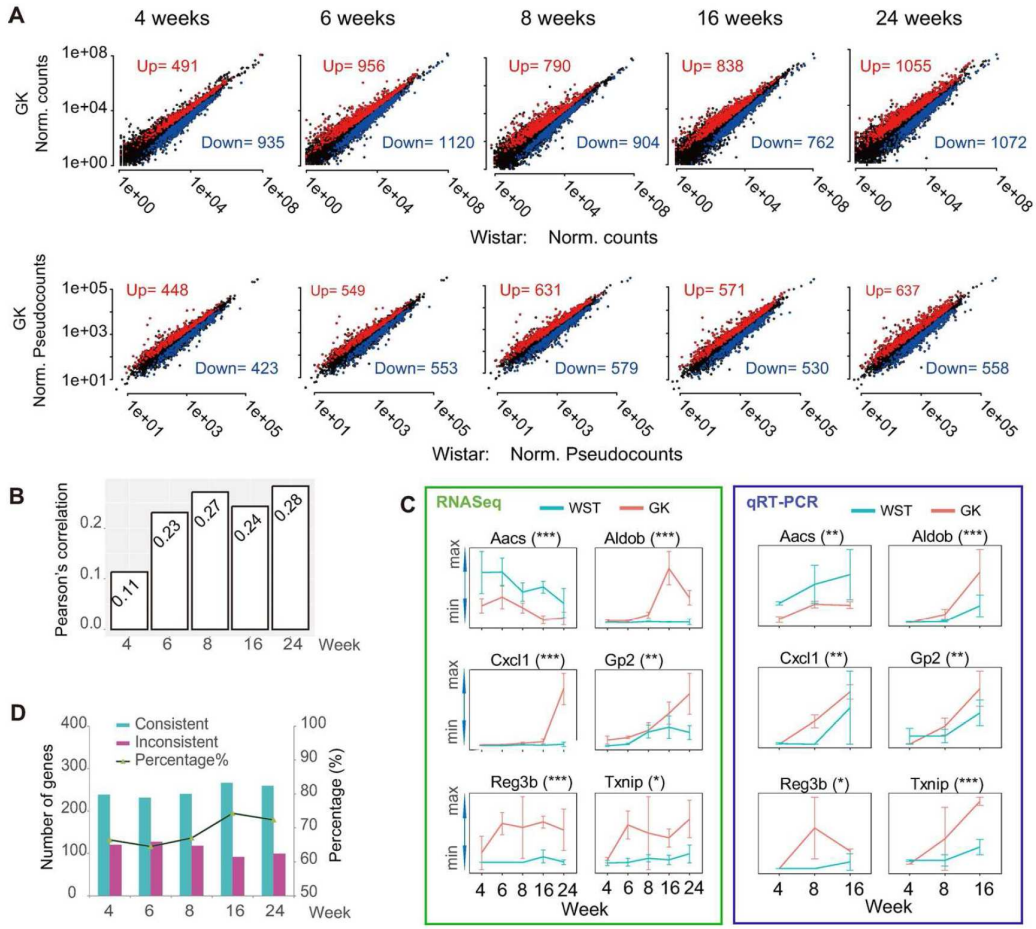
Supplementary Figure 2. Extension of data analysis and morphological analysis of islets. (**A**) Comparison of hierarchical clustering results between TMT ratios and protein normalized pseudocounts. (**B**) Spearman correlation analysis between mRNA normalized counts and protein normalized pseudocounts for each sample. (**C**) Immunohistochemistry micrographs of GK and WST rat islet paraffin sections co-stained for insulin (INS, red) and glucagon (GCG, green) at different time points. Scale bar, 50 μm.



Extension of data analysis and morphological analysis of islets.

SUPPLEMENTARY DATA

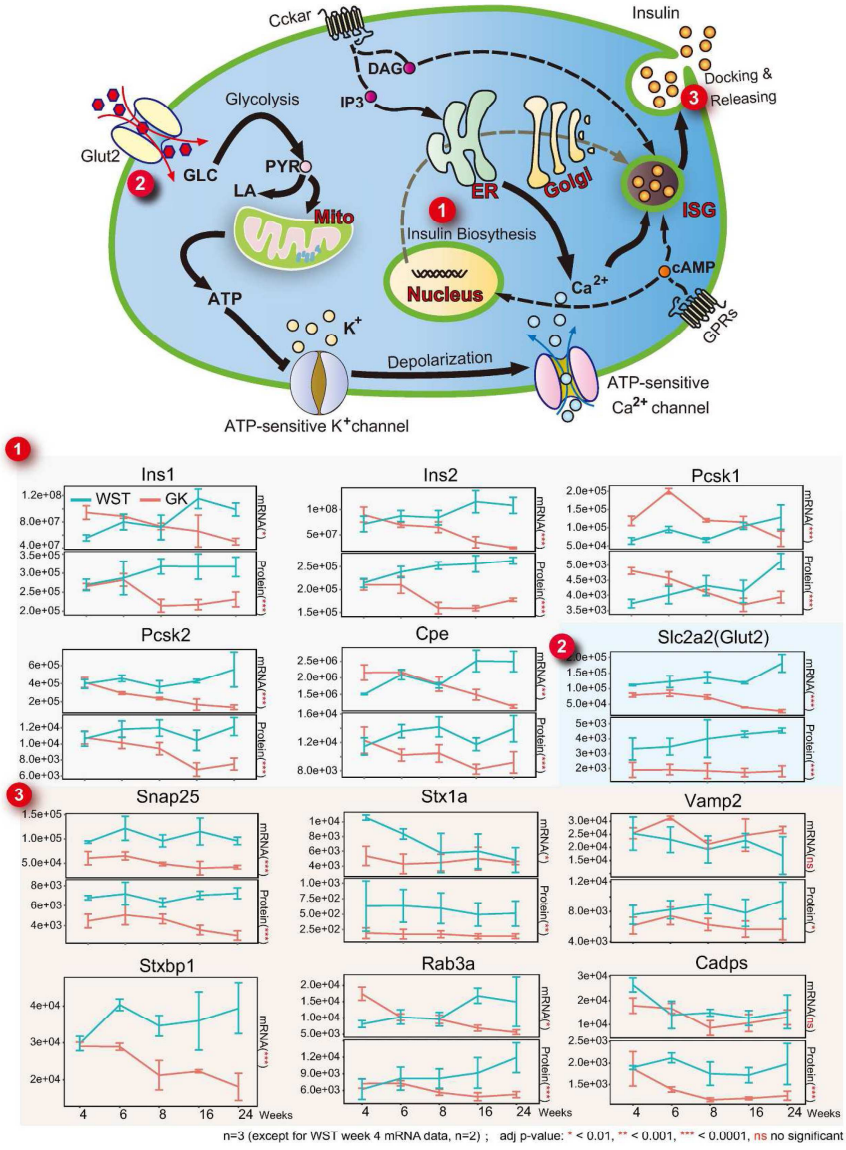
Supplementary Figure 3. Transcriptome and proteome comparisons between GK and WST rats at different time points. (**A**) Scatter plots of quantifiable genes from the transcriptomic and proteomic data at each time point. Up- and down-regulated genes are highlighted in red and blue. (**B**) Bar plot of Pearson's correlation coefficients from the transcriptome and proteome at each time point. (**C**) Random selection of 6 genes for the validation of our omics data by qRT-PCR methods (n = 3-5). The annotation of "*" is the same as described in Fig. 6. (D) The histogram of overlapping DE genes both in our mRNA dataset and recent publically available dataset of islets from normal and T2D individuals (Supplementary Table 1, Ref. 10). The genes either up-regulated or down-regulated in two datasets are considered as consistent ones, otherwise as inconsistent ones. The line indicates the percentage of the consistent genes at five different time points.



Transcriptome and proteome comparisons between GK and WST rats at different time points.

SUPPLEMENTARY DATA

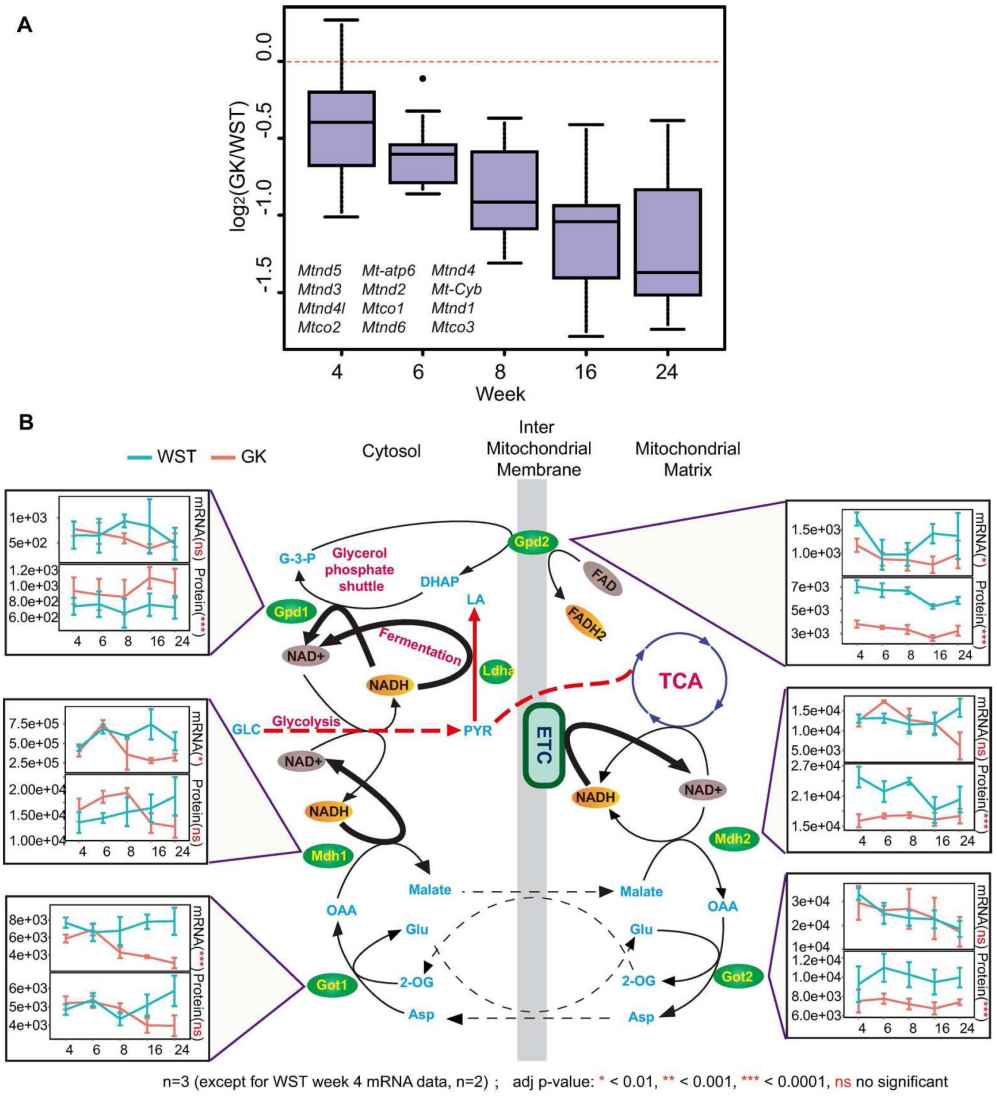
Supplementary Figure 4. Insulin secretion was impaired in GK β-cells over the duration of the evaluated time course. The cartoon depicts the primary signal flow of GSIS by pancreatic β-cells, including glucose transport, glucose metabolism, Ca^{2+} flux, insulin secretion granule docking and release, and insulin biosynthesis. The time course gene expression data for several key genes are selectively displayed as a line chart with the same figure annotations as in Fig. 6.



Insulin secretion was impaired in GK β-cells over the duration of the evaluated time course.

SUPPLEMENTARY DATA

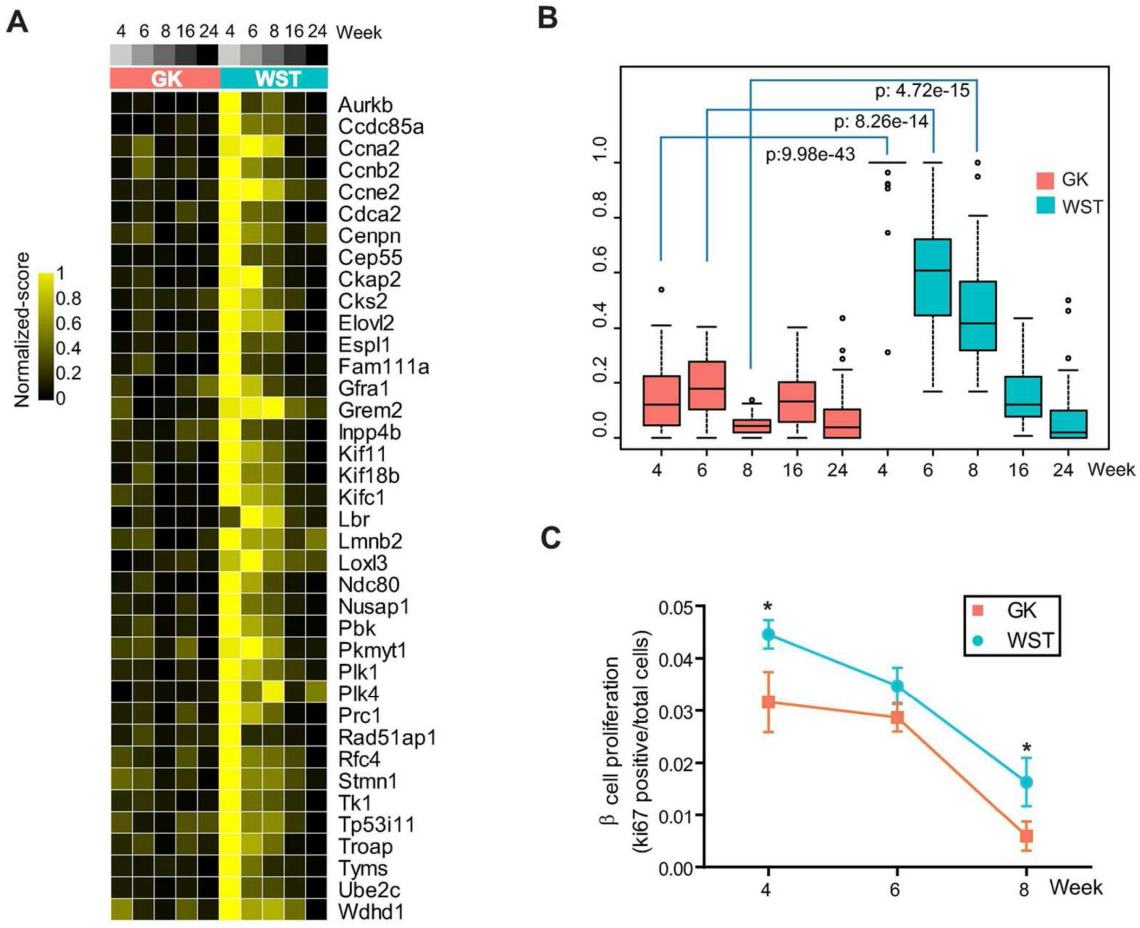
Supplementary Figure 5. Mitochondrial dysfunction in GK islets. (**A**) Box plot of log2-fold-changes (GK vs WST) from 12 mitochondria-encoded mRNAs at each time point. (**B**) The malate-aspartate shuttle and glycerol phosphate shuttle were abnormal in GK rat islets. Got1, Got2, Mdh1, Mdh2 and Gpd2 were significantly down-regulated during diabetes development in GK rats, resulting in an increased NAD^+/NADH ratio.



Mitochondrial dysfunction in GK islets.

SUPPLEMENTARY DATA

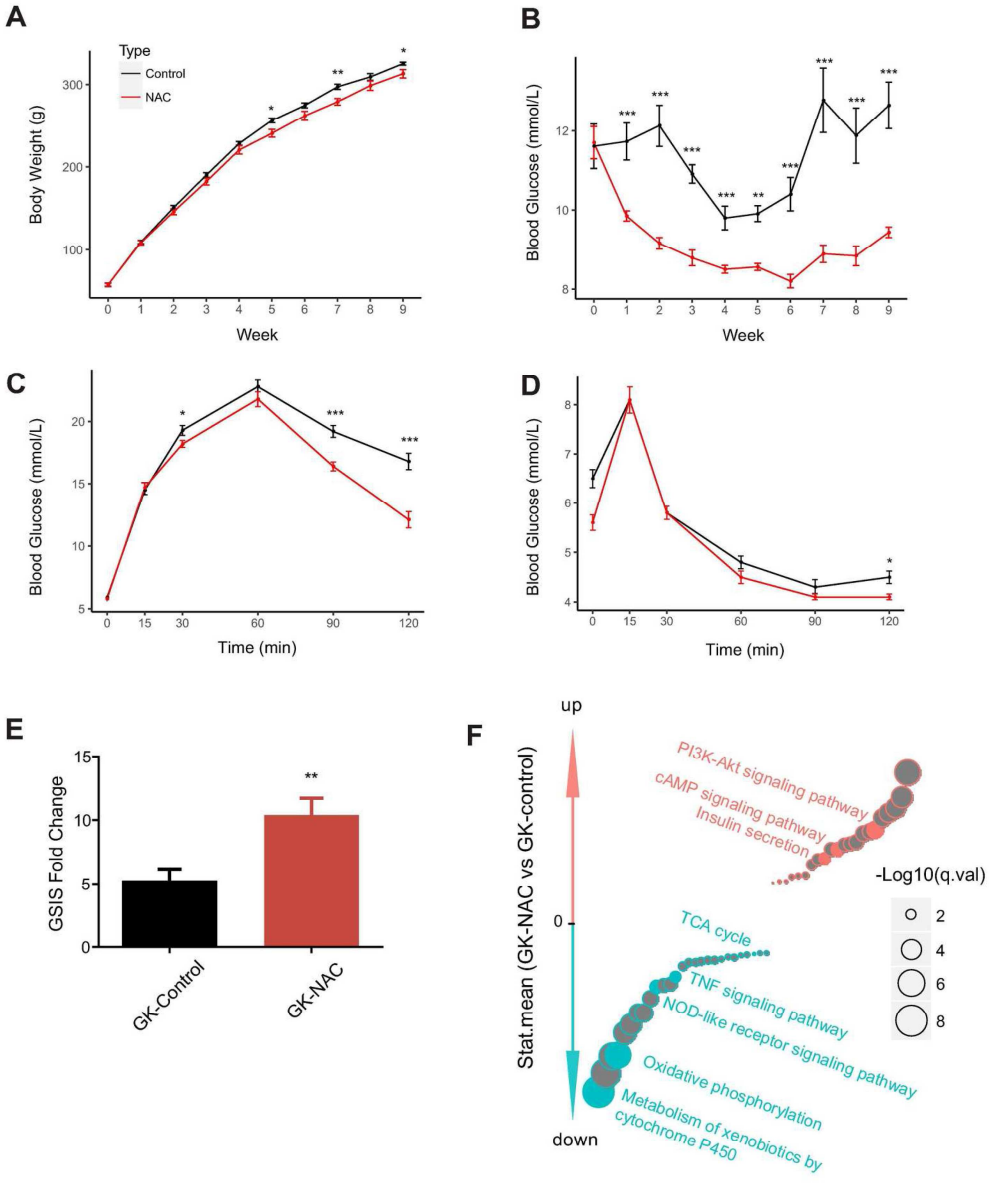
Supplementary Figure 6. Defective β-cell proliferation in GK islets. (**A**) Heatmap of the mRNA expression of 38 genes functionally related to the cell cycle. The mRNA expression data for each gene were normalized to a range of 0-1 via the min-max normalization method. (**B**) Box plot of the mRNA expression data in (**A**). A two-tailed Student's t test was applied to the pairwise comparisons, revealing significantly different mRNA expression levels between GK and WST islets at 4, 6, and 8 weeks. (**C**) The β-cell proliferation rate was measured based on Ki67 (cell proliferation marker) and insulin staining. All data are presented as the mean ± SEM. "*" indicates $p < 0.05$. n = 3 independent experiments.



Mitochondrial dysfunction in GK islets.

SUPPLEMENTARY DATA

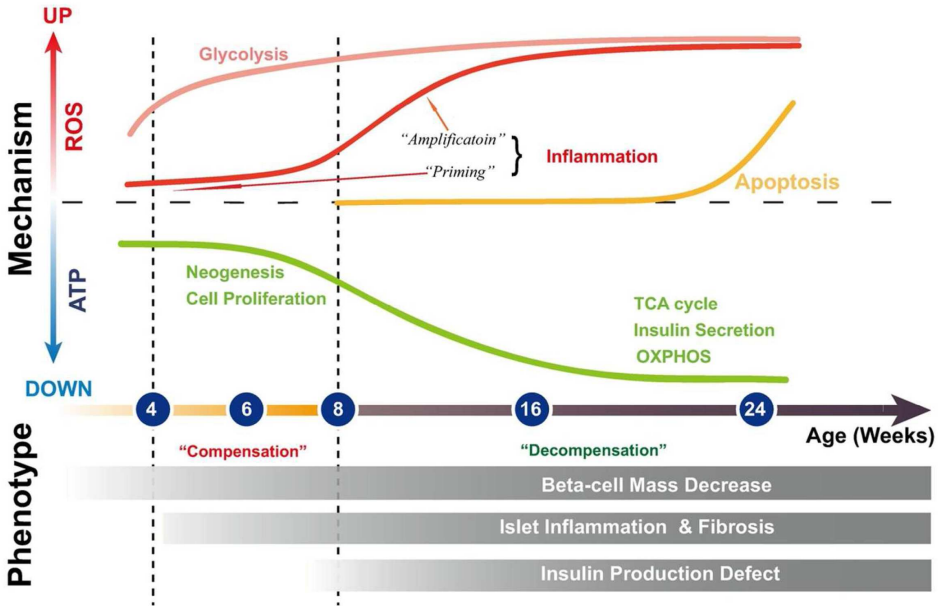
Supplementary Figure 7. NAC treatment experiment. (**A**-**D**) Body weights (**A**), random blood glucose (**B**), GTT (**C**) and ITT blood glucose (**D**) were measured (n = 8). (**E**) After NAC-treatment for 2 weeks, GSIS of isolated islets in GK-NAC and GK-control was evaluated by calculating the fold change of secreted insulin amount under high (16.7 mM) and low (2.8 mM) glucose stimuli (n=7). All data are presented as the mean ± SEM. "*" indicates $p < 0.05$. "**" indicates $p < 0.01$. "***" indicates $p < 0.001$. (F) Selected plotting of enriched KEGG pathways analyzed by GAGE comparing the transcriptome between GK-NAC and GK-control ($q < 0.05$). The "Stat.mean" values represent the averaged magnitude and direction of fold-changes at the gene-set level.



NAC treatment experiment.

SUPPLEMENTARY DATA

Supplementary Figure 8. Time course of pancreatic islet deterioration in GK rats. The early stage (4-6 weeks) is characterized by up-regulated anaerobic glycolysis, inflammation priming, down-regulation of the TCA cycle, defective insulin secretion due to impaired mitochondrial metabolism, and compensation for insulin synthesis and OXPHOS gene expression at the transcription level. The late stage (8-24 weeks) is characterized by inflammation amplification, reduction in insulin production and compensation failure. No significant apoptosis occurred during the early stage, but apoptosis was elevated at 24 weeks at the mRNA level. Neogenesis and cell proliferation were impaired starting during the early stage, resulting in β-cell mass reduction.



Time course of pancreatic islet deterioration in GK rats

SUPPLEMENTARY DATA

Supplementary Table 1. The reported transcriptomes and proteomes of islets from various T2D animal models and human cadaver donors.
Supplementary Table 2. All the information about identification and quantification of mRNAs and proteins.
Supplementary Table 3. Pearson's correlation analysis of the temporal mRNA and protein expression of 746 overlapping DE genes, related to Figure 2A.
Supplementary Table 4. Time course dynamic expression clustering analysis of DE genes, related to Figure 2C.
Supplementary Table 5. KEGG pathway enrichment analysis.
Supplementary Table 6. Unsupervised hierarchical clustering analysis of mitochondria-related genes
Supplementary Table 7. DE genes that were identified in NAC experiment by RNASeq, and the result of KEGG pathway enrichment analysis.

* All the tables can be accessed via the website of http://www.ibp.cas.cn/ibplwfjtj/xutao/201611/t20161117_4697642.html.

SUPPLEMENTARY DATA

Supplementary Experimental Procedures

Preparation of pancreatic islets from GK and Wistar rats

Pancreatic islets from male diabetic GK and control Wistar rats were isolated by collagenase digestion. Briefly, pancreases were inflated by instilling 5 ml of 0.5 mg/ml collagenase P (Roche Applied Science, Mannheim, Germany) in Hanks' buffered saline solution (HBSS) through the pancreatic duct, dissected out, and incubated in a water bath at 37 °C for 20 min. The digested pancreases were rinsed with HBSS, and islets were separated on a Ficoll density gradient. After three washes with HBSS, islets were carefully hand-picked under a dissection microscope to remove extra-islet tissues as much as possible. The purity of islets was evaluated by checking the expression of Amy1a (a marker of acinar cells) and Krt19 (a marker of ductal cells). In our omics dataset, Amy1a was slightly higher at protein level (only at 16 weeks) in GK islet, but not statistically different at mRNA level at any time points. For Krt19, it was only detected at mRNA level with no statistical difference. This indicates that the contaminants in islets were fairly minimized.

Random blood glucose assay

Random-fed blood glucose levels were measured weekly from the tail vein with an Accu-Chek glucometer (Roche Diagnostics, Mannheim, Germany).

Glucose tolerance test (GTT)

Rats were fasted overnight (14 h) prior to the administration of glucose (1.5 g/kg body weight) by oral gavage. Glucose measurements were taken at 0, 15, 30, 60, 90 and 120 min post-administration using an Accu-Chek glucometer (Roche Diagnostics, Mannheim, Germany). Blood was collected from the tail vein at each time point during the glucose tolerance test. Rats were denied access to food during the study.

Insulin tolerance test (ITT)

After 6-h fasting, rats were injected with regular human insulin (NovolinR, Novo Nordisk) at a dose of 0.75 U/kg of body weight. Blood samples were collected from the tail vein at 0, 15, 30, 60, 90 and 120 min post-injection to determine blood glucose concentrations.

Glucose-stimulated insulin secretion (GSIS)

The insulin secretion induced by glucose in primary islets was assessed as previously described (1). The primary islets of rats were isolated through collagenase (Roch) perfusion method. The islets were cultured in RPMI160 medium containing 100 U/ml penicillin, 100 μg/ml streptomycin and 10% fetal bovine serum overnight, and then incubated for 1 h in Krebs-Ringer buffer (KRB) containing 2.8 mM and 16.7 mM glucose serially. The insulin concentration in the collected supernatant was measured using an insulin ELISA kit (Millipore). The amount of protein was determined using a BCA kit.

RNA-Seq library preparation, sequencing and data processing

Isolated islets (the number was around 100-150 at 4-6 weeks and 100-120 after 8 weeks) were immediately placed in TRIzol reagent (Invitrogen, cat. no. 15596018). RNA was extracted by adding chloroform, followed by isopropanol precipitation. The RNA pellet was washed twice with 70% ethanol, dried, and resolved in RNase-free water. A260/A280 and 28s/18s rRNA ratios were determined to ensure that the RNA samples were highly purified and not degraded.

The MAPS (Multiplex Analysis of PolyA-linked Sequences) protocol has been published previously (3). Briefly, TRIzol-isolated RNA (1 μg) was enriched for Poly(A+) RNA with biotinylated oligo(dT) and converted to cDNA with Superscript III (Invitrogen, cat. no. 18080-051). Next, a terminal transferase (NEB, cat. no. M0315S) was used to add a ddNTP to the 3'-end of the cDNA. After second-strand cDNA synthesis, 23 PCR cycles were performed to amplify the cDNA. PCR products in the range of 200-400 nt were subjected to deep sequencing on a HiSeq 2500 (Illumina).

The analysis was performed using scripts generated in-house. Fastq files were aligned (4) to merged transcript sequences obtained from the UCSC RefGene database (5) with the mapping program Bowtie (v1.1.1) using the following parameter: "-l25 -n2 -M3 -e 200 --best --strata --phred33-quals --trim5 4".

SUPPLEMENTARY DATA

The unmapped tags in the first round were then mapped to the rat genome (rn6). For each gene, only tags that were uniquely mapped and localized in exons or exon-exon junctions were counted.
Due to limit amount of islets from one animal, the islets for transcriptomics and proteomics (see below) were not from the same animal, but from the same offspring to minimize inter-animal variations.

Protein sample preparation for proteome analysis

Rat islets (the number was 150-200 at 4-6 weeks and around 100-150 after 8 weeks) were lysed in a buffer containing 8 M urea/100 mM NH_4HCO_3 (Sigma, cat. no. A6141), pH 8.3, and a protease inhibitor cocktail (Sigma, cat. no. P8340-1ML). Proteins (50 μg) were first reduced by adding a final concentration of 10.0 mM dithiothreitol for 30.0 min at 37 ºC and were then immediately alkylated by incubating with a final concentration of 20 mM iodoacetamide for 45 min at 37 ºC in the dark. Resulting mixtures were diluted to less than 2 M urea with 100 mM NH_4HCO_3 prior to digestion overnight with sequencing-grade trypsin (Promega, cat. no. V5111) at a substrate/enzyme ratio of 100:1 (w:w). Digestion was quenched via acidification with formic acid (FA). Peptide desalting was subsequently performed via solid phase extraction (Sep-pack Vac C18 cartridges, Waters, cat. no. Wat020515), followed by vacuum-drying.
The resulting peptide samples were re-suspended in 100 mM triethylammonium bicarbonate (TEAB) buffer. An aliquot of 50% of each sample was chemically labeled with TMT reagents (Thermo Fisher, cat. no. 90066) according to the manufacturer's instructions. To accommodate all 10 samples in the analysis of a single biological replicate, we created a reference sample by mixing 1/5 of the amount of each sample (10 in total) and then labeling with TMT reagent 131. The remainder of the TMT reagents were used to label the other samples: 126 for GK, WST_4w, 127 for GK, WST_6w, 128 for GK, WST_8w, 129 for GK, WST_16w and 130 for GK, WST_24w. The reference sample was used for data normalization and dataset combination. To ensure that each of the samples contained the same amount of protein, a small 1:1:1:1:1:1 aliquot was prepared and analyzed by MS. Summed reporter ion ratios informed mixing ratios of the remaining labeled digests. In this way, two sets of 6-plex TMT-labeled peptide mixtures were equally pooled and were treated in parallel throughout the following steps.
Prior to MS analysis, TMT-labeled peptide mixtures were fractionated using high-pH reversed-phase chromatography. Briefly, the samples were first desalted using Sep-Pak Vac C18 SPE cartridges (Waters, Massachusetts, USA) and dried in a vacuum concentrator. Desalted peptides were dissolved in solution A (2% Acetonitrile, pH 10, pH adjusted with ammonium hydroxide) and were then loaded automatically onto a YMC-Triart C18 basic reversed-phase liquid chromatography column (250 × 4.6 mm, 5 μm particles) (cat. no. TA12S05-2546WT, YMC, Kyoto, Japan). Peptides were separated in a binary buffer system of solution A and solution B (98% Acetonitrile, without pH adjustment, solution B) in an Ultimate-3000 LC system (Thermo Scientific, Massachusetts, USA). The gradient of buffer B was set as follows: 0-5% for 1 min, 5-15% for 5 min, 15-26% for 32 min, 26-40% for 22 min, 40-95% for 2 min, and 95% for 3 min. The first fraction was collected starting at 5 min, and the remaining eluates were collected at intervals of 84 s. A total of 45 fractions were obtained, and these were concatenated into 15 fractions by merging fractions 1, 16, 31; fractions 2, 17, 32; and so on. Then, all 15 peptide fractions were dried in a vacuum concentrator and stored at -20 °C until subsequent nanoLC-MS/MS analysis was performed.

NanoLC-MS/MS analysis

NanoLC-MS/MS experiments were performed on a Q Exactive mass spectrometer (Thermo Scientific) coupled to an Easy-nLC 1000 HPLC system (Thermo Scientific). The peptides were loaded onto a 100-μm id × 2-cm fused silica trap column packed in-house with reversed-phase silica (Reprosil-Pur C18 AQ, 5 μm, Dr. Maisch GmbH) and then separated on a 75-μm id × 20-cm C18 column packed with reversed-phase silica (Reprosil-Pur C18 AQ, 3 μm, Dr. Maisch GmbH). The loaded peptides were eluted with a 78-min gradient. Solvent A consisted of 0.1% FA in water solution, and solvent B consisted of 0.1% FA in acetonitrile solution. The following segmented gradient was used at a flow rate of 280

nl/min: 4-12% B, 5 min; 12-22% B, 50 min; 22-32% B, 12 min; 32-95% B, 1 min; and 95% B, 7 min. The mass spectrometer was operated in data-dependent acquisition mode, and full-scan MS data were acquired in the Orbitrap with a resolution of 70,000 (m/z 200) across a mass range of 300-1600 m/z. The target value was 3.00E+06 with a maximum injection time of 60 ms. After the survey scans, the top 20 most intense precursor ions were selected for MS/MS fragmentation with an isolation width of 2 m/z in the HCD collision cell and an optimized normalized collision energy of 32%. Subsequently, MS/MS spectra were acquired in the Orbitrap with a resolution of 17,500 (m/z 200) and a low-mass cut-off setting of 100 m/z. The target value was set as 5.00E+04 with a maximum injection time of 80 ms. The dynamic exclusion time was 50 s. The following nanoelectrospray ion source settings were used: spray voltage of 2.0 kV, no sheath gas flow, and a heated capillary temperature of 320 °C. Each fraction was repeatedly analyzed.

MS data processing

The raw MS data were processed with Proteome Discovery (version 1.4, Thermo Scientific). Briefly, peptide identification was performed with both Sequest HT and the Mascot 2.3 search engine comparing against a UniProt database (version 5.62) supplemented with all frequently observed MS contaminants. The following parameters were used for database searching: 10 ppm precursor mass tolerance, 0.02 Da fragment ion tolerance, up to two missed cleavages, carbamidomethyl cysteine, TMT modification on amino (N)-term and lysine as fixed modifications, and oxidized methionine as a variable modification. The peptide confidence was set to a high level (q-value < 0.01) for peptide filtering.

To improve the accuracy and confidence of protein quantification, optimized data processing was developed using freely accessible tools and in-house written scripts (available upon request). 1) msconvert (http://proteowizard.sourceforge.net) was first used to perform a deconvolution of the high-resolution MS2 spectra, in which all fragment ion isotopic distributions were converted to an m/z value corresponding to the monoisotopic single charge. The signals of TMT reporter ions were extracted with the following requirements: maximum mass accuracy of 15 ppm, detection of all 6 TMT reporter ion channels required. 2) The summed reported ion intensity from each channel for all acquired MS2 spectra was used for sample normalization. The intensities of 131 channels representing reference samples in two 6-plex TMT experiments were utilized as conjunct factors for data normalization such that comparison between different time point samples from GK and WST rats achieved the same level. 3) To minimize ratio distortion due to the presence of more than one peptide species within a precursor ion isolation width, we also rejected the quantification of MS/MS spectra based on the precursor intensity fraction (PIF). For our dataset, a PIF of 50% was selected as the optimal trade-off value for both identification and quantification. 4) If a specific peptide was quantified multiple times, the peptide with the lowest PIF was selected to produce the representative quantitative TMT ratio. The median values of the TMT ratios of peptides from the same protein were calculated as the protein ratios. 5) To maintain experimental design power for additional statistical analyses, the protein TMT ratio was further processed with the following step. The normalized spectral abundance factor (NSAF) for each protein was calculated using an in-house Perl script. After multiplying by an amplification factor (1.0E+06 was adopted in this study), NSAF values were transformed into natural integer numbers, which were considered summed 6-sample pseudo-abundances that were distributed to each sample based on individual TMT ratios, hence the name “pseudocount”. ANOVA was adopted to analyze differentially expressed proteins between GK and WST islets.

Hierarchical clustering and principal component analysis (PCA)

To assess the similarity among all samples using the transcriptomic and proteomic datasets, unsupervised hierarchical clustering was performed using the “pheatmap” package in R. PCA analysis was performed using the built-in R function “prcomp()”.

Differential gene expression analysis

Both mRNA raw counts and protein pseudocounts were normalized by implementing the RUV (remove

unwanted variation) approach. Differential gene expression analysis was carried out using the two-way ANOVA method with the built-in R function “anova()”. Two variables, time (5 different time points) and type (GK and WST), were taken into consideration, and genes with FDRs (adjusted p value with “Benjamini & Hochberg” method) of less than 0.01 were considered differentially expressed genes.

Gene ontology (GO) enrichment analysis

The NCBI ENTREZ gene ID list was processed in batches using the DAVID Web Service API Perl Client to identify enriched Gene Ontology categories. Rattus norvegicus was set as the background. KEGG category terms with FDRs of less than 0.05 were considered significant hits.

Dynamic gene expression patterns

To investigate time course patterns of gene expression at both the mRNA and protein levels, we first calculated the log2 ratios of all differentially expressed mRNAs and proteins. To maintain the expression scale, we created ten series of values, including five ‘0’ values (4, 6, 8, 16, and 24 weeks, representing expression values in WST rats) and the log2 ratios of GK/WST (4, 6, 8, 16, 24 weeks, representing expression values in GK rats), and then performed data scaling via Z-transformation for the following clustering analysis.

Different expression patterns of mRNAs and proteins were clustered using the built-in R function “kmeans()”. The optimal results for clustering analysis were obtained by iteratively performing k-means clustering with a range of cluster numbers, followed by DAVID functional annotation analysis to assess their biological relevance.

KEGG signaling pathway analysis

The Generally Applicable Gene-set Enrichment (GAGE) package in R was utilized to perform pathway analysis. GAGE was generally applicable to gene expression datasets with different sample sizes and experimental designs. The KEGG Pathway Database was used as a reference. Pathways with q-values of less than 0.05 were considered significantly different between GK and WST rats.

Immunohistochemistry, antibodies and microscopy

Pancreases from 4- to 24-week-old rats were fixed in 4% paraformaldehyde and were embedded in paraffin. The primary (i) and secondary (ii) antibodies (Ab) used in this study were (i) guinea pig anti-insulin (prepared by our lab, 1:200), mouse anti-glucagon (Abcam, ab10988, 1:200), and rabbit anti-Ki67 (Abcam, ab15580, 1:1000 dilution) and (ii) goat anti-guinea pig IgG Rhodamine (Jackson, 1:200), donkey anti-rabbit IgG Alexa Fluro 488 (Invitrogen, 1:200), donkey anti-mouse IgG DyLight 488 (Jackson, 1:200). Fluorescent images were acquired with a confocal laser scanning microscope (Olympus, FV 1200) operated in multitrack mode using the following objective: UPLSAPO40X2 40×/0.95. The microscope system was drived by Olymbus Fluoview Ver.2.0a software.

1. Lei, L., Liu, Q., Liu, S., Huan, Y., Sun, S., Chen, Z., Li, L., Feng, Z., Li, Y., and Shen, Z. (2015) Antidiabetic potential of a novel dual-target activator of glucokinase and peroxisome proliferator activated receptor-gamma. *Metabolism: clinical and experimental* **64**, 1250-1261
2. Avall, K., Ali, Y., Leibiger, I. B., Leibiger, B., Moede, T., Paschen, M., Dicker, A., Dare, E., Kohler, M., Ilegems, E., Abdulreda, M. H., Graham, M., Crooke, R. M., Tay, V. S., Refai, E., Nilsson, S. K., Jacob, S., Selander, L., Berggren, P. O., and Juntti-Berggren, L. (2015) Apolipoprotein CIII links islet insulin resistance to beta-cell failure in diabetes. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E2611-2619
3. Fox-Walsh, K., Davis-Turak, J., Zhou, Y., Li, H., and Fu, X. D. (2011) A multiplex RNA-seq strategy to profile poly(A+) RNA: application to analysis of transcription response and 3' end formation. *Genomics* **98**, 266-271
4. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**, R25
5. Karolchik, D., Hinrichs, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler, D., and Kent, W. J. (2004) The UCSC Table Browser data retrieval tool. *Nucleic acids research* **32**, D493-496