UCSF
UC San Francisco Electronic Theses and Dissertations

Title

Investigating the genetic factors underlying the evolution of frugivory using comparative and functional genomics

Permalink

https://escholarship.org/uc/item/5hb22435

Author

Gordon, Wei Elizabeth

Publication Date

2023

Supplemental Material

https://escholarship.org/uc/item/5hb22435#supplemental

Peer reviewed|Thesis/dissertation

Investigating the genetic factors underlying the evolution of frugivory in mammals using comparative and functional genomics

by
Wei Gordon

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
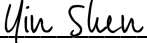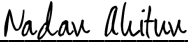DOCTOR OF PHILOSOPHY

in

Genetics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:

Yin Shen _____
AD18806FAF2442B...

Yin Shen
_____
Chair

DocuSigned by:

Jeff Wall _____
DocuSigned by: 4C8...

Jeff Wall
_____

Nadav Ahituv _____
77BA96E7DAE34F4...

Nadav Ahituv
_____

_____

_____

Committee Members

Dedication

I dedicate this dissertation to my Dad, Gary Gordon. My Dad has encouraged me through every

pursuit in life, worked hard to ensure that I could pursue everything I wanted, and provided

continuous advice to keep me pushing through graduate school.

Acknowledgements

First and foremost, I thank my P.I., Dr. Nadav Ahituv. This dissertation would not be possible without Dr. Ahituv's patience, support, and kindness. For the past several years, Dr. Ahituv gave me direction when I did not know where to go, space when I needed, thoughtfulness about both my mental and physical health, and excitement. Dr. Ahituv allowed me to pursue anything and everything that motivated me, whether it was outside teaching opportunities, lab morale events, my NSF GRFP, traveling to Belize, Grad Slam, or my next steps at Menlo College. Dr. Ahituv has been with me every step of the way, and I cannot thank him enough. Though I will miss meeting regularly, I know that we will remain in close contact. I will always seek his mentorship.

I thank my many collaborators. First and foremost, thank you to Seungbyn Baek. I would not be graduating now without you, and I hope we get to meet in person someday. Thank you to Dr. Martin Hemberg for helping move the bat multiome project forward. Thank you to Dr. Sharon Swartz and her graduate student Andrea Rummel at Brown University for providing my lab with bat tissues and letting us know about the Belize "Bat-a-thon." Thank you to Dr. Lisa Cooper at Northeast Ohio University and Dr. Tony Schountz at Colorado State University for providing my lab with bat tissues as well. Thank you to Dr. Nancy Simmons at the American Museum of Natural History and Dr. Melissa Ingala at the Smithsonian of Natural History for teaching Dr. Ahituv and I how to capture and release wild bats and for providing us with bat tissues. Thank you to Dr. Yien-Ming Kuo at UCSF for the consistent dissertation help and personal support.

Thank you to my qualifying exam committee: Dr. Kaveh Ashrafi (chair), Dr. Guo Huang, Dr. Yin Shen, and Dr. Jeff Wall. Thank you for meeting with me as much as I wanted and for providing me with constructive feedback to succeed in one of the first Zoom qualifying exams. I additionally thank Dr. Ashrafi and Dr. Huang for allowing me to rotate in their labs and construct

Contributions

Chapter 2: Diana Moreno Santillán with the Bat1K made bat genome alignments. Ilias Georgakopoulos-Soares, PhD created a pairwise genome alignment script. Yien-Ming Kuo, PhD sectioned mouse embryos for LacZ staining analyses. Aki Ushiki, PhD performed injections for mouse genomic sequence swapping and assisted with Southern blots. Rory Sheng and Liz Murray performed Southern blots. Mai Nobuhara is conducting phenotyping experiments with the mouse knock-ins and is co-author for this future article.

Chapter 3: Lisa Cooper, PhD and Alex Galazyuk, PhD provided *E. fuscus* tissues, and Tony Schountz, PhD, Nancy Simmons, PhD, and Melissa Ingala, PhD provided *A. jamaicensis* tissues. Hai Ngyuen, PhD helped perform the multiome experiments. Ilias Georgakopoulos-Soares, PhD performed transcription factor analyses. Yien-Ming Kuo, PhD sectioned bat tissues for immunofluorescence experiments. Seungbyn Baek conducted cell composition analyses, gene expression analyses, made figures, and is co-author for this future article. Martin Hemberg, PhD was co-corresponding author.

Citation for Chapter 3:

Gordon, Wei E., Seungbyn Baek, Hai P. Nguyen, Yien-Ming Kuo, Rachael Bradley, Alex Galazyuk, Insuk Lee, et al. 2023. "Integrative Single-Cell Characterization of Frugivory Adaptations in the Bat Kidney and Pancreas." *bioRxiv*. https://doi.org/10.1101/2023.02.12.528204.

Investigating the genetic factors underlying the evolution of frugivory in mammals using comparative and functional genomics

Wei Elizabeth Gordon

Abstract

Specialization on fruit, or frugivory, is a dietary adaptation that has evolved multiple times in primates and bats. While there are some genes that have been linked to this dietary specialization, the genetic factors, particularly gene regulatory factors, underlying mammalian frugivory are largely unknown. In this work, I utilized a combination of comparative and functional genomics to identify and functionally examine genetic factors for frugivorous adaptation in mammals. First, I used accelerated regions analysis to identify regions of the genome that are highly conserved across mammals but have significantly changed in fruit-eating bats and primates, termed frugivorous accelerated regions (FARs), and tested selected FARs *in vitro* and *in vivo*. I found that FARs drive transcriptional activity in luciferase reporter assays and mouse transgenic reporter assays, and we are now examining FAR phenotypes from mouse genomic sequence swapping. Next, I conducted the first combined single-cell RNA- and ATAC-seq on fasted and fed frugivorous and insectivorous bat kidneys and pancreases. We characterized cell composition, gene expression, and gene regulation in these heterogeneous tissues and revealed many novel cellular and molecular traits in frugivorous bats. Importantly, we found frugivorous bats share many traits with human diabetes and demonstrate how our integrative single-cell profilings can be used for therapeutic purposes.

Table of Contents

List of Figures

List of Tables

List of Supplemental Files

Supplemental File 2.1: FARs for each dataset.

Supplemental File 3.1: Differentially expressed genes between fasted and fed states in kidney for each species.

Supplemental File 3.2: Differentially expressed genes between fasted and fed states in pancreas for each species.

Supplemental File 3.3: Differentially expressed genes between species in the kidney.

Supplemental File 3.4: Gene enrichment analyses of differentially expressed genes between species in the kidney.

Supplemental File 3.5: Cell-type-specific ATAC peaks in the insectivore kidney.

Supplemental File 3.6: Cell-type-specific ATAC peaks in the frugivore kidney.

Supplemental File 3.7: T1 and T2 diabetes-associated SNP list in *hg38* coordinates.

Supplemental File 3.8: T1 and T2 diabetes-associated SNP overlaps with bat kidney scATAC-seq peaks in *hg38* coordinates.

Supplemental File 3.9: Differentially expressed genes between species in the pancreas.

Supplemental File 3.10: Gene enrichment analyses of differentially expressed genes between species in the pancreas.

Supplemental File 3.11: Cell-type-specific ATAC peaks in insectivore pancreas.

Supplemental File 3.12: Cell-type-specific ATAC peaks in frugivore pancreas.

Supplemental File 3.13: T1 and T2 diabetes-associated SNP overlaps with bat pancreas scATAC-seq peaks in *hg38* coordinates.

Supplemental File 3.14: Orthogroup IDs from Orthofinder for re-naming bat genome features.

Chapter 1: Introduction

1.1 Introduction

The ability to eat fruit provides a number of advantages to mammals. Fruit is a quick energy source which is rich in easily digestible carbohydrates. Fruit also offers quick hydration as it is extremely high in water content. Furthermore, fruits are loaded with fiber, vitamins like vitamin A, minerals like potassium, and antioxidants. However, fruits are low in protein, fat, sodium, and calcium. Within mammals, fruit-eating is observed in carnivores like wolves, foxes, and bears, in ungulates like deer and boar, in primates like macaques and orangutans, and in bats (*1–5*). In particular, specialization on fruit, or frugivory, evolved multiple times in both primate and bat orders (*4*, *6*, *7*). In bats, which have the broadest range of dietary specializations known to mammals, frugivory is the only dietary specialization to evolve independently in two families: Pteropodidae (Old World Fruit Bats; OWFBs) and Phyllostomidae (New World Fruit Bats; NWFBs) (*3*). The common ancestor to all placental mammals was an insect-eater (insectivore) as was the common ancestor of the primate and bat orders (*8*, *9*). Although shifting to fruit specialization was likely advantageous for many mammals, as frugivory is linked with larger brain size in both primates and bats (*4*, *10*) and more social interactions (*11–13*), the dietary shift from insectivory to frugivory in mammals is not understood at a molecular level.

Adaptations to eat fruit include metabolic, morphological, and sensory adaptations. Metabolic adaptations encompass gene losses (*7*, *14*, *15*), such as uricase, an enzyme lost specifically in primates which likely contributes to increased fat storage from fructose of ephemeral fruit (*16*). Despite the high sugar amount fruit bats consume, fruit bats are able to assimilate sugar rapidly and lower their blood sugar faster than non-fruit bats (*3*, *17–20*), which may be due to evolutionary selection in glucose transporters (*3*, *21*) and/or loss of hormone inhibitors (*15*). Furthermore, fruit bats have increased intestinal sucrase and maltase activities, indicators of

2

abilities to assimilate plant sugars, accompanied by decreased intestinal trehalase activity, an indicator of reduced need to digest insect sugar (*22*). Morphological adaptations encompass changes in gastrointestinal structures, such as a longer small intestine relative to body length and greater microvillous surface area in fruit bats compared to non-fruit bats and a homogenous gut structure (similar surface area for stomach, large intestine and small intestine) in frugivorous primates (*23–27*). Frugivorous bats and primates also have specialized craniofacial features (*26–32*) and tongue morphologies (*33*, *34*). There have also been several changes to kidney morphology in fruit bats relative to non-fruit bats, including an increased renal cortex and a decreased renal medulla (*22*, *35*). Sensory adaptations encompass color vision (*14*, *36–40*) and olfactory senses (*41–45*).

While there have been some genetic investigations on understanding the evolution of frugivory in mammals, most genetic candidates were identified through biased targeting of metabolic genes (*21*, *46–52*). For example, the rate-limiting enzyme in gluconeogenesis, phosphoenolpyruvate carboxykinase 1 (*PCK1*), was compared across a variety of bat species due to its role in glucose homeostasis and was found to have four parallel amino acid substitutions in both OWFBs and NWFBs (*51*). Parallel evolution was similarly observed for both glycogen synthase enzymes in skeletal muscle (*GYS1*) and the liver (*GYS2*) (*48*, *49*). More recent work has been done using genome-wide approaches. To identify frugivorous-associated genes, numerous tissues including tissues mixed together were used to generate transcriptomes which were compared for 58 bats across all diets (*53*), liver transcriptomes were generated and compared across four insect bats and one fruit bat to focus specifically on differences in a key metabolic tissue (*54*), and comparative genomics was used to identify potential frugivorous-associated gene losses, finding some renal transporters and metabolic genes to potentially be associated with this trait (*15*). Using comparative genomics on new bat

3

genome assemblies also provided evidence for frugivorous convergent evolution through olfactory gene family composition, gene loss of digestive enzymes and bitter taste receptors, and convergent amino acid substitutions in metabolic genes (*7*).

In addition to genes, the elements that regulate their transcription can also be modified due to dietary selective pressures. These regulatory elements include promoters (immediately upstream of a gene and promote transcription), enhancers (activate the promoter and are tissue specific), silencers (silence the promoter) and insulators (prevent interactions between adjacent elements). Regulatory sequences can control tissue development, sensory perception, and metabolic flux (*55–57*). For example, in humans, two variants in the regulatory region of the lactase-encoding gene (*LCT*) are associated with lactase persistence and enhance its transcription beyond infancy (*55*). In bats, an 11 base pair (bp) deletion in the promoter of the Solute Carrier Family 2 Member 2 (*SLC2A2)* gene, encoding glucose transporter 2 (GLUT2), in all examined fruit bats correlated with reported differences in liver *SLC2A2* expression between frugivorous and insectivorous bats (*3*). While loss of this sequence was predicted to disrupt binding of transcriptional repressor *ZNF354C (3)*, no functional studies were done.

Here, I set out to identify DNA sequences that could be responsible for the evolution of frugivory in mammals. First, I utilized a comparative genomics approach to identify frugivorous accelerated regions (FARs) in both bats and primates. We tested FARs for adaptive function using reporter assays and mouse genomic sequence swapping. Next, we utilized integrative single-cell RNA-seq and ATAC-seq on adult insect- (*E. fuscus*) and fruit- eating (*A. jamaicensis*) bats to identify cell populations, genes and regulatory element differences that could be associated with frugivory adaptations in the kidney and pancreas.

Chapter 2: Accelerated regions analysis identifies gene regulatory sequences in fruit-eating bats and primates

2.1: Abstract

Here, we set out to identify DNA sequences that could be responsible for the evolution of frugivory in mammals. First, we tested the 11 bp deletion in the proximal promoter of *SLC2A2* in frugivorous bats for predicted gene regulatory activity using luciferase reporter assays and found no increased reporter gene expression from frugivore *SLC2A2* promoters compared to non-frugivores. To expand gene regulatory candidates involved in frugivory evolution, we utilized a comparative genomics approach to identify frugivorous accelerated regions (FARs), regions of the genome that evolved rapidly in frugivores but are highly conserved across mammals, in both bats and primates. We identified thousands of FARs in each mammalian order, including one FAR shared between frugivorous primates and bats within 1 Megabase (Mb) of the H2.0 like homeobox (*HLX*) gene. We characterize this shared FAR as well as FARs within 1 Mb of the fibroblast growth factor 9 (*FGF9*) gene utilizing luciferase reporter assays, finding that FARs drive reporter expression. We then tested these FARs for enhancer activity using mouse transgenic assays and swapping endogenous mouse loci with FARs. Mouse FAR knock-ins (KIs) are being analyzed for intestinal, kidney, and metabolic phenotypes. Together, these results demonstrate that our comparative genomics approach identifies gene regulatory regions associated with frugivory and unveils developmental features necessary for frugivorous adaptation.

2.2: Methods

Luciferase assays for *SLC2A2* promoters:

To construct reporter plasmids containing bat, mouse, and primate genomic regions, DNA sequences were synthesized with either pGL4.23[minP] (Promega) (**Supplementary Table 2.1**) or pGL4.11b (Promega) vector primers for cloning (**Supplementary Table 2.2**). All sequences were validated by Sanger sequencing. Human liver cells (HEPG2) were obtained from ATCC and maintained with EMEM (ATCC 30-2003) supplemented with 10% FBS (ATCC 30-2020) and 1% penicillin-streptomycin (ThermoFisher Scientific 15140122). 90 ng of reporter plasmids and 10 ng of the Renilla luciferase plasmid pGL4.73 (Promega), to correct for transfection efficiency, were co-transfected in each well of a 96-well plate using X-tremeGENE HP DNA Transfection Reagent (Sigma-Aldrich) according to the manufacturers' protocol. 48 hours post-transfection, cells were harvested, and reporter activity was measured using the Dual-luciferase reporter assay system (Promega) according to the manufacturers' protocol. Reporter activity was quantified using the Glomax 96-well plate luminometer (Promega). Relative firefly/Renilla luciferase values were determined, and results of luciferase assays represent mean ± SEM derived from 2 independent experiments of quadruplet measurements (n = 8). *P* values were calculated using unpaired Student's t-test.

Detection of FARs:

To identify FARs, we utilized the open source software package PhyloAcc (*58*), which employs a Bayesian method to test conserved genomic regions for shifts in DNA substitution patterns in multiple lineages across a phylogeny. The input to PhyloAcc is a multiple sequence alignment file of conserved genomic regions, a phylogenetic model of the species in the alignment, and a list of locations of each conserved genomic region within the multiple sequence alignment file.

For detection of FARs in bats, we used two datasets: 1) publicly available whole genome

alignments and conserved regions of 120 mammals (*59*), which included 11 bats and is

hereafter referred to as the "*120 mammals dataset*"; 2) whole genome alignments of 37 bat

species we created with novel high quality genomes from the *Bat1K* (*60*) and hereafter referred

to as the "*Bat1K dataset*" (**Supplementary Table 2.3**). For the *Bat1K dataset*, we generated, in

collaboration with the *Bat1K*, whole genome alignments of 37 bat species with ancestral

outgroup Florida manatee (*triMan1*; GenBank: GCA_000243295.1) to root the phylogenetic tree.

Briefly, pairwise alignments for each species to reference genome *R. aegyptiacus*

(*mRouAeg1.p*) were computed with Lastz (*61*) on tandem repeat-masked genome assemblies,

created with trfBig from UCSC Genome Bioinformatics Group's suite of biological analysis tools

(*62*), before whole genome alignments were assembled with Multiz (*63*). Using the Phast

package (*64*), a nonconserved phylogenetic model was estimated with phyloFit from four-fold

degenerate sites genome-wide. We then adjusted the background frequencies of our model with

modFreqs from UCSC Genome Bioinformatics Group's suite of biological analysis tools (*62*).

Conserved regions were then called by phastCons from the Phast package. As it was previously

observed that known functional elements were often tiled with multiple conserved regions

separated by short gaps (*65*), we merged adjacent elements until the ratio of the distance

between the elements merged over the total length of the region was ≤ .01. A total of 1,070,648

conserved genomic regions 50 bp or greater in length were extracted.

For detection of FARs in frugivorous primates, we created whole genome alignments by utilizing

available pairwise alignments of primate genomes to human (*hg38*) as well as two ancestral

outgroups mouse (*mm10*) and cat (*felCat9*). To add more frugivorous primate genomes, we

downloaded the newest available assemblies from NCBI Genome (**Supplementary Table 2.4**)

and created pairwise alignments to hg38 using Lastz (K = 2400, L = 3000, Y = 9400, H = 2000, and the lastz default scoring matrix) (*61*). We then used all human pairwise alignments to create whole genome alignments with Multiz (*63*). A nonconserved phylogenetic model was created and conserved genomic regions were extracted as described above for a total of 894,539 conserved regions.

For FAR detection in the public *120 mammals dataset*, conserved regions were tested for significant sequence changes in the three available fruit bat genomes, all of which belong to the Old World Fruit Bat (OWFB) family: Egyptian rousette (*rouAeg1*), large flying fox (*HLpteVam2*) and the black-flying fox (*pteAle1*). For fAR detection in the *Bat1K dataset*, conserved regions were tested for significant sequence changes in 9 fruit bats spanning both frugivorous bat families: OWFBs *R. aegyptiacus*, *M. sobrinus, E. helvum, P. alecto*, *P. vampyrus* and New World Fruit Bats (NWFBs) *C. perspicillata*, *G. soricina*, *A. jamaicensis, A. caudifer*. For FAR detection in primates, primates were classified into diet categories from public data (*4*, *66*), and only highly frugivorous primates (>60% fruit-eating) for which genomes were available for [orangutan (*ponAbe3*), Northern white-cheeked gibbon (*nomLeu4*), silvery gibbon *(HMol_V2/hylMol1)*, Sclater's lemur (*eulFla1*), black lemur (*eulMac1*), Ma's night monkey (*aotNan1*), black-handed spider monkey *(AteGeo_v1_BIUU/ateGeo1*), pig-tailed macaque (*macNem1*) and bonobo (*panPan2*)] were selected as target species for FAR detection. We used the following PhyloAcc options for both analyses: BURNIN=500, MCMC=1000, CHAIN=1, CONSERVE_PROP=0.8, VERBOSE=0. We called accelerated regions with Bayes factors log-BF1>20 and log-BF2>0 whereby the false discovery rate (FDR) drops below 5% (*58*). There were 1,363 FARs in the 120 mammals dataset, 1,440 FARs in the *Bat1K* dataset, and 4,634 FARs in the primates dataset (**Supplemental File 2.1**). Acceleration pattern and sequence alignment plots were created using R functions from Hu et al. 2019 (*67*).

Associating FARs with human and mouse phenotypes:

FARs were associated with genes and human and mouse phenotypes via Genomic Regions

Enrichment of Annotations Tool (GREAT) (*68*). With default settings, GREAT assigns each gene

a basal regulatory domain 5 kb upstream and 1 kb downstream of the TSS. GREAT then

extends this domain in both directions to the nearest gene's basal regulatory domain but no

more than 1000 kb in one direction. Input regions that fall within the gene's regulatory domain

are associated with that gene. For the *Bat1K* dataset, only 740 FARs that lifted over from

mRouAeg1.p to *hg38* using liftOver from UCSC Genome Bioinformatics Group's suite of

biological analysis tools (*62*) could be inputted into GREAT.

Luciferase assays for FARs:

To construct reporter plasmids containing bat, mouse, and primate genomic regions, DNA

sequences were synthesized into the pGL4.23 (Promega) vector for candidate enhancers or

amplified from genomic DNA samples before being cloned into pGL4.23 (**Supplementary Table

2.5**). All sequences were validated by Sanger sequencing. Human kidney cells (HEK293T) were

obtained from ATCC and maintained with DMEM high glucose (Sigma D5796) supplemented

with 10% fetal bovine serum (FBS; ATCC 30-2020), 1% L-glutamine (Corning 25-005-CI) and

1% penicillin-streptomycin (ThermoFisher Scientific 15140122). Rat small intestine cells (IEC-6)

were obtained from the UCSF Cell and Genome Engineering Core and maintained with DMEM

(ATCC 30-2002) supplemented with 0.1 Unit/ml human insulin, 10% fetal bovine serum (FBS;

ATCC 30-2020) and 1% penicillin-streptomycin (ThermoFisher Scientific 15140122). Human

foreskin fibroblasts were obtained from ATCC and maintained with EMEM (ATCC 30-2003)

supplemented with 10% FBS (ATCC 30-2020) and 1% penicillin-streptomycin (ThermoFisher

Scientific 15140122). 240 ng of reporter plasmids and 60 ng of the Renilla luciferase plasmid

pGL4.73 (Promega), to correct for transfection efficiency, were co-transfected into each cell line

in each well of a 48-well plate using X-tremeGENE HP DNA Transfection Reagent (Sigma-Aldrich) according to the manufacturers' protocol. 48 hours post-transfection, cells were harvested, and reporter activity was measured using the Dual-luciferase reporter assay system (Promega) according to the manufacturers' protocol. Reporter activity was quantified using the Glomax 96-well plate luminometer (Promega). Relative firefly/Renilla luciferase values were determined, and results of luciferase assays represent mean ± SEM derived from 2 independent experiments of triplet measurements (n = 6). *P* values were calculated using Student's t-test.

Enriched TFBM analyses

Relatively enriched TFBMs were estimated from AME software within the MEME suite (*69*) using default parameters and the JASPAR 2022 core redundant dataset for vertebrates (*70*). FARs were input as primary sequences, and non-frugivorous sequences homologous to the FARs were input as control sequences.

Mouse Transgenic Enhancer Assays:

PCR was carried out on *R. aegyptiacus* DNA using primers that were designed to amplify candidate enhancer peak sequence with additional 100–500 bp outside of the predicted region (**Supplementary Table 2.5**). Synthesis was carried out for *P. vampyrus* and *E. flavifrons* DNA. PCR product and synthetic sequences were cloned into the Hsp68-LacZ vector (Kothary et al. 1988) and verified with Sanger sequencing. All transgenic mice were generated by Cyagen Biosciences using standard procedures (Nagy et al. 2002), and harvested and stained for LacZ expression at E10.5, E12.5, or E14.5 as previously described (*71*). Embryos were sectioned on sagittal planes at 14uM. Sectioned embryos were counterstained with Nuclear Fast Red Solution (Electron Microscopy Sciences, #26078-05). Pictures were obtained using an M165FC stereo microscope and a DFC500 12-megapixel camera (Leica). To be designated as an

enhancer, we required consistent spatial expression patterns present in at least three embryos.

Mouse genomic sequence swapping with FARs:

To carry out mouse genomic sequence swaps, we used Precise Integration into Target Chromosome (PITCh) CRISPR-based knockin technology (*72*). Briefly, two sgRNA were designed to target the 5' and 3' ends of target regions that were swapped in mice. A donor plasmid was generated that contains the bat sequence that was swapped in along with ~1kb mouse homology arms on either side flanked by a target site for one of the two sgRNA (**Supplementary Figure 2.9, Supplementary Table 2.7-2.7**). Cas9 nuclease, sgRNAs and the donor plasmid were injected into the pronucleus of mouse fertilized eggs via the Gladstone Transgenic Gene Targeting Core. Subsequent embryos were genotyped via PCR, Sanger sequencing and Southern blots and both heterozygous and homozygous mice maintained. Genotyping primers can be found in **Supplementary Table 2.9**. Southern blot analysis was performed after cleavage with *StyI-HF* enzyme for 153011-13 knock-in and *BstXI* enzyme for 54670/3752 knock-in as previously described (*73*) (**Supplementary Figure 2.9, Supplementary Table 2.10)**. Mouse work was approved by the UCSF Institutional Animal Care and Use Committee (protocol number AN100466) and was conducted in accordance with AALAC and NIH guidelines.

2.3: Results

Frugivore *SLC2A2* promoters do not promote transcriptional activity in luciferase assays

To test whether the 11 bp deletion in the proximal promoter of *SLC2A2* in frugivorous bats

conferred increased transcriptional activity reported in fruit bat livers (*3*), we conducted

luciferase reporter assays. We synthesized two OWFB, two NWFB, and two insect bat proximal

promoter sequences, sequenced by Meng et al. 2016 (~300 bp) (**Supplementary Table 2.1**),

and cloned them into reporter vector pGL4.23[minP] (Promega), which contains a minimal

promoter followed by a luciferase reporter gene. We also synthesized larger *SLC2A2* promoter

sequences (~800 bp), which capture the 11 bp deletion region and end before transcription start

sites (TSSs), from frugivorous bats and primates and non-frugivorous mammals

(**Supplementary Table 2.2**) to examine for reporter activity in the pGL4.11b vector (Promega),

which contains only the luciferase reporter gene. Reporter constructs were co-transfected into

human liver cells (HepG2) with the *Renilla* luciferase reporter gene vector

(pGL4.73[hRluc/SV40]; Promega) to normalize for transfection efficiency. Transfected cell

extracts were surveyed for relative luciferase activity 48 hours after transfection. In both reporter

assay designs, only promoter sequences from insectivorous bat genus *Hipposideros* drove

reporter activity (**Figure 2.1-2.2**).

Computational analysis identifies thousands of bat and primate FARs

To identify gene regulatory sequences that could be responsible for the evolution of frugivory in

mammals in an unbiased and comprehensive manner, we implemented a comparative

genomics approach to detect frugivorous-accelerated regions (FARs), sequences in the

mammalian genome that are highly conserved across evolution but have significantly changed

in fruit-eating species. We analyzed three different datasets (**Figure 2.3A**). For detecting bat

13

FARs, we utilized whole genome alignments of 120 mammals (*59*) which include 11 bats, 3 of which are Old World Fruit Bats (OWFBs), with human as the reference genome (*hg38*). This dataset will hereafter be referred to as the "*120 mammals* dataset." We also analyzed whole genome alignments we created across 35 bats, including both OWFBs and New World Fruit Bats (NWFBs), using the newest available bat genomes with the help of the *Bat1K* (Table S#) and reference genome *Rousettus aegyptiacus* (*mRouAeg1.p*). We term this dataset the "*Bat1K* dataset." For detecting frugivorous primate ARs, we aligned 27 of the newest available primate genomes (**Supplementary Table 2.4**) with reference genome *hg38*. We term this dataset the "*primates* dataset." Conserved regions in each dataset were extracted using phastCons (*64*). 687,261 conserved regions were detected from the *120 mammals* dataset; 1,070,648 regions from the *Bat1k* dataset; and 894,539 regions from the *primates* dataset (**Figure 2.3A**). These regions were then tested for statistically significant changes in nucleotides substitution rates in fruit-eaters with PhyloAcc (*58*). Target fruit bats included all OWFBs (*E. helvum, M. sobrinus, P. alecto*, *P. vampyrus, R. aegyptiacus*) and NWFBs (*C. perspicillata*, *G. soricina*, *A. jamaicensis* and *A. caudifer*). Target frugivorous primates included Sumatran orangutan, Northern white-cheeked gibbon*,* silvery gibbon, Sclater's lemur, black lemur, Ma's night monkey*,* black-handed spider monkey, pig-tailed macaque and bonobo. Accelerated regions were called with Bayes factors log-BF1>20 and log-BF2>0 [false discovery rate (FDR) < .05]. For the fruit bats, 1,363 FARs were detected from the *120 mammals* dataset, 1,440 FARs were detected from the *Bat1K* dataset, and 4,634 FARs were detected from the *primates* dataset (**Figure 2.3A; Supplemental File 2.1**). Using multiple datasets and targeting multiple fruit-eating species increases sensitivity of detection sequences that could be responsible for frugivory adaptation in mammals.

To determine what genes and phenotypes bat FARs are associated with, we input these regions into the Genomic Regions Enrichment of Annotations Tool (GREAT) (*68*) (see **Methods**). As GREAT cannot be used with custom reference genomes, we extracted 740 FARs from the *Bat1K* dataset from *mRouAeg1.p* to *hg38* coordinates (**Figure 2.1A**) (see **Methods**). More than 97% of all bat FARs input into GREAT had predicted associations with one or two genes (**Supplementary Figure 2.1**). From the *120 mammals* dataset, 10% region-gene associations were within 5 kilobases (kb) of the transcription start site (TSS), 30% region-gene associations were within 5 to 50 kb of the TSS, and 60% region-gene associations > 50 kb from the TSS. From the *Bat1K* dataset, 14% region-gene associations were within 5 kb of the TSS, 23% region-gene associations were within 5 to 50 kb of the TSS, and 63% region-gene associations > 50 kb from the TSS. Both bat datasets had strong enrichments for gene ontology (GO) Biological Process terms related to kidney and nephron development, accompanied by mouse phenotype terms for the general urinary system (**Figure 2.3B**), which is likely indicative of the observed morphological differences in kidneys between bats of frugivorous and non-frugivorous diets (*22*, *35*). Ear morphogenesis was also a top GO Biological Process term for the region-gene associations from the *120 mammals* dataset, and many Mouse Phenotype Single KO terms related to ear development were highly ranked in both bat datasets (**Figure 2.3B**), likely indicative of the loss of laryngeal echolocation in OWFBs (*74*). Interestingly, the *Bat1K* dataset returned several Mouse Phenotype Single KO terms related to liver development, as well as "decreased circulating insulin level" (**Figure 2.3B**), which is indicative of gene changes associated with the high sensitivity to insulin fruit bats have (*75*). Together, our comparative genomics approach identifies bat genomic regions that are associated with human and mouse terms suggestive of frugivorous adaptations.

To determine what genes and phenotypes primate FARs are associated with, we input these regions into GREAT (*68*) (see **Methods**). 97% of which had predicted associations with one or more genes (**Supplementary Figure 2.1**). 12% region-gene associations were within 5 kb of the TSS, 32% were within 5 kb of the TSS, and 56% region-gene associations > 50 kb from the TSS. Intriguingly, the most enriched GO Biological Process term was "endoplasmic reticulum (ER) tubular network organization" (**Figure 2.3B**). ER tubules may be the main sites for lipid synthesis and calcium cation regulation (*76*, *77*), fitting with the high carbohydrate and low calcium content of fruit. "Response to vitamin A" was also an enriched term, and most dietary vitamin A is found in fruits and leafy greens (*78*). Many mouse phenotype terms were related to bone development with an emphasis on jaw and cranium development (**Figure 2.3B**), which is consistent with the craniofacial differences known between frugivores and non-frugivores (*26–32*). Investigating the mouse phenotype terms related to gut development like "abnormal liver bud morphology," we found a cluster of 9 FARs within 500 kb of H2.0 Like Homeobox (*HLX*) gene (**Figure 2.4A**), an essential transcription factor for embryonic gut development (*79*). Strikingly, one of these FARs intersected a FAR from each of the bat datasets (**Figure 2.4A**). In summary, FARs analysis uncovers many unbiased genomic candidates in the evolution of frugivory in mammals.

FARs drive differential enhancer activity *in vitro*

To determine whether FARs are functional regulatory elements, we selected a few FARs to test for adaptive function. We focused on the FARs that 1) reside within 1 Mb of known developmental genes, 2) had the strongest acceleration patterns, and 3) overlapped regulatory marks in the human and mouse genome. In the *120 mammals* dataset, a cluster of 9 FARs resided within 1 Mb of the fibroblast growth factor 9 (*FGF9*) gene (**Figure 2.5B**). This gene plays a key role in embryonic development of many organs, including small intestine and kidney,

which have observable differences between frugivores and non-frugivores (*22–27*, *35*). Mouse *Fgf9* homozygous knockouts have a shortened small intestine (*80*), and *Fgf9* heterozygous knockouts crossed with *Fgf20* homozygous knockouts have nephron progenitor depletion and renal dysplasia (*81*). Of the 9 FARs around *FGF9*, we selected the 5 FARs with the greatest acceleration patterns (**Supplementary Figure 2.2-2.3**) to test. As *120 mammals* FARs 153011, 153012 and 153013 were separated by < 100 bp, we combined them into one sequence for functional analysis which we term FAR_153011-13. We also chose to examine the FAR downstream of *HLX* that intersected all three datasets (**Supplementary Figure 2.4-2.5**). *HLX* is known to be expressed in the visceral mesenchyme and drives embryonic expansion of the liver and gut, where frugivores are known to have morphological adaptations encompassing intestine length and surface area (*23–27*). Homozygous *Hlx* mouse knockouts show severe size reductions in liver and digestive tract (*79*).

To test *FGF9* and *HLX* FARs for adaptive function, we first examined transcriptional activity with luciferase reporter assays. Regions spanning each of the FAR candidates (1-3 kb) were either amplified from available bat genomic DNA (*HLrouAeg4*) or synthesized (*HLpteVam2*, *eulFla1*) (**Supplementary Table 2.5**) and cloned into the pGL4.23 vector (Promega), which contains a minimal promoter followed by a luciferase reporter gene. In addition, the regions corresponding to FARs in the human genome (*hg38*), mouse genome (*mm10*), and insect bat genomes (*minNat1*) were also amplified from genomic DNA and cloned into pGL4.23 (**Supplementary Table 2.5**). Reporter constructs for *120 mammals* FAR_152985, FAR_153011-13 and FAR_153033 and non-frugivorous mammals were transfected into human kidney cells (HEK293T) and rat small intestine cells (IEC-6), and reporter constructs for *120 mammals* FAR_54670/*Bat1K* FAR_3752/*primates* FAR_29615 and non-frugivorous mammals were transfected into human foreskin fibroblasts (BJ cells) and IEC-6 cells. In HEK293T cells, *120*

mammals FARs 152985, 153011-13 and 153033 all induced transcriptional activity (**Figure 2.5B**). FAR_153011-13 induced the most transcriptional activity, 753-fold higher and 94-fold higher than corresponding insect bat and mouse regions, respectively. In IEC-6 cells, all three FARs again induced transcriptional activity (**Figure 2.5C**). FAR_153011-13 induced the most transcriptional activity over empty vector control (4-fold). As FAR_153011-13 induced the most transcriptional activity in both HEK293T and IEC-6 cells, we conducted new luciferase assays capturing the FAR within a smaller genomic region (1 kb), termed FAR_153011-13_short, to determine if we still observe transcriptional activity from a smaller sequence. FAR_153011-13_short still induced transcriptional activity in both HEK293T and IEC-6 cells. FAR_153011-13_short resulted in 4-fold higher transcriptional activity relative to the corresponding shorter mouse sequence in HEK293T cells (**Figure 2.5D**) and 7-fold higher transcriptional activity relative to the corresponding shorter mouse sequence in IEC-6 cells (**Figure 2.5E**). In BJ cells, *120 mammals* FAR_54670/*Bat1K* FAR_3752 and *primates* FAR_29615 each resulted in a 6-fold induction of transcriptional activity as compared to empty vector control (**Figure 2.5F**). The fruit bat region resulted in 6-fold greater luciferase activity relative to the respective insect bat region, and the lemur region resulted in 19-fold greater luciferase activity relative to the respective human region. Both fruit bat and lemur regions resulted in 11-fold greater luciferase activity relative to the respective mouse region. In IEC-6 cells, the lemur region resulted in 56-fold greater luciferase activity relative to the respective human region, 6-fold greater luciferase activity relative to the respective mouse region, and 4-fold greater luciferase activity relative to the empty vector control (**Figure 2.5G**). The fruit bat region did not induce transcriptional activity in IEC-6 cells. Combined, our data demonstrate that computational analysis identifies FARs that drive transcriptional activity *in vitro*.

To determine what regulatory factors in FARs might be driving increased reporter activity, we performed transcription factor binding motif (TFBM) enrichment analyses with AME (*69*) and identified motifs that are relatively enriched in FARs compared to non-frugivorous homologous sequences. FAR_153011-13_short sequences (N = 3), as compared to present, non-frugivorous sequences in *120 mammals* dataset (N = 114) (**Supplementary Figure 2.2**), are enriched for transcriptional activators, *E2F3*, *ZFX*, and *TFAP2A* **(Supplementary Figure 2.6, Supplementary Table 2.6)**. These transcription factors (TFs) are known to regulate the expression of genes involved in cell proliferation and play roles in the progression of many cancers, including colorectal and renal cancers (*82–87*). Notably, *TFAP2A* regulates differentiation of nephron precursor populations (*88–90*), and complete loss of this TF in mouse nephron progenitors results in absence of distal convoluted tubules (*88–90*), which are largely found in the renal cortex. Fruit bats have an increased renal cortex as compared to insectivorous bats (*22*), so FAR_153011-13_short may enhance *FGF9* transcription to increase the nephron progenitor pool in the developing renal cortex. *120 mammals* FAR_54670/*Bat1K* FAR_3752/*primates* FAR_29615 sequences (N = 6), as compared to present, non-frugivorous sequences in *120 mammals* dataset (N = 116) (**Supplementary Figure 2.4**), are enriched for *ZBTB7C* and *ZBTB7A* **(Supplementary Figure 2.7, Supplementary Table 2.6)**. Interestingly, *ZBTB7C* is a critical regulator of gluconeogenesis, having upregulated expression in the liver during mammalian fasting (*91*). *ZBTB7A* is a regulator of glycolysis as well as of hematopoietic lineage fate (*92*). *HLX* regulates hematopoietic differentiation and the metabolic state of hematopoietic cells (*93*), and *HLX* is a target of *ZBTB7A* from ENCODE ChIP-seq data (*94*, *95*). These TFBM enrichment analyses indicate that FARs are regulating their predicted targets and provide insight into regulatory element evolution in frugivorous adaptation.

19

FARs are functional enhancers in mouse.

To determine whether FARs are functional regulatory elements *in vivo*, we tested our selected

FARs for enhancer activity using a mouse transgenic assay. Regions spanning each of the FAR

candidates were either amplified from available bat genomic DNA (*HLrouAeg4*) or synthesized

(*HLpteVam2*, *eulFla1*) (**Supplementary Table 2.5**) and cloned into the Hsp68-LacZ vector,

which contains an Hsp68 minimal promoter followed by the LacZ reporter gene (*96*). Constructs

were injected into single-cell mouse embryos. Transgenic embryos were harvested at E10.5 for

fruit bat and primate FARs near *HLX* and E12.5 and E14.5 for fruit bat FAR near *FGF9*. These

stages were chosen from reported expression patterns of *Hlx* and *Fgf9 (79–81)*. *120 mammals*

FAR_54670/*Bat1K* FAR_3752 and (3/5 embryos) and *primates* FAR_29615 (3/6 embryos)

showed gut enhancer activity in our transgenic mouse assay (**Supplementary Figure 2.8**), but

FAR_153011-13_short did not show kidney or gut enhancer activity at E12.5 (3/3 embryos),

showing inconsistent expression patterns, or E14.5 (3/3 embryos), showing limb and facial

expression patterns, which is consistent with the known role *Fgf9* plays in skeletal development

(*97, 98*).

To determine whether FARs are functional regulatory elements *in vivo* and within the

appropriate genomic location, we tested our selected FARs for enhancer activity by swapping

the endogenous mouse locus with a FAR. *120 mammals* FAR_153011-13_short, hereinafter

referred to as FAR_153011-13, and *120 mammals* FAR_54670/*Bat1K* FAR_3752 were

synthesized with ~1000 bp mouse sequences that flank the endogenous mouse locus into the

pGEM-T-Easy vector (Table S#), along with a gRNA that was also synthesized at each end of

the flanking mouse sequences (**Supplementary Figure 2.9A**). Founder mice and germ line

transmission in F1 offspring with the desired knock-in were validated by PCR, Sanger

sequencing, and Southern blot (**Supplementary Figure 2.9B-C**) (see **Methods**). Mice are being

analyzed for frugivory-associated morphological phenotypes associated with the small intestine (length, microvillous surface area) and kidney (nephron progenitor population), as well as for physiological phenotypes (glucose and insulin tolerance).

2.4: Concluding Remarks

The *SLC2A2* promoter in fruit bats, which has an 11 bp deletion predicted to disrupt the binding of transcriptional repressor *ZNF354C (3)*, did not result in greater reporter activity than the *SLC2A2* promoter in insectivorous bats, as predicted for liver cells (*3*). Frugivorous primate *SLC2A2* promoters recapitulated the results from frugivorous bats. While HepG2 cells may not have provided the necessary cellular context to see transcriptional activity (may need bat liver cells and/or glucose stimulation), this work reinforces the need for functional testing of predicted regulatory elements. Future examination of this putative regulatory element should also include *ZNF354C* ChIP with insectivorous and frugivorous bat livers to determine whether this transcriptional regulator indeed binds this element and acts as a repressor.

Biased computational targeting of metabolically-relevant genes is both low-throughput and limited in adaptive scope. With increased availability of bat genomes and tissues, new genome-wide experiments are possible. Using comparative genomics, we identified thousands of sequences uniquely changed in frugivores across mammalian evolution, termed FARs, and we showed that FARs predicted to affect *HLX* and *FGF9* gene expression drive transcriptional activity *in vitro* and *in vivo*. A limitation of our investigation includes the qualities of genome assemblies and genome alignments used, as these impact FAR detection. However, we obtained the newest available genome assemblies to make primate and bat alignments, and we called FARs with stringent detection parameters (*58*). In addition, we only functionally examined FARs near two developmental genes. Future studies could examine all FARs in a massively parallel reporter assay (MPRA) (*57*), with multiple cell lines, including bat cells, and stimulation (ie. glucose), to identify more FARs with regulatory activity in an unbiased manner. Nonetheless,

we demonstrate that molecular adaptations to frugivory can be identified with our comparative genomics approach.

To our knowledge, we provide some of the first functional validations for frugivorous genetic candidates. Phenotyping analyses of FAR knock-in mice will reveal whether FARs are functional gene regulatory elements in an organismal context. Preliminary results indicate that both FARs tested in this study affect intestinal length, and *Fgf9*-associated FAR knock-ins have lower fasting blood glucose and improved glucose and insulin tolerance. *Fgf9* overexpression in the liver was recently shown to improve insulin sensitivity and glucose intolerance in diet-induced obese mice (*99*). Of note, we did not demonstrate that the examined FARs directly regulate *Fgf9* or *Hlx* in knock-in mice, and qPCR analyses will be done in the future. Moreover, TF ChIP-seq is needed to determine whether the enriched TFBMs estimated in the examined FARs are responsible for driving transcriptional activity. As frugivory presents many metabolic challenges for mammals, including high sugar, FARs have the potential to be used as a design guide for treatments such as *cis*-regulation therapy (CRT) (*100*) of metabolic diseases, like diabetes.

2.5: Figures & Tables
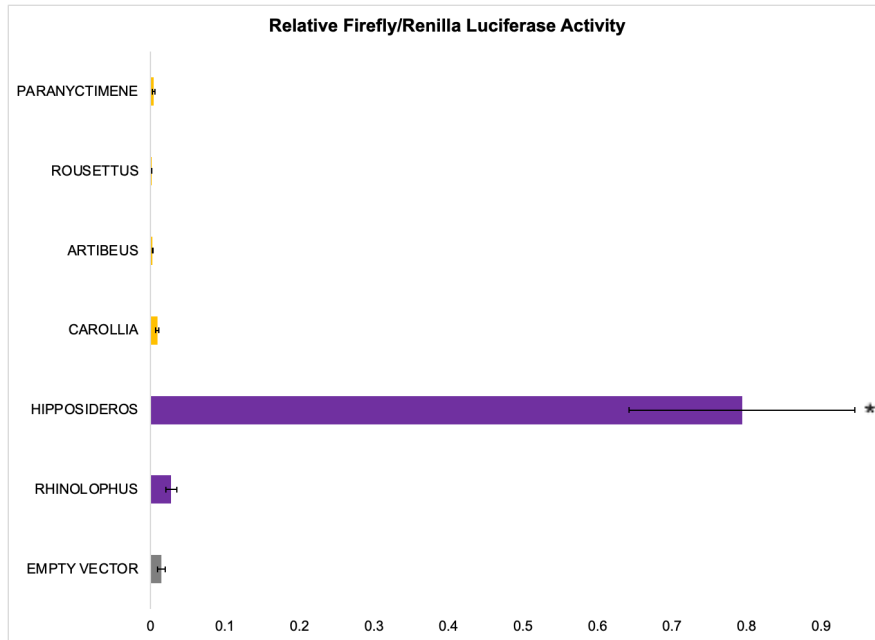
Figure 2.1: Luciferase assays in HepG2 with *SLC2A2* proximal promoters.
Frugivores are colored in yellow, insectivores are colored in purple, omnivores are colored in blue, and the gummivore is colored in green. Student's t-test vs. empty vector *p < .0001
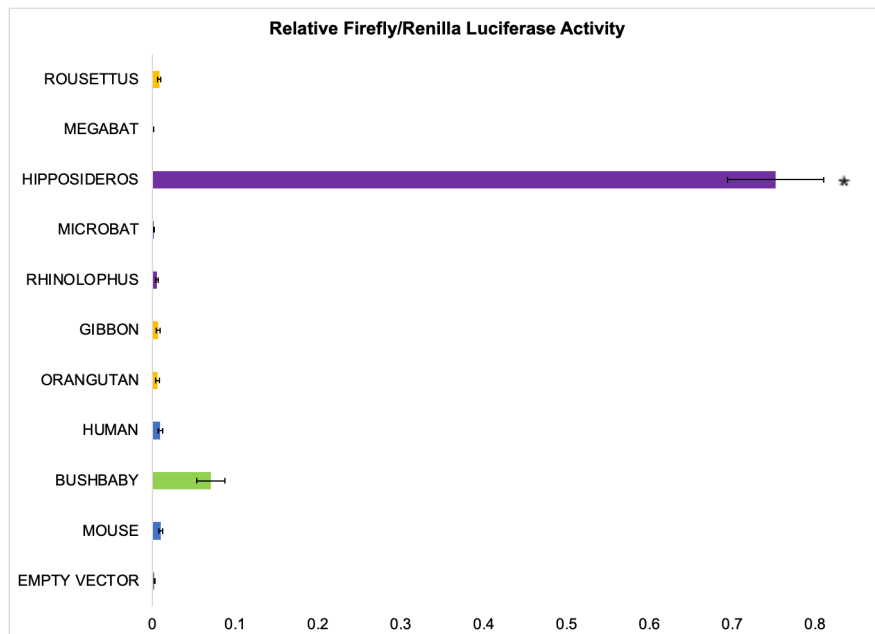
Figure 2.2: Luciferase assays in HepG2 with *SLC2A2* promoters.
Frugivores are colored in yellow, insectivores are colored in purple, omnivores are colored in blue, and the gummivore is colored in green. Student's t-test vs. empty vector *p < .0001
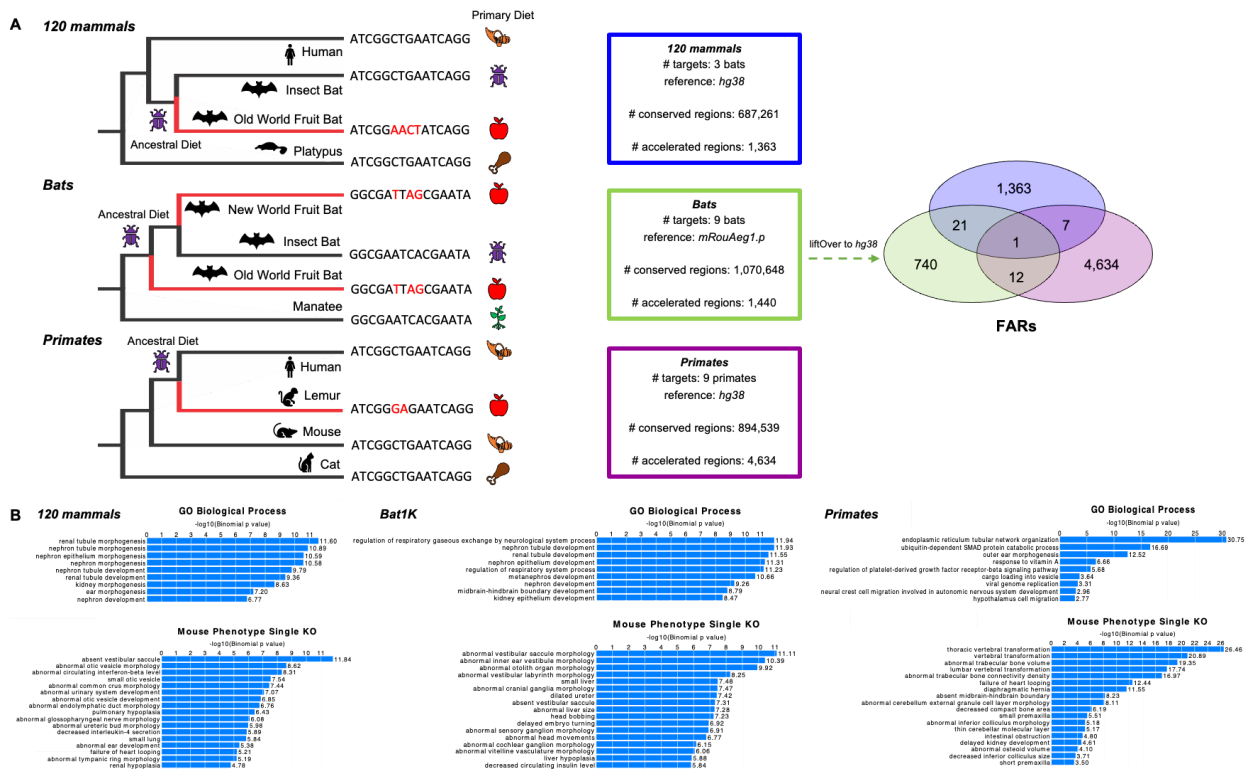
Figure 2.3: Comparative genomics design and analysis.
(**A**) Phylogenetic trees depicting mammals, mock genomic sequences, and diets from three whole genome alignments examined in this study (120 mammals (Hecker and Hiller 2020), Bats (Bat1K), and Primates). Cornucopia (orange) represents omnivory, beetle (purple) represents insectivory, apple (red) represents fruit, meat (brown) represents carnivory, and plant (green) represents herbivory. (**B**) (Left) Whole genome alignments with the number of conserved and accelerated regions identified. (Right) Venn diagram showing the number of FARs that lift to human coordinates (*hg38*) that are shared and not shared between whole genome alignments examined in this study. (**C**) GREAT Go Biological Process and Mouse Phenotype Single KO terms identified from human coordinate FARs (**B**) from each whole genome alignment.

25

Figure 2.4: UCSC Genome Browser snapshots of FARs near *HLX* and *FGF9*.
(**A**) Snapshot of FARs predicted to be associated with *HLX*. Zoom-in shows sequence alignments of FAR shared between all three whole genome alignments. (**B**) Snapshot of FARs predicted to be associated with *FGF9*. Zoom-in shows sequence alignments of 5 FARs.
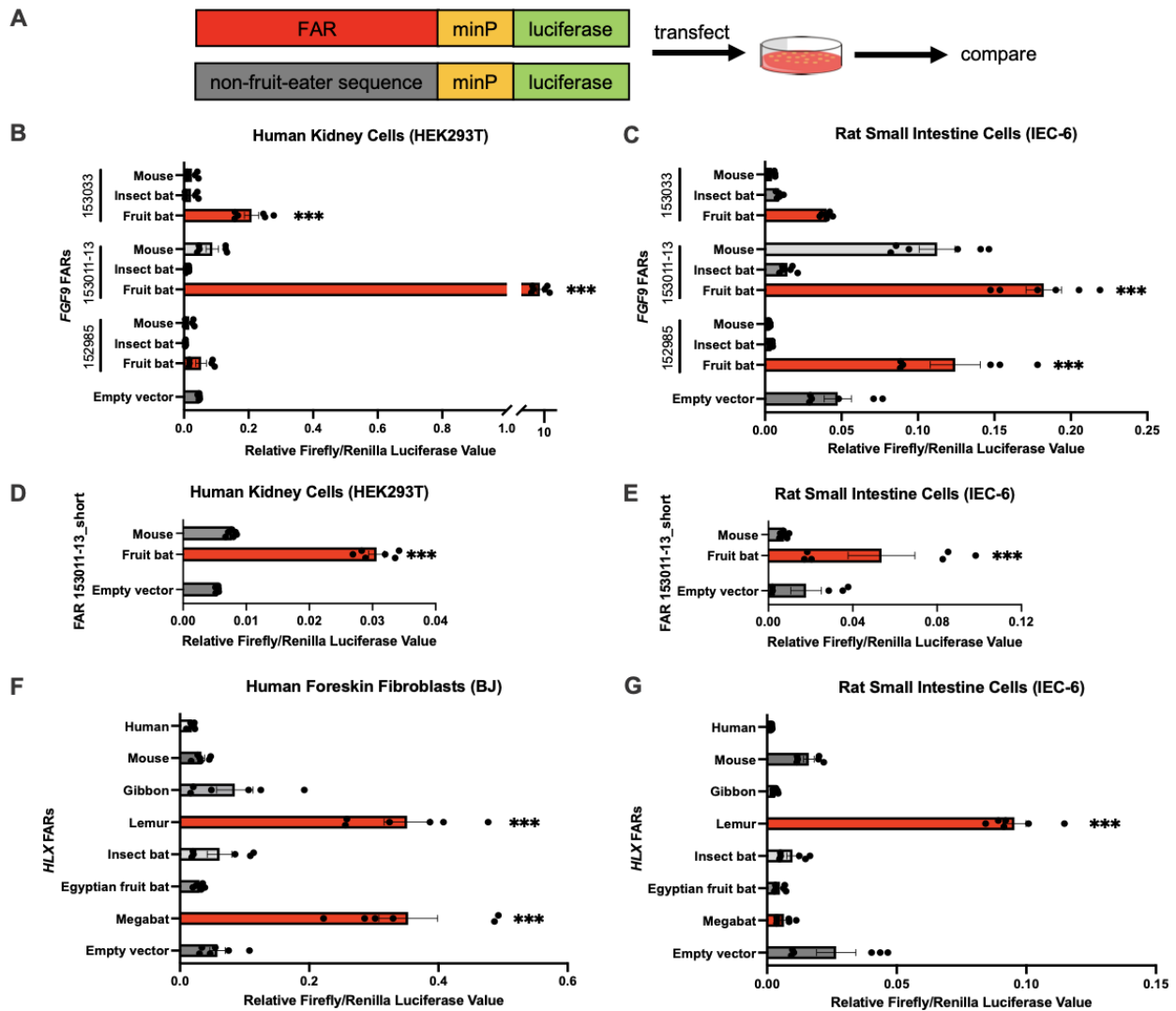
Figure 2.5: Luciferase reporter assays with FARs.
(**A**) Schematic of luciferase reporter assay design. (**B-C**) Bar charts showing relative firefly/renilla luciferase means ± SEM of FARs associated with *FGF9* and non-frugivore corresponding regions in HEK293T (**B**) or IEC-6 (**C**). Student's t-test vs. empty vector ***p < .001 (**D-E**) Bar charts showing relative firefly/renilla luciferase means ± SEM of shortened FAR associated with *FGF9* and non-frugivore corresponding region in HEK293T (**D**) or IEC-6 (**E**). Student's t-test vs. empty vector ***p < .001 (**F-G**) Bar charts showing relative firefly/renilla luciferase means ± SEM of FARs associated with *HLX* and non-frugivore corresponding regions in BJ cells (**F**) or IEC-6 (**G**). Student's t-test vs. empty vector ***p < .001.
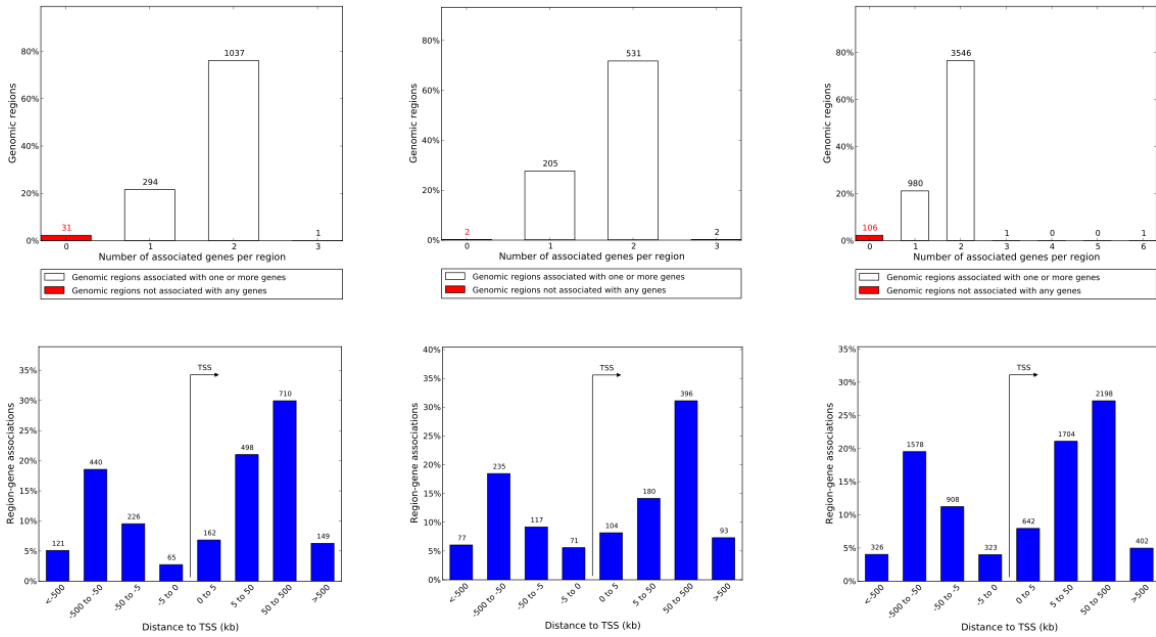
Supplementary Figure 2.1: Number of associated genes per FAR and FAR-gene associations binned by orientation and distance to TSS from GREAT analyses.

Supplementary Figure 2.2: Acceleration patterns in FARs associated with *FGF9*.
(**A-C**) Acceleration patterns in (**A**) *120 mammals* FAR_153011, (**B**) FAR_153012, and (**C**) FAR_153013. Extended branch lengths and red or brown color denote accelerated substitution rate (red > brown). Shorter branch lengths and green color denote conserved substitution rate or purple color for background substitution rate. Names in gray indicate that the sequence is not present.

Supplementary Figure 2.3: Sequence alignments of FARs associated with *FGF9*.
Sequence alignments of *120 mammals* FAR_153011 (118 bp), FAR_153012 (168 bp), and
FAR_153013 (258 bp). Assemblies in blue are frugivorous targets.
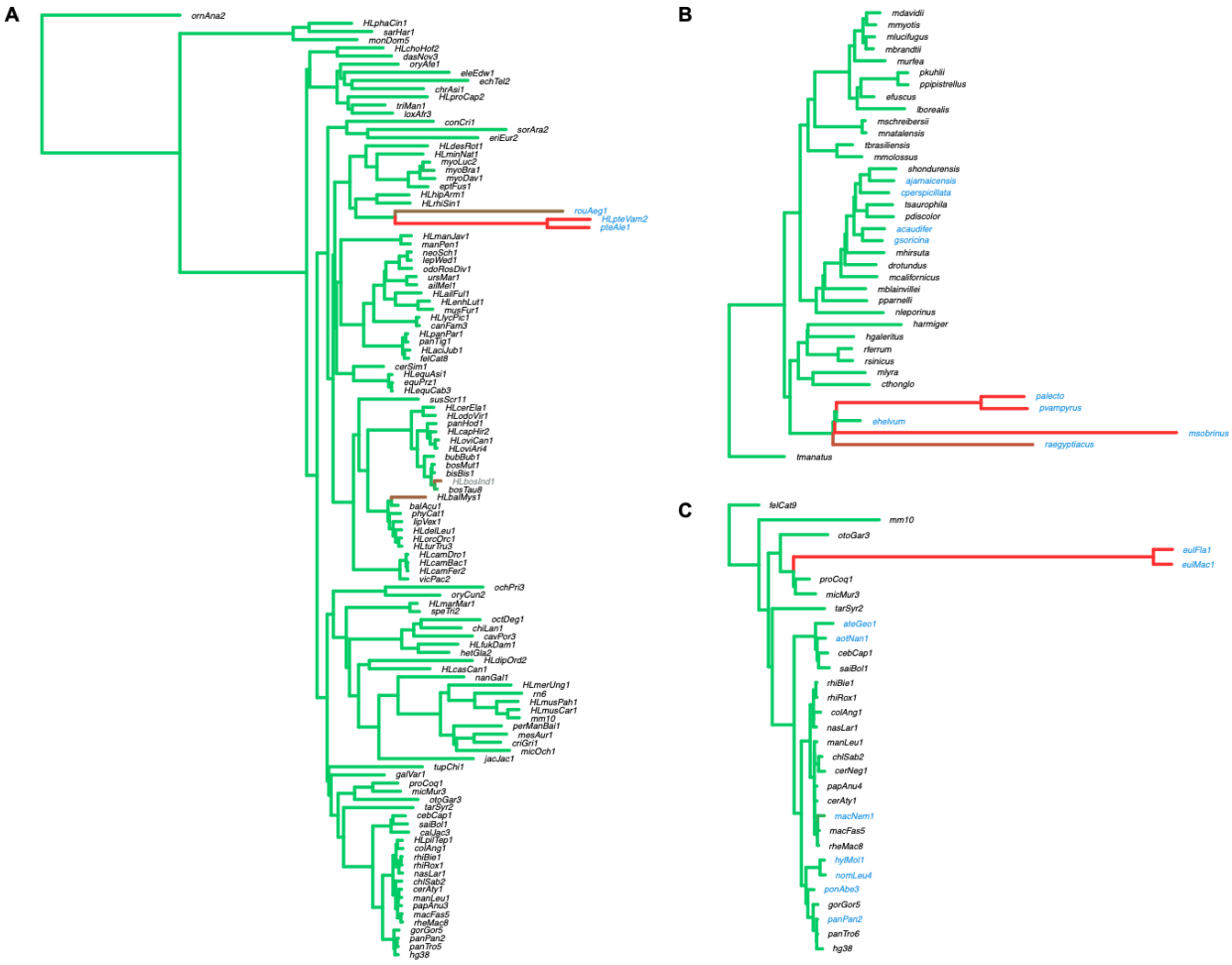
Supplementary Figure 2.4: Acceleration patterns in FARs associated with *HLX*.
(**A-C**) Acceleration patterns in (**A**) *120 mammals* FAR_54670, (**B**) *Bat1K* FAR_3752, and (**C**) primates FAR_29615. Extended branch lengths and red or brown color denote accelerated substitution rate (red > brown). Shorter branch lengths and green color denote conserved substitution rate or purple color for background substitution rate. Assemblies in blue are frugivorous targets. Assemblies in gray indicate that the sequence is not present.
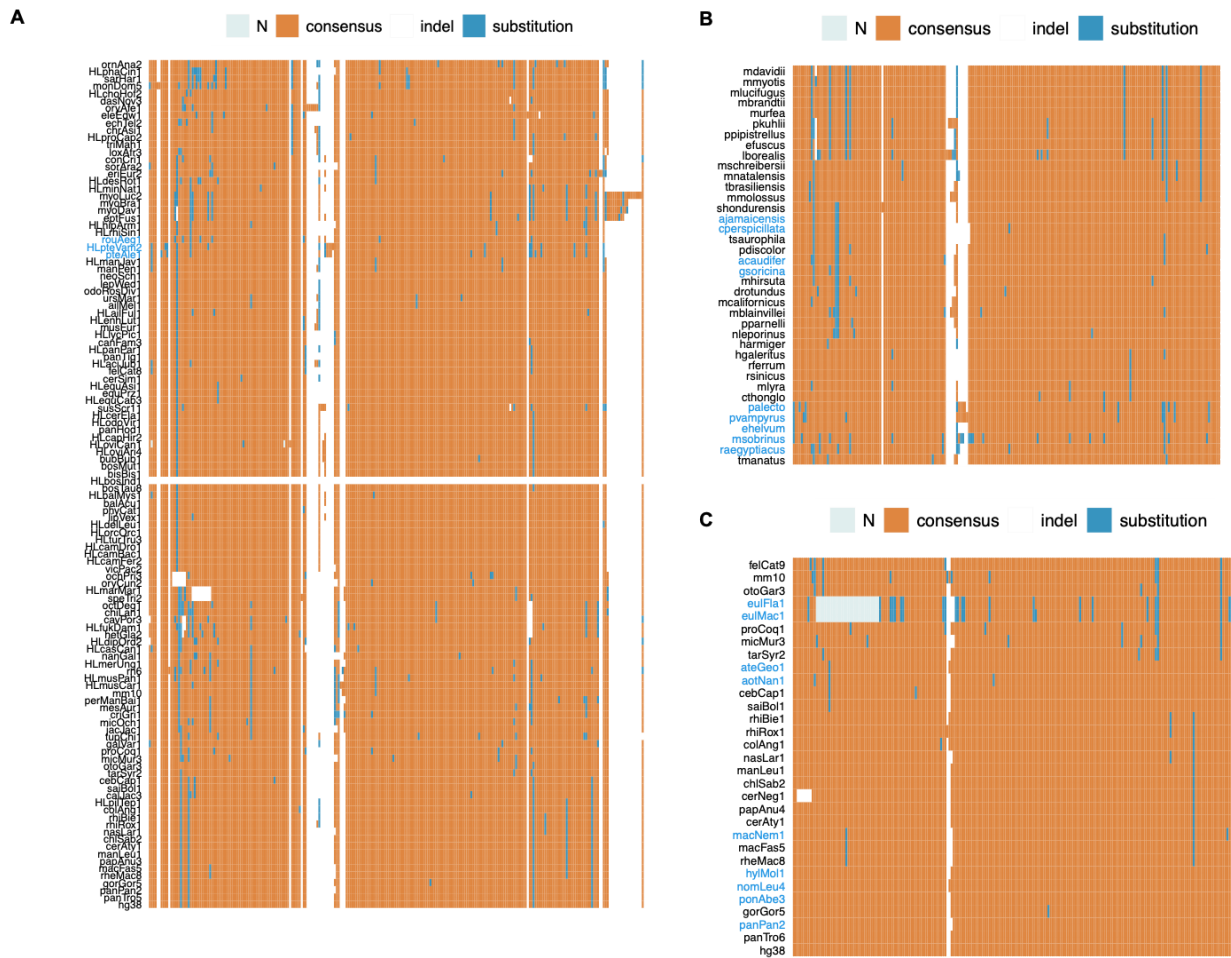
Supplementary Figure 2.5: Sequence alignments of FARs associated with *HLX*.
(**A-C**) Sequence alignments of **A**) *120 mammals* FAR_54670 (254 bp), (**B**) *Bat1K* FAR_3752 (212 bp), and (**C**) primates FAR_29615 (211 bp). Assemblies in blue are frugivorous targets.

Supplementary Figure 2.6: Examples of enriched TFBMs in *120 mammals* FAR_153011-13_short.
Red highlight indicates *TFAP2A* motif. Blue highlight indicates *ZFX* motif. Purple highlight indicates *E2F3* motif. Asterisks indicate matched bases.

```
                            ACCACc
                         CGA    CcxxX

eulFla1     -tgtttgcacggggaacacccacccccgaccc---------------------------- 31
eulMac1     -TGTTTGCACGGGGAACACCCACCCCCGACCC---------------------------- 31
msobrinus   GAGTTCATCGCGGGCTCTCCAAAACCGGTTCCCGTGGGACGGTAACGCGTACACCATCT 120
rouAeg1     AAGTTCATCGTGGGCTCATCCAAAACCGGTTCCCATGGGACGGTAACATTTACACCATAT 74
pteAle1     AAGTTCATCGTGGGCTCATCCAAAACCGGTTCCCATGGGACGGTAACATTTACACCATAT 74
HLpteVam2   AAGTTCATCGTGGGCTCATCCAAAACCGGTTCCCATGGGACGGTAACATTTACACCACAT 74
                *** ***   ** ***   *** *

eulFla1     gggccccgcggaaaatgacttcagcacagagtttatttaatagatcctccaaatc---gt 164
eulMac1     GGGCCCCGCGGAAAATGACTTCAGCACAGAGTTTATTTAATAGATCCTCCAAATT---GT 164
msobrinus   CTAATGAGCACAATTAGCCTCGGCCACCGCATTTGAATTATTAATGCTCCCTAATAGGAT 272
rouAeg1     CTAATGAGCACAATTAACCACAACCACCACATTTGAATTATTAATGCTTCCTAAT---A- 250
pteAle1     CTAATGAGCACAATTAACCCACGACCACCGCAATTGAATTATTAATGCTTTCCTAA---TA 251
HLpteVam2   CTAATGAGCACAATTAACCACGACCACCGCAATTGAATTATTAATGCT-TCCTAA---TA 250
                  ** **    /* ***   *** **   * **  ** **

                            CCACc
                         CGA    C

eulFla1     gggccccgcggaaaatgacttcagcacagagtttatttaatagatcctccaaatc---gt 164
eulMac1     GGGCCCCGCGGAAAATGACTTCAGCACAGAGTTTATTTAATAGATCCTCCAAATT---GT 164
msobrinus   CTAATGAGCACAATTAGCCTCGGCCACCGCATTTGAATTATTAATGCTCCCTAATAGGAT 272
rouAeg1     CTAATGAGCACAATTAACCACAACCACCACATTTGAATTATTAATGCTTCCTAAT---A- 250
pteAle1     CTAATGAGCACAATTAACCACGACCACCGCAATTGAATTATTAATGCTTTCCTAA---TA 251
HLpteVam2   CTAATGAGCACAATTAACCACGACCACCGCAATTGAATTATTAATGCT-TCCTAA---TA 250
                  ** **    *   ***    **  * **  ** **

eulFla1     ccagagctccgagcggctttgctcccgagaccgg----------------cgtccaccg 754
eulMac1     CCAGAGCTCCGAGCGGCTTTGCTCCCGAGACCGG----------------CGTCCACCG 754
msobrinus   GCTGGTCTGCAAATTCCATTTGTAAAACACTCCCGCGGCTCA-AGAAGAGAGTGCTTTCT 890
rouAeg1     GCCGGTCTGCAAATTCCATTTATAAAACACTCCCGCTGCTCAGAAAAGAGAGTGCTTTCT 834
pteAle1     GCCGGTCTGCAAATTCCATTTATAAAACCCTCCTGCTGCTCAA-AAAGAGAGTGCTTTCT 838
HLpteVam2   GCCGGTCTGCAAATTCCATTTATAAAACACTCCCGCTGCTCAA-AAAGAGAGTGCTTTCT 837
                * *  ** **  *     * **  *        ** *              *   *

eulFla1     gggcggggccgaagcattttcctgcaggacctaccccccacccagcccacctccggacga 874
eulMac1     GGGCGGGGCCGAAGCATTTTCCTGCAGGACCTACCCCCCACCCAGCCCACCTCCGGACGA 874
msobrinus   ---CTGTGCTAAAGTCTTTGTCAGCCGGGC-----------------CCTCTGTGGAC-- 977
rouAeg1     ---CTGTGCTAAAGCCATTTTCGGCTGGGC------------------CCTCGTGTGGAC-- 921
pteAle1     ---CTGTGCTAAAGTCATTTTCGGCCGGGC------------------CCTCTGTGGAC-- 925
HLpteVam2   ---CTGTGCTAAAGTCATTTTCAGCCGGGC------------------CCTCTGTGGAC-- 924
                * * **  ***    **   * ** ** *                 *  **  ****

eulFla1     ggcattcagggcagttggggccacgtccgaggagcagaactacgtaaaacgcacccgtgt 934
eulMac1     GGCATTCAGGGCAGTTGGGGCCACGTCCGAGGAGCAGAACTACGTAAAACGCACCCGTGT 934
msobrinus   -GCATTTTCCTCTCGCTTTACCACG-----------AGCCCACCGCACGACCCTCCCTGGG 1026
rouAeg1     -ACATTCCCCTCGCACTTCACTGCT----------AGCCCAGCACACGGCTTTCCCTGGG 970
pteAle1     -ACATTTTCCTCTGACTTCACCACG----------AGCCCTCCACACGACTTCCCTG-GG 973
HLpteVam2   -ACATTTTCCTCTGACTTCACCACG----------AGCCCTCCACACGACTTTCCCTGGG 973
                ****    *        *  *       ** *  *  *    *  **    *
```

Supplementary Figure 2.7: Examples of enriched TFBMs in *120 mammals* FAR_54670/*Bat1K* FAR_3752/*primates* FAR_29615.
Yellow highlight indicates *ZBTB7C* motif. Orange highlight indicates *ZBTB7A* motif. Asterisks indicate matched bases.

Supplementary Figure 2.8: Mouse transgenic enhancer assays with FARs.
(**A**) Schematic of reporter assay design. (**B**) E10.5 mouse embryos with LacZ expression from *primates* FAR_29615 and *120 mammals* FAR_54670/*Bat1k* FAR_3752. Zoom-ins show LacZ-stained embryo section with nuclear fast red counter-staining. (C) Mouse embryos at E12.5 (top) and E14.5 (bottom) with LacZ expression from *120 mammals* FAR_153011-13.
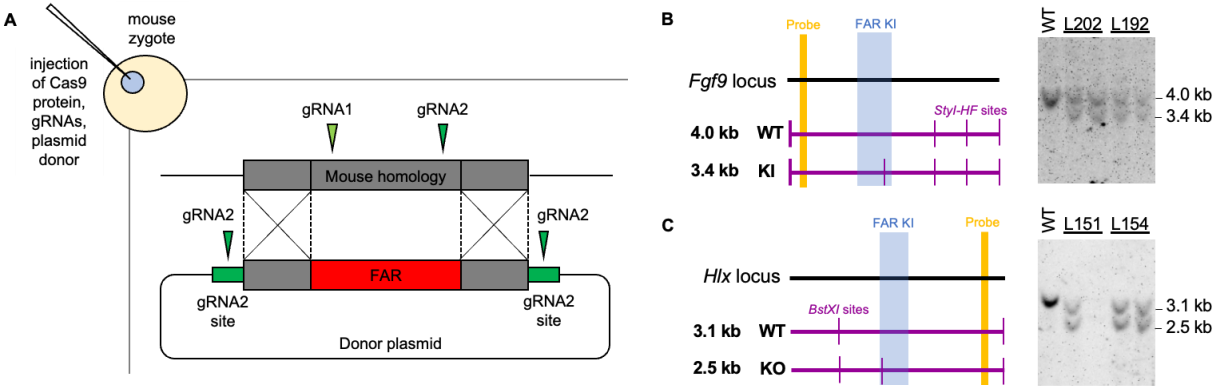
Supplementary Figure 2.9: Mouse genomic sequence swap with FARs design and Southern blotting.
(**A**) Schematic of FAR KI design. (**B-C**) Southern blot probe design (left) and blot (right) for FAR 153011-13 (**B**) and FAR 54670/3752 (**C**).

Supplementary Table 2.1: *SLC2A2* proximal promoter sequences synthesized for luciferase assays with pGL4.23[minP].

Species	Diet	NCBI Nucleotide Accession
Rhinolophus ferrumequinum	Insectivore	KT961105.1
Hipposideros armiger	Insectivore	KT961103.1
Carollia sowelli	Frugivore	KT961109.1
Artibeus lituratus	Frugivore	KT961111.1
Rousettus leschenaultii	Frugivore	KT961116.1
Paranyctimene raptor	Frugivore	KT961115.1

Supplementary Table 2.2: *SLC2A2* promoter sequences synthesized for luciferase assays with pGL4.11b.

Species	Diet	NCBI GenBank Accession	Coordinates
Rhinolophus sinicus	Insectivore	GCA_001888835.1	NW_017738930.1:2002522-2003277
Hipposideros armiger	Insectivore	GCA_001890085.1	NW_017731433.1:3839357-3840121
Myotis lucifigus	Insectivore	GCA_000147115.1	GL429820:1726740-1727486
Rousettus aegyptiacus	Frugivore	GCA_001466805.2	NW_015495092.1:82528-83266
Pteropus vampyrus	Frugivore	GCA_000151845.1	scaffold_6977:81577-82330
Nomascus leucogenys	Frugivore	GCA_000146795.3	chr11:95482790-95483559
Pongo abelii	Frugivore	GCA_002880775.3	chr3:168604456-168605243
Homo sapiens	Omnivore	GCA_000001405.29	chr3:171026671-171027445
Mus musculus	Omnivore	GCA_000001635.2	chr3:28697318-28698016
Otolemur garnettii	Gummivore	GCA_000181295.3	GL873603:8317729-8318416

Supplementary Table 2.3: Species, diets, and genomes used in the *Bat1k* dataset.

Family	Species	Diet	GenBank Assembly Accession or Assembly Group
Craseonycteridae	*Craseonycteris thonglongyai*	Insectivore	GCA_004027555.1
Hipposideridae	*Hipposideros armiger*	Insectivore	GCA_001890085.1
Hipposideridae	*Hipposideros galeritus*	Insectivore	GCA_004027415.1
Megadermatidae	*Megaderma lyra*	Carnivore	GCA_004026885.1
Miniopteridae	*Miniopterus natalensis*	Insectivore	GCA_001595765.1
Miniopteridae	*Miniopterus schreibersii*	Insectivore	GCA_004026525.1
Molossidae	*Molossus molossus*	Insectivore	GCA_014108415.1
Molossidae	*Tadarida brasiliensis*	Insectivore	GCA_004025005.1
Mormoopidae	*Mormoops blainvillei*	Insectivore	GCA_004026545.1
Mormoopidae	*Pteronotus parnellii*	Insectivore	GCA_000465405.1
Noctilonidae	*Noctilio leporinus*	Piscivore	GCA_004026585.1
Phyllostomidae	*Carollia perspicillata*	Frugivore	Bat1K
Phyllostomidae	*Glossophaga soricina*	Frugivore/Nectarivore	Bat1K
Phyllostomidae	*Phyllostomus discolor*	Omnivore	GCA_014049915.1
Phyllostomidae	*Artibeus jamaicensis*	Frugivore	GCA_004027435.1
Phyllostomidae	*Desmodus rotundus*	Sanguivore	GCA_002940915.2
Phyllostomidae	*Tonatia saurophila*	Omnivore	GCA_004024845.1
Phyllostomidae	*Anoura caudifer*	Frugivore/Nectarivore	GCA_004027475.1
Phyllostomidae	*Sturnira hondurensis*	Nectarivore	Genome Warehouse in BIG Data Center
Phyllostomidae	*Macrotus californicus*	Insectivore	GCA_007922815.1
Phyllostomidae	*Micronycteris hirsuta*	Omnivore	GCA_004026765.1
Pteropodidae	*Rousettus aegyptiacus*	Frugivore	GCA_014176215.1
Pteropodidae	*Eidolon helvum*	Frugivore	GCA_000465285.1
Pteropodidae	*Macroglossus sobrinus*	Frugivore/Nectarivore	GCA_004027375.1
Pteropodidae	*Pteropus alecto*	Frugivore	GCA_000325575.1

Family	Species	Diet	GenBank Assembly Accession or Assembly Group
Pteropodidae	*Pteropus vampyrus*	Frugivore	GCA_000151845.2
Rhinolophidae	*Rhinolophus ferrumequinum*	Insectivore	GCA_014108255.1
Rhinolophidae	*Rhinolophus sinicus*	Insectivore	GCA_001888835.1
Vespertilionidae	*Myotis myotis*	Insectivore	GCA_014108235.1
Vespertilionidae	*Pipistrellus kuhlii*	Insectivore	GCA_014108245.1
Vespertilionidae	*Eptesicus fuscus*	Insectivore	GCA_000308155.1
Vespertilionidae	*Lasiurus borealis*	Insectivore	GCA_004026805.1
Vespertilionidae	*Murina feae*	Insectivore	GCA_004026665.1
Vespertilionidae	*Myotis brandtii*	Insectivore	GCA_000412655.1
Vespertilionidae	*Myotis davidii*	Insectivore	GCA_000327345.1

Supplementary Table 2.4: Species, diets, and genomes used in the *primates* dataset.

Species	Common name	Diet	GenBank Assembly Accession
Aotus nancymae	Ma's night monkey	Frugivore	GCA_000952055.2
Ateles geoffroyi	Black-handed spider monkey	Frugivore	GCA_004024785.1
Cebus capucinus imitator	White-faced sapajou	Omnivore	GCA_001604975.1
Cercocebus atys	Sooty mangabey	Granivore	GCA_000955945.1
Cercopithecus neglectus	De Brazza's monkey	Omnivore	GCA_004027615.1
Colobus angolensis palliatus	Angolan colobus	Folivore	GCA_000951035.1
Eulemur flavifrons	Blue-eyed black lemur	Frugivore	GCA_001262665.1
Eulemur macaco macaco	Black lemur	Frugivore	GCA_001262655.1
Felis catus	Cat	Carnivore	GCA_000181335.5
Gorilla gorilla gorilla	Gorilla	Folivore	GCA_900006655.1
Homo sapiens	Human	Omnivore	GCA_000001405.15
Hylobates moloch	Silvery gibbon	Frugivore	GCA_009828535.2
Macaca fascicularis	Crab-eating macaque	Omnivore	GCA_000364345.1
Macaca mulatta	Rhesus	Omnivore	GCA_000772875.3
Macaca nemestrina	Pig-tailed macaque	Frugivore	GCA_000956065.1
Mandrillus leucophaeus	Drill	Omnivore	GCA_000951045.1
Microcebus murinus	Mouse lemur	Omnivore	GCA_000165445.3
Mus musculus	Mouse	Omnivore	GCA_000001635.2
Nasalis larvatus	Proboscis monkey	Folivore	GCA_000772465.1
Nomascus leucogenys	Northern white-cheeked gibbon	Frugivore	GCA_006542625.1
Otolemur garnettii	Bushbaby	Omnivore	GCA_000181295.3
Pan paniscus	Bonobo	Frugivore	GCA_000258655.2
Pan troglodytes	Chimp	Omnivore	GCA_002880755.3
Papio anubis	Baboon	Omnivore	GCA_000264685.2
Pongo abelii	Sumatran orangutan	Frugivore	GCA_002880775.3
Propithecus coquereli	Coquerel's sifaka	Folivore	GCA_000956105.1
Rhinopithecus bieti	Black snub-nosed monkey	Lichenivore	GCA_001698545.1
Saimiri boliviensis	Squirrel monkey	Omnivore	GCA_000235385.1
Tarsius syrichta	Tarsier	Insectivore	GCA_000164805.2

Supplementary Table 2.5: Genomic primers used for luciferase and mouse transgenic enhancer assays (underlined).

FAR	Species (assembly)	Coordinates	Primers
152985	Egyptian fruit bat (HLrouAeg4)	scaffold_m13_p_11:715,658-716,721	Forward: GGGCACTGGAACTTTTCTTG Reverse: ATGCAAGGCTCATTCTCC
152985	Natal long-fingered bat (HLminNat1)	LDJU01000412:664,773-666,367	Forward: CCGGTCAAGGCAGTAACAATC Reverse: TTTGTGTGCATGGTTGTGTG
152985	Mouse (mm10)	chr14:58,386,213-58,389,161	Forward: CACTCTCAATCCACCCAC Reverse: CACAAAGGCTGTGTCTC
153011-13	Egyptian fruit bat (HLrouAeg4)	scaffold_m13_p_11:747,845-749,981	Forward: CACGGACACACTGGTCTCAG Reverse: TGACGTGACGTACCTTCTGG
153011-13	Natal long-fingered bat (HLminNat1)	LDJU01000412:507,590-509,839	Forward: TGCCCATAACTGGACTGTG Reverse: CAACAGGGGAAGCAGAGAAG
153011-13	Mouse (mm10)	chr14:58,498,589-58,500,691	Forward: CAATGGCAGGAAACTGGG Reverse: GGAAGTGACGCTTTCACAC
153033	Egyptian fruit bat (HLrouAeg4)	scaffold_m13_p_11:767,776-770,744	Forward: ACATGGCCCTTGTGTTTCTC Reverse: TTTGTTTAAACGCCACACCA
153033	Natal long-fingered bat (HLminNat1)	LDJU01000412:361,854-363,632	Forward: TCTCTCACCCTAGCCTTC Reverse: TTTTCTGCAGCCCATTAC
153033	Mouse (mm10)	chr14:58,655,253-58,658,247	Forward: CCCTCCTGTTACAACCCAG Reverse: TGGAGACACACATGGCTG
<u>153011-13_short</u>	<u>Egyptian fruit bat (HLrouAeg4)</u>	<u>scaffold_m13_p_11:748,651-749,587</u>	<u>Forward: GCCCGGGGCCCGACGGGT Reverse: CAGGGGCTCTGCCCAGGGGG</u>

FAR	Species (assembly)	Coordinates	Primers
153011-13_short	Mouse (hub_158535_FVB_NJ_dna)	chr14:50732782-50733771	Forward: GCTTGGGGCCCGGAACAG Reverse: CAATGGCTGTGCCCCCAG
54670/3752	Megabat (HLpteVam2)	KN525965:50475-51574	synthesis
54670/3752	Natal long-fingered bat (HLminNat1)	NW_01554859.1:4011884-4012831	Forward: ACACCGTGCATGTGTTCATT Reverse: ACTTCCACAGATCCGTACCC
29615	Eulemur flavifrons isolate Harlow scaffold6622	gb|LGHW01006703.1|:52538-53582	synthesis
54670/3752/29615	Mouse (hub_158535_FVB_NJ_dna)	chr1:184583650-184584648	Forward: CCAATGCCTCTGACCAGATTC Reverse: GGGAACACGCACCTCTATG
54670/3752/29615	Human (hg38)	chr1:221,029,922-221,030,921	Forward: CCTCTGTTTCTCTGCAGAATC Reverse: CAATACCTGGGTCCAAATTC
54670/3752	Egyptian fruit bat (HLrouAeg4)	scaffold_m13_p_18:12,148,408-12,149,505	Forward: GCTCTACCCCTTAACCTG Reverse: TGTCCATGTATAAGTGATTCTTC
29615	Gibbon (nomLeu3)	chr5:26,818,912-26,820,011	Forward: CCACAATGGATGACTTTTCC Reverse: CCCACAGCCATGGTAGTTTTAC

Supplementary Table 2.6: TFBM enrichment results.
P-value is determined from Fisher's exact test and adjusted for multiple tests using a Bonferroni correction. True positive (TP) is the fraction of FAR sequences enriched for the motif. False positive (FP) is the fraction of non-frugivorous, homologous FAR sequences enriched for the motif.

FAR	JASPAR ID	TF	Adjusted *p*-value	TP	FP
153011-13_short	MA0469.1	*E2F3*	1.54e-5	3/3	0/114
153011-13_short	MA0146.2	*ZFX*	1.92e-5	3/3	0/114
153011-13_short	MA0146.1	*ZFX*	1.92e-5	3/3	0/114
153011-13_short	MA0003.1	*TFAP2A*	5.00e-5	3/3	0/114
54670/3752/29615	MA0659.1	*ZBTB7C*	2.25e-8	6/6	1/116
54670/3752/29615	MA0750.1	*ZBTB7A*	4.19e-6	5/6	2/116

Supplementary Table 2.7: Mouse homology arm sequences for mouse genomic sequence swapping.

FAR	Mouse homology arm	hub_158535_FVB_NJ_dna_ Coordinates
153011-13	Left	chr14:50731782-50732781
153011-13	Right	chr14:50,733,772-50,734,511
54670/3752	Left	chr1:180322089-180323088
54670/3752	Right	chr1:180324088-180325087

Supplementary Table 2.8: sgRNA sequences for mouse genomic sequence swapping.

sgRNA	Sequence (PAM)	hub_158535_FVB_NJ_dna_ Coordinates
153011-13-gRNA2	ACCAGCTTGACCCACGAAGCAGG	chr14:50732849-50732871
153011-13-gRNA1	CTCCCATTTGTCAAATCCTCAGG	chr14:50733661-50733683
54670/3752-gRNA2	CCTCTGACCAGATTCATCACAGG	chr1:180323095-180323117
54670/3752-gRNA1	CCTTGGACTGCTATCTTCTAGGG	chr1:180323702-180323724

Supplementary Table 2.9: FAR knock-in genotyping primers.
WT = wild-type; KI = knock-in.

Target	hub_158535_FVB_NJ_dna_ Coordinates	Primers	Expected Band Size
153011-13	chr14:50732731-50733126	Forward: AGGTGCATGTTTGGTGAGTG Reverse: TTACAGACGCTCAGCCATGA	WT: 396 KI: 202
54670/3752	chr14:180323320-180324166	Forward: CTCCCTAATAGGGTGCATTC TC Reverse: ACCGTGGATGTGTTTGTTGG	WT: 984 KI: 216

Supplementary Table 2.10: Southern blot probe genomic primers and sequences.

Probe	hub_158535_FVB_NJ_dna_Coordinates	Primers	Sequence
153011-13	chr14:50730290-50730592	Forward: GGGGCAGTCTTGATATGGCC Reverse: CCTCAGTGGGCCCTGCCTGCATTTC	GGGGCAGTCTTGATATGGCCATGAATTTC CCCTTCATTTCTAAGTGCAGTTCAGATTG GATTGGGCATTTACAAACAGACAGCTTTT GTTCTACCTCCAGGATCATAGAAGGAATA ATTTGTCCACTGAATAATATACAACCTGGC CCTAACTATGATTCTTCCCATTCCTCTAGC ATCAAGGCATCTCATAGGACCCTCTAGAG CAATCTTATTTTGAGCTTGAGACCTCTCC CTGTCTGGTGCCCCGTCCCATCCCATATT TCCAGGATTTCTCTGAAATGCAGGCAGG GCCCCACTGAGG
54670/3752	chr1:180325115-180325357	Forward: AATTTATGTCCCAAGATATTCACC Reverse: ACATGCTCAGACAGTCCTTATTTC	AATTTATGTCCCAAGATATTCACCTGCCTC ACAATATCATGGGCTTCTACCTCAGCCAG GAAAGAGGCCTATGCCATTCACAAACCA CAGTAGTTTGAACCGCATATTAAAATGATT AGTTCTGGGACTGCGAAGGTAAGTTCGT ACCTTCTTCAGGTGAATGGAGATAATTTT GTTTCATTTGTGAGAGCCAACAACAGTCA TAATTTGGTAAAAGGAGAAATAAGGACTG TCTGAGCATGT

Chapter 3: Integrative single-cell characterization of frugivory adaptations in the bat

kidney and pancreas

3.1: Abstract

Here, we used integrative single-cell RNA-seq and ATAC-seq on adult insectivorous (*Eptesicus fuscus*) and frugivorous (*Artibeus jamaicensis*) bats to identify cell populations, genes, and regulatory element differences that could be associated with frugivory adaptations in the kidney and pancreas. In total, we analyzed over 34,696 cells from eight insectivorous and seven frugivorous bats. We developed a cross-species integrated analysis framework for non-model organism genomes, and utilized human and mouse single-cell kidney and pancreas markers to annotate cell-types. Our data reveal cell-type, gene expression, and gene regulation differences between insectivorous and frugivorous bats. These include more collecting duct cells and gene expression changes involved in fluid and electrolyte balance in the frugivore kidney, and a relative decrease in exocrine cells and gene expression changes involved in insulin secretion and glucose response in the frugivore pancreas. Cell composition differences were further validated with immunofluorescence on insectivorous and frugivorous bat tissues, confirming a reduction in renal medulla and an expansion of endocrine tissue in fruit bats. Transcription factor (TF) analyses of single-cell ATAC-seq found divergent TF binding site (TFBS) usage between dietary phenotypes. Together, our single-cell multi-omics approach indicates that the frugivore kidneys and pancreases exhibit many signatures of diabetes, such as increased potassium secretion, gluconeogenesis, and glucose reabsorption in the kidney and hyperinsulinemia and hyperglycemia. In summary, our work provides the first single-cell datasets for bats that also compares kidneys and pancreases between closely related mammals of contrasting diets, providing insight into cell composition, gene expression, and gene regulation adaptations to frugivory that are also related to human disease, in particular with diabetes.

3.2: Background

One of the main challenges a fruit-centered diet poses is high blood sugar levels, which can lead to diabetes. To overcome this, an animal must rapidly control high sugar levels. Interestingly, although fruit bats consume more sugar than non-frugivorous bats, they can lower their blood sugar faster(*3*, *17*, *19*, *20*). Fruit bats demonstrate high sensitivity to glucose and insulin(*75*) and can directly fuel their metabolic needs with exogenous sugars(*18*). The pancreas is responsible for generating hormones that regulate blood sugar and appetite, such as insulin and glucagon(*101*), as well as secreting enzymes and digestive juices into the small intestine(*102*). The high sensitivity to sugar levels in fruit bats is thought to be supported by an expansion of endocrine tissue in the pancreas(*75*, *103*) and by loss of genes involved in insulin metabolism and signaling(*15*).

Fruit is rich in water and low in electrolytes like sodium and calcium. The kidney is responsible for maintaining water and salt balance, filtering the blood of waste, and maintaining blood pressure(*104*). The kidney is also involved in metabolizing sugars, generating glucose, and clearing insulin from circulation(*105*). Several kidney modifications evolved in fruit bats that lead to dilute urine production(*106*), including a larger renal cortex and a smaller renal medulla relative to insect-eating bats(*22*, *35*) as well as several transporter losses(*15*).

3.3: Methods

Bat samples and dietary treatment

For Chromium single-cell Multiome ATAC + Gene Expression experiments, kidneys and pancreases were obtained from four adult male *Artibeus jamaicensis* fruit-eating bats (Colorado State University [CSU]) and five adult male *Eptesicus fuscus* insect-eating bats (6-7 years old; Northeast Ohio Medical University [NEOMED]). All Jamaican fruit bats were fasted overnight during their night cycle (approximately 12 hours), and "fasted" bats were removed from their enclosures before feeding for euthanasia. "Fed" Jamaican fruit bats were provided unlimited cantaloupe and banana and were euthanized 30 minutes later. Tissues were harvested immediately for flash-freezing (May 2021) and stored at -80 Celsius until nuclei isolation. Because big brown bats feed on mealworms (*Tenebrio molitor*), we increased the fruit content of these mealworms to create a fruit "treatment" for insect-eating bats. These mealworms were fed a modified fruit diet for four days of sweet potatoes, apples, mango, and cane sugar. "Fasted" big brown bats were removed from their enclosures before feeding for euthanasia, and three "fed" big brown bats were fed an unlimited supply of fruit-enriched mealworms and were euthanized 30 minutes later. Tissues were harvested immediately for flash-freezing (all big brown bats were euthanized before the winter (September 2020) to exclude hibernation effects) and stored at -80 Celsius until nuclei isolation. All experimental procedures were approved under IACUC protocols (NEOMED IACUC #20-06-721, CSU IACUC #1034).

For immunofluorescence experiments, kidneys and pancreases were obtained from 3 adult male big brown bats of unknown age (June 2022; NEOMED IACUC #20-06-721), 2 adult male Jamaican fruit bats (August 2022; CSU IACUC #1034), and 1 adult female Jamaican fruit bat captured on an American Museum of Natural History (AMNH) field expedition to the Lamanai

Archaeological Reserve in Orange Walk District of Belize in November 2021. The individual

sampled (field number BZ701, voucher number AMNH.Mammalogy.281145) was caught in a

ground-level mist net set near the High Temple at Lamanai (17.76750 N, 88.65207 W) on 17

November 2021. The bat was subjected to minimal handling after capture, and it was held in a

clean cloth bag after capture as per best practices for field containment of bats. After species

identification, the individual was euthanized humanely by isoflurane inhalation the same night it

was captured. Capture and sampling were conducted under Belize Institute of Archaeology

Permit IA/S/5/6/21(01) and Belize Forest Department Permit FD/WL/1/21(16), and samples

were exported under Belize Forest Department permit FD/WL/7/22(08). All work was conducted

with approval by the AMNH Institutional Animal Care and Use Committee (AMNH

IACUC-20191212). Tissues were removed from the subject individual immediately following

euthanasia and were flash-frozen in a liquid nitrogen dry shipper, with the cold chain maintained

from field to museum to laboratory.

Chromium single-cell multiome ATAC and gene expression sample processing and

sequencing

For Chromium single-cell Multiome ATAC + Gene Expression experiments, we followed the

manufacturer's protocols for complex tissues (10x Genomics: CG000375, CG000338). Nuclei

were sorted on a FACS Aria II (Becton Dickinson) using a capture target of 10,000 nuclei per

sample to prepare libraries. ATAC libraries and GEX libraries were pooled together by tissue.

Tissue-specific libraries were sequenced PE150 on two lanes of NovaSeq 6000 (Illumina)

(**Supplementary Table 3.1**). The pancreas of a fed Jamaican fruit bat was not used due to low

quality sequencing data.

Bat genome modifications for joint scRNA and scATAC analysis

Jamaican fruit bat and big brown bat genomes and annotations were downloaded from NCBI

(GenBank Assembly Accession: GCA_014825515.1, GCA_000308155.1). Scaffolds smaller

than 50 kb for each genome were removed. Gene information in annotation files was collapsed

so that each gene was represented by a single "exon" transcript. A mitochondrial genome was

generated for the big brown bat using GetOrganelle(*107*), with default parameters for animal

mitochondria assembly, and the output fasta was input into MITOS(*108*), with vertebrate genetic

code, to create the corresponding gtf file (GenBank TPA: BK063052). Each species annotation

was modified with one-to-one orthologue IDs created from Orthofinder(*109*) (**Supplemental File

3.14**), which takes proteomes as input and assigned 89.9% of the total bat genes.

Individual sample scRNA-seq and scATAC-seq processing and quality control (QC)

A total of 15 scRNA and scATAC FASTQs were input into cellranger-arc 2.0 (10x Genomics)

whereby raw feature barcode matrices were generated with "count". ATAC matrices were loaded

as an object into Seurat version 4.0(*110*). A common peak set was created within a species with

R package GenomicRanges 1.50.2(*111*) "reduce". RNA matrices were then combined with

ATAC matrices using only cells that overlapped the existing object. Cells were filtered for

mitochondrial percentage < 25%; ATAC peak count 500 < x < 100,000; RNA count 200 < x <

25,000; nucleosome signal < 2; TSS enrichment >1 (terms defined by ENCODE(*95*)).

Joint scRNA-seq and scATAC-seq bioinformatics workflow within species

Using post-QC Seurat objects as generated above, SCTransform was used to normalize each

sample with mitochondrial percentage as a regressed variable and to find the top 3,000 variable

features from each sample. Species replicates were then merged. To integrate RNA (SCT)

across all samples within a species, 3,000 repeatedly variable features across each sample

were selected with "SelectIntegrationFeatures", and Harmony was run on the PCA dimensions 1:30, removing original sample identification as a variable. To integrate ATAC across all samples within a species, peaks were normalized with Signac function "RunTFIDF", the most frequently observed features were identified with Signac version 1.8.0(*111*) function "FindTopFeatures" with "min.cutoff = 'q0'", and Harmony was run on the singular value decomposition (SVD) dimensions 2:30, removing original sample identification as a variable. To integrate the integrated RNA and ATAC modalities, weighted nearest neighbor analysis(*110*) was applied on the Harmony reductions using 30 dimensions for each modality. Seurat function "FindClusters" was used with SLM algorithm and 0.6 resolution to identify clusters. Seurat function "FindAllMarkers" was used to identify cluster-specific markers and manually assign clusters according to shared markers to known mouse and human cell-types.

Joint scRNA-seq and scATAC-seq bioinformatics workflow across species

Using post-QC Seurat objects mentioned above, gene activity matrices were created for each sample with the Seurat function "GeneActivity", and ATAC peaks were removed from each Seurat object. R package SCTransform(*112*) was used to normalize each sample with mitochondrial percentage as a regressed variable and to find the top 3,000 variable features from each sample. Samples of the same species were then merged while retaining SCT matrices. Merged gene activity matrices were then log-normalized with the Seurat function "NormalizeData" with "scale.factor" set to the median of the gene activity counts. Two-thousand variable features were detected for normalized gene activity scores for each species with Seurat function "FindVariableFeatures" (vst method). Species were then merged. To integrate RNA (SCT) across species, we used Seurat function "SelectIntegrationFeatures" to select the 3000 top scoring variable features across all samples and manually added the selected variable features to the merged samples, and Harmony(*113*) was run on the principal component

analysis (PCA) dimensions 1:30 with 20 maximum iterations and removing original sample and

species variables. To integrate gene activity scores across species, we used Seurat function

"SelectIntegrationFeatures" to select the 2,000 top scoring variable features across all samples

and manually added the selected variable features to the merged samples. The features were

scaled with Seurat function "ScaleData", and Harmony was run on the PCA dimensions 1:30,

removing original sample identification and species variables. To integrate the RNA and gene

activity score modalities, weighted nearest neighbor analysis(*110*) was applied on the Harmony

reductions using 30 dimensions for each modality. Seurat function "FindClusters" was used with

SLM algorithm and 0.6 resolution to identify clusters, and "FindSubClusters" was used with

original Louvain algorithm and 0.1 resolution to identify subclusters in pancreatic acinar cells.

Seurat function "FindAllMarkers" was used to identify cluster-specific markers and manually

assign clusters according to shared canonical markers to known mouse and human cell-types.

Cell composition analyses

We calculated cell-type proportions from each sample and compared proportion percentages

between the species by calculating significance with the Wilcoxon signed-rank test. The

resulting boxplots for proportion percentages for each species and p-values were visualized with

R package ggplot2 version 3.4.0(*114*). For calculating species or condition associations per cell,

we used R package CNA version 0.0.99(*115*). For species associations, we used species

variables as testing variables and used the neighbor graph generated from Seurat's Weighted

Nearest Neighbor Analysis from above. For condition associations, we used condition variables

as testing variables and regressed out effects from species variables. For mosaic plots and

calculation of p-values with pearson residuals, we used R package VCD version 1.4.10(*116*)

with species information and cell-type annotations.

Automated cell-type annotations with human and mouse references and comparison

For the automated cell-type annotation method using human and mouse single-cell references, we used R package Azimuth version 0.4.6(*110*) with built-in Azimuth human kidney dataset(*110*);(*117*), mouse p0 and adult kidney(*118*), Azimuth human pancreas dataset(*110*, *119*);(*120–125*), and human adult pancreas(*119*). For reference datasets with several levels of annotations, the most major cell-types were selected. For comparing similarity between manual annotations and automated annotations, we calculated Jaccard similarity coefficients between cells from each cell-type with different annotation methods. The resulting similarity heatmaps were visualized with R package pheatmap version 1.0.12(*126*).

Immunofluorescence staining, imaging and analysis

Flash-frozen samples were embedded in Tissue-Tek OCT Compound (Sakura Finetek, Torrance, Ca, USA), cut as 14uM cryo-sections, and fixed in 4% paraformaldehyde for 10 minutes at room temperature. Sections were rinsed in Phosphate Buffered Saline (PBS) and blocked in a humidity chamber in 3% normal donkey serum (Sigma-Aldrich, 566460) and 0.3% Triton X-100 (Sigma-Aldrich, T8787) in PBS. Sections were then incubated overnight at 4 degrees Celsius in a humidity chamber with one of the following protein-targeting primary antibodies from Thermo Scientific™ at respective dilutions in blocking buffer: *SLC12A1* (18970-1-AP; 1:100), *AQP2* (PA5-78808; 1:100), *SLC26A4* (PA5-115911; 1:100), *INS* (15848-1-AP; 1:500), *GCG* (15954-1-AP; 1:500). Epitope matching can be found in **Supplementary Table 3.4**. Sections were washed with PBS and incubated in a humidity chamber for two hours at room temperature with secondary antibody Donkey Anti-Rabbit IgG NorthernLights™ NL557-conjugated Antibody (Bio-techne, NL004; 1:200). Sections were rinsed with PBS and either incubated with Hoechst 33342 Solution (20mM) (Thermo Scientific™,

62249; 1:10000) for 5 minutes before being mounted with ProLong™ Diamond Antifade

Mountant (Invitrogen™, P36970) or mounted with ProLong™ Diamond Antifade Mountant with

DAPI (Invitrogen™, P36962). Tissue sections were imaged on a CSU-W1 Spinning Disk/High

Speed Widefield microscope with a Plan-Apochromat 20x objective and the Andor Zyla 4.2

sCMOS camera for confocal imaging or the Andor DU-888 EMCCD camera at the UCSF Center

for Advanced Light Microscopy (CALM). On a section of tissue, ten random z-stack images

(2-22 sections at .92uM thickness) were collected using the same imaging specifications

respective to the target antibody. Fluorescence illumination was kept to a minimum to avoid

photobleaching. All images shown are single z-sections processed with ImageJ on the FIJI

platform(*127*) using the same processing parameters respective to the target antibody.

Tissue section nuclei count and total antibody fluorescence intensity were measured for every

z-section for each image with CellProfiler(*128*) using the same detection framework respective

to the organ and to the target antibody. A fluorescence intensity threshold was created for each

antibody to reduce background signal. Nuclei counts and total antibody fluorescence intensities

(arbitrary units [AU]) were summed across each image z-stack. The sum total antibody intensity

was normalized to the sum nuclei count to obtain total antibody intensity (AU)/nuclei for each of

the 10 images per tissue section. These 10 normalizations were then averaged to get the tissue

section average of total antibody intensity (AU)/nuclei. To obtain the species average of total

antibody intensity (AU)/nuclei, we averaged the tissue section averages by species. Results of

immunofluorescence experiments represent mean ± standard error of the mean (SEM) derived

from 3 individual insectivorous bats and 3 frugivorous bats (n = 3/phenotype). *P* values were

calculated using a mixed effects model with restricted maximum likelihood and Satterthwaite

approximation for degrees of freedom.

Differential expression and gene set enrichment analyses

We used Seurat function "FindMarkers" for calculation of differentially expressed genes. We used a threshold of average log base2 fold change > 0.25, adjusted p-values < 0.01, and minimum detected gene fraction (min.pct) > 0.25. For DEGs between species, we additionally filtered out genes that are not expressed in both species. The differentially expressed genes were visualized with volcano plots using R package ggplot2 version 3.4.0(*114*). We visualized selected genes on UMAP embeddings with Seurat function "FeaturePlot" and "VlnPlot" for violin plots. For gene set enrichment analysis using DEGs, we selected top genes with smaller counts of 100 or all DEG counts for each condition. We used R package EnrichR version 3.0(*129*) for stated pathway databases and *p*-value calculations. The gene set enrichment results were visualized with ggplot2 with adjusted *p*-values calculated with EnrichR.

Peak calling and enriched motif and GREAT analyses

ATAC data was added back into the integrated Seurat objects (**Supplementary Figure 3.1D**) following separation by species via the Seurat function "SplitObject()". Peaks were called for each cell-type by MACS2(*130*) with "--nomodel --extsize 200 --shift -100" and effective genome size respective to each bat genome. Cell-type-specific peaks were associated genes and human and mouse phenotypes with the Genomic Regions Enrichment of Annotations Tool (GREAT;(*68*). GREAT only takes human or mouse coordinates for input, therefore big brown bat peaks were lifted to the human genome *hg38(131)* using UCSC Genome Bioinformatics Group tools chainSwap and liftOver(*62*) with -minMatch=.1 and chain file "hg38.eptFus1.all.chain". To lift Jamaican fruit bat peaks to *hg38* for input into GREAT, a chain file was created between *hg38* chromosomes 1-22 and tandem repeat-masked Jamaican fruit bat genome via Lastz(*61*) and UCSC Genome Bioinformatics Group tools trfBig, mafToPsl and pslToChain(*62*). GREAT defines a basal regulatory domain for a gene as 5 kb upstream and 1 kb downstream of the

gene's TSS, regardless of other nearby genes, and extends this domain in both directions to the nearest gene's basal domain, but no more than 1,000 kb in one direction. We viewed outputs from "Significant By Region-based Binomial" view.

Differentially enriched TFBSs from cell-type-specific peaks were estimated from AME software within the MEME suite(*69*), setting each time the other species' cell-type-specific peaks as control sequences and default parameters. TFBS prediction on regulatory sequences was executed with FIMO software within the MEME suite(*69*) and filtered for *p*-value < .0000006 and *q*-value < .0003. For AME and FIMO motif inputs, we used the JASPAR 2022 core redundant dataset for vertebrates(*70*).

Diabetes SNP curation and intersection with scATAC-seq peaks

SNPs associated with T1D and T2D were downloaded from the NHGRI-EBI GWAS Catalog(*132*). These SNPs were filtered for noncoding regions and proxied for LD with SNiPA(*133*) and threshold r^2 > .8. The *hg38* coordinates for the resulting SNPs were intersected with the human genome coordinates of the bat cell-type-specific ATAC peaks. Bat scATAC-seq coverage over these human variants was examined by lifting the human genome coordinates back to bats using liftOver(*62*) and the chain files described in GREAT analyses. Hypergeometric p-values for SNP enrichment in bat ATAC peaks were calculated as follows: population size (N) = 611,564,276 noncoding SNPs from NCBI dbSNP 151(*134*); number of success in the population (M) = 10,059 or 9,700 noncoding dbSNPs overlapping bat kidney or pancreas peaks, respectively; sample size (s) = 3,460 T1D/T2D SNPs; number of successes (k) = 381 or 421 T1D/T2D SNPs overlapping bat kidney peaks or pancreas peaks, respectively.

Luciferase assays

To construct reporter plasmids containing bat genomic regions, DNA sequences were amplified from genomic DNA generated from *E. fuscus* and *A. jamaicensis* skeletal muscle, via the Easy-DNA™ gDNA Purification Kit (Invitrogen™ K180001), using primers listed in **Supplementary Table 3.3** and cloned into pGL4.11b (Promega) or pGL4.23[minP] (Promega). All sequences were validated by Sanger sequencing. Human kidney cells (HEK293T) were obtained from ATCC and maintained with DMEM high glucose (Sigma D5796) supplemented with 10% fetal bovine serum (FBS; ATCC 30-2020), 1% L-glutamine (Corning 25-005-CI) and 1% penicillin-streptomycin (Thermo Scientific™ 15140122). Human liver cells (HEPG2) were obtained from ATCC and maintained with EMEM (ATCC 30-2003) supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin (Thermo Scientific™ 15140122). 80 ng of reporter plasmids and 20 ng of Renilla luciferase plasmid pGL4.73 (Promega), to correct for transfection efficiency, were co-transfected into each cell line in each well of a 96-well plate using X-tremeGENE HP DNA Transfection Reagent (Sigma-Aldrich) according to the manufacturer's protocol. 48 hours post-transfection, cells were harvested, and reporter activity was measured using the Dual-luciferase reporter assay system (Promega) according to the manufacturer's protocol. Reporter activity was quantified using the Glomax 96-well plate luminometer (Promega). Relative firefly/Renilla luciferase values were determined, and results of luciferase assays represent mean ± SEM derived from two independent experiments of quadruplet measurements (n = 8). *P* values were calculated using Student's unpaired t-test.

3.4: Results

Integrative single-cell profiling of insect- and fruit-eating bat kidneys and pancreases

To characterize the cell-types, genes and regulatory elements that differ between insectivorous

and frugivorous bats, we conducted integrative single-cell sequencing (RNA-seq and ATAC-seq)

on the kidneys and pancreases of four adult male insectivorous big brown bats (*Eptesicus*

fuscus) (family: Vespertilionidae) and four adult male Jamaican fruit bats (*Artibeus jamaicensis*)

(family: Phyllostomidae). The families of these bats diverged approximately 53.8 million years

ago (**Figure 3.1A**)(*135*). We used these bat species as they have publicly available genomes

and are colonized in research labs. Big brown bats were fed a regular diet of mealworms in

captivity, whereas Jamaican fruit bats were fed a variety of non-citrus fruits in captivity, such as

cantaloupe and banana. We subjected these bats to an overnight fasting regime, followed by

two big brown bats fed fruit-fed mealworms (to maximize fruit content) and two fruit bats fed fruit

thirty minutes before euthanasia (**Figure 3.1B**). We chose this time point because fruit bats

digest food quickly and pass material within 30 minutes(*136*) and lower their blood sugar within

30 minutes(*3*, *17*, *19*, *20*). Tissues were harvested immediately and flash-frozen. Nuclei were

then isolated, subjected to fluorescence-activated cell sorting (FACS), and processed using the

10X Genomics Chromium single-cell Multiome ATAC + Gene Expression kit following

established protocols (see Materials and Methods).

As these bats are not widely used model organisms and have poorly assembled and annotated

genomes, we made several modifications to analyze their multimodal data. These include (see

Materials and Methods for more detail): 1) Removal of scaffolds < 50 kilo bases (kb) in length.

Removal of these short scaffolds still allowed us to capture > 90% of the total sequences and

genes for each genome (**Supplementary Figure 3.1A**). 2) Collapse of gene information in

annotation files so that each gene is represented by a single "exon" transcript. 3) Due to technical reasons, ATAC-seq can contain a large number of mitochondrial reads(*137*), whose removal is needed for scATAC-seq analyses. While the Jamaican fruit bat has an annotated mitochondrial genome, the big brown bat does not. We thus used GetOrganelle(*107*) for assembly and MITOS (Bernt et al. 2013) for annotation to generate a big brown bat mitochondrial genome. Reassuringly, all mitochondrial genes identified in the Jamaican fruit bat genome were detected in the big brown bat genome. 4) To allow comparison of scRNA-seq and scATAC-seq between the two species, we used Orthofinder(*109*), using only one-to-one orthologues for subsequent comparisons to increase the number of shared features by 3.11% across the big brown bat genome and 3.24% across the Jamaican fruit bat genome.

We used cellranger-arc 2.0 (10X Genomics) to generate raw ATAC and RNA counts and Seurat(*110*) and Signac(*111*) for quality control and downstream analysis (**Supplementary Figure 3.1B-C; Supplementary Table 3.1**). To merge samples, we created a common peak set across every sample within a species and within a tissue (**Supplementary Figure 3.1D**). To increase our sample sizes for phenotypic comparisons between species, we merged all samples within a species, hereinafter referred to as replicates. We combined two fasted and two fed insectivorous bat samples for each tissue (8,527 cells for kidney and 7,213 cells for pancreas), two fasted and two fed frugivorous bat samples for the kidney (9,315 cells), and two fasted and one fed fruit bat sample (due to one sample providing low quality sequencing data) for the pancreas (9,641 cells). To jointly analyze scRNA-seq and scATAC-seq within a species for each tissue, we used the R package Harmony(*113*) to correct for batch effects across replicates and applied "weighted nearest neighbor" analysis(*110*) (see Materials and Methods). To jointly analyze scRNA-seq and scATAC-seq across species for each tissue, we used gene activity scores to maximize the number of shared features for ATAC integration across species.

We then employed the same methods used for analyzing within species as for cross-species analysis (**Supplementary Figure 3.1D**; see **Methods**). Cell-types were identified using canonical mouse and human markers. As bats diverged from human and mouse lineages approximately 75 million years ago(*138*), and our study presents the first functional genomics and gene expression dataset for bat kidney and pancreas, the number of shared markers per cell-type identified between bat, human and mouse was limited, restricting our ability to detect novel cell-types that could be unique to bats. Of note, while we observed some significant differences in gene expression between fasted and fed states in each species (**Supplemental File 3.1-3.2**), these were minimal compared to species gene expression differences (**Supplementary Figure 3.2A-B**) (longer treatments are likely needed for many transcriptional differences between fasted and fed(*139*)), with the exception of acinar cells, which are known to dominate total mRNA population from the pancreas(*140*) and permit rapid exocytosis of enzymes for digestion(*141*). Given our small sample sizes for treatment (N=1-2/treatment), we focused on subsequent analyses on species differences (N=3-4/species). We also did not observe significant differences in cell compositions between fasted and fed states in each species by co-varying neighborhood analysis (CNA) (**Supplementary Figure 3.2C**)(*115*), which was expected as cell-type differences are not likely to transpire following a 30 minute treatment.

Insectivorous and frugivorous bat kidneys have different cell compositions

For the bat kidney, we initially annotated all major known cell-types using previously reported scRNA-seq markers from human and mouse kidney(*118*, *142*, *143*). These included proximal tubules (*CUBN*), connecting tubules (*SCNN1G*), distal convoluted tubules (*SLC12A3*), loop of Henle cells, thick ascending limbs (*SLC12A1*) and descending thin limbs (*SLC14A11*), podocytes (*NPHS2*), type A intercalated cells (*KIT*), type B intercalated cells (*SLC26A4*), principal cells (*AQP2*), mesangial cells (*PDGFRB*), endothelial cells (*PECAM1*), fibroblasts

(*FBN1*), and immune cells (*CD36*, *IL7R*, *CD74*) (**Figure 3.1C-G**). Comparison of our cell-type

annotations with automated annotations, using single-cell kidney reference databases from

humans (Azimuth) (**Figure 3.1H-I**)(*110*)(*117*) and mouse (**Supplementary Figure 3.3A-B**)(*118*),

showed high similarity. We also annotated a "proximal tubules-like" cell cluster, as this cluster

expresses the proximal tubule markers *MIOX*, *SLC34A1*, and *LRP2* but at lower levels

compared to the proximal tubules cluster (**Supplementary Figure 3.3C**). We confirmed

consistency of cell-type annotations before and after cross-species integration by observing

high similarity between the annotations determined separately for each species before

integration and the annotations determined jointly after integration (**Figure 3.1J**).

We detected several cell composition differences between frugivorous and insectivorous bat

kidneys. Between the two bats, the insectivorous bat renal epithelial cell distribution most

closely resembles that of the mouse (**Supplementary Figure 3.3D**)(*144*). In fruit bats, we found

fewer thick ascending limbs (TAL) and significantly fewer descending thin limbs (DTL) cell-types

(**Figure 3.1G**), which make up the loop of Henle and are largely found in the renal medulla. The

loop of Henle cluster was correlated with the insectivorous bat by CNA (**Supplementary Figure

3.3E**). Although TAL was not significantly more abundant in the insectivorous bat by Wilcoxon

rank-sum test, likely due to small sample size, a Chi-square test of independence indicates high

confidence of TAL enrichment in insectivorous bats (**Supplementary Figure 3.3F**). This is in

line with previous reports that showed that fruit bats have a larger renal cortex and a smaller

renal medulla compared to insectivorous bats(*22*, *35*, *145*). Additionally, we observed fruit bats

to have significantly more type A intercalated cells, which are involved in acid secretion into the

urine, and type B intercalated cells (**Figure 3.1G, Supplementary Figure 3.3F**), which mediate

bicarbonate secretion while reabsorbing sodium chloride(*146*), fitting with the high amounts of

bicarbonate and the low amounts of sodium in fruit, which has a negative risk of renal acid

load(*147*). The fruit bat also has significantly more principal cells, which reabsorb sodium and excrete potassium, than the insectivorous bat (**Figure 3.1G, Supplementary Figure 3.3F**), in line with fruit containing low sodium and high potassium levels. All of these collecting duct cell-type clusters correlate with the fruit bat in CNA (**Supplementary Figure 3.3E**). The fruit bat also has significantly more connecting tubules, which are largely found in the renal cortex, and this cluster correlates with the fruit bat by CNA (**Figure 3.1G, Supplementary Figure 3.3E-F**). Connecting tubules together with late distal convoluted tubules and the cortical collecting duct (principal cells, type A and B intercalated cells) are often called the aldosterone-sensitive distal nephron (ASDN). Aldosterone increases sodium reabsorption and promotes potassium secretion in the final step of the renin-angiotensin-aldosterone system (RAAS)(*148*). As hyperkalemia (excessive potassium in blood) stimulates aldosterone release(*148*), the greater abundance of connecting tubules in fruit bats is in line with a high potassium diet. The insectivorous bat has significantly more proximal tubule-like cells as compared to the fruit bat, and this cluster also shows significant association with insectivorous bats (**Figure 3.1G, Supplementary Figure 3.3E-F**). Decreased proximal tubule count is also observed in human diabetic kidneys(*149*), although proximal tubular growth occurs in early diabetic nephropathy(*150*). In summary, we find that frugivorous and insectivorous kidneys differ in many nephron components, particularly in proximal tubule-like cells, connecting tubules, the loop of Henle and the collecting duct.

To further validate these cell composition differences, we performed immunofluorescence on various cell-type specific markers on kidney tissue sections from adult big brown bats and Jamaican fruit bats (**Supplementary Table 3.4**). For type B intercalated cells, we used an antibody against pendrin, encoded by *SLC26A4*, finding significantly more pendrin-expressing cells in fruit bats than in insectivorous bats (**Figure 3.2A-B**). For TAL, we used an antibody

against the $Na^+/K^+/2Cl^-$ co-transporter NKCC2, encoded by *SLC12A1*, finding significantly fewer NKCC2-expressing cells in fruit bats than in insectivorous bats (**Figure 3.2C**). For principal cells, we used an antibody against aquaporin 2, encoded by *AQP2*, finding significantly more aquaporin 2-expressing cells in fruit bats than in insectivorous bats (**Supplementary Figure 3.4A-B**). Combined, our immunofluorescence results validate the single-cell composition differences observed for these cell-types.

Gene expression differences between insectivorous and frugivorous bat kidneys

We next analyzed our scRNA-seq datasets to identify molecular changes associated with adaptation to these dietary differences. We conducted gene ontology (GO) enrichment analyses for the genes that showed differential expression between frugivorous and insectivorous bat kidneys (**Supplemental File 3.3-3.4**). As there are significant cell-type composition differences in the kidney for collecting duct cells, we first focused on enriched terms identified in these cells. For type B intercalated cells, the top KEGG 2021 Human Pathway term for fruit bats was "collecting duct acid secretion", involving the components of vacuolar H^+-ATPase (**Figure 3.2D**). For example, ATPase H^+ transporting V1 subunit G3 (*ATP6V1G3*) was highly expressed in type B intercalated cells in fruit bats but not in insectivorous bats (**Figure 3.2E**). Reabsorption of sodium in type B intercalated cells is fueled by H^+-ATPase(*146*), so fruit bats most likely require more vacuolar H^+-ATPase to reabsorb scarce sodium in their diet. For principal cells, the top GO Biological Process terms were all related to maintaining sodium-potassium balance (**Supplementary Figure 3.4C-D**), which includes fruit bat upregulated lysine-deficient protein kinases *WNK1* and *WNK4* (**Supplementary Figure 3.4E**). The WNK signaling pathway is regulated by sodium intake via aldosterone as well as insulin(*151*), and mutations in humans in both these genes lead to hyperkalemia and hypertension(*152*). Combined, our single-cell RNA-seq analyses reveal important gene expression adaptations to high dietary glucose and

potassium and low dietary sodium and calcium that occurred within the collecting duct, many of which are associated with human diseases such as hyperkalemia and hypertension.

For type A intercalated cells, the top KEGG 2021 Human Pathways for fruit bats were "endocrine and other factor-related calcium reabsorption" and "renin secretion" (**Figure 3.2F**). These genes included kallikrein 1 (*KLK1*), an endocrine-responsive gene that is induced by high glucose(*153*), which was found to be highly expressed in fruit bat type A intercalated cells but not in insectivorous bats (**Figure 3.2G**). Interestingly, mutations in this gene or its promoter have been found to be associated with hypertension in humans, and its overexpression in mice leads to hypotension, protection from diabetic cardiac damage, renal fibrosis and renal vasodilation, and increased nitrate levels in urine(*154*), which are found naturally in high concentrations in plants(*155*). Furthermore, *KLK1* protects against hyperkalemia after a high potassium load in mice(*156*). Another highly expressed gene in the fruit bat type A intercalated cells that was not expressed in the insectivorous bat was the potassium calcium-activated channel subfamily M alpha 1 (*KCNMA1*) (**Figure 3.2G**). This gene is responsive to renin-mediated regulation of body potassium levels and localizes to the apical membrane of these cells under a high potassium diet to secrete potassium(*146*). Furthermore, high potassium stimulates the synthesis and function of the *KCNMA1*-encoded transporter(*146*), fitting the lack of expression in insectivorous bats. Increased expression of *KLK1*, *KCNMA1*, and vacuolar H^+-ATPase in type A intercalated cells in fruit bats is likely an adaptation to fruit specialization that helps to address high glucose and high potassium intake.

To determine if TAL gene expression differed substantially between bat species, we used the STRING database(*157*) to identify genes related to *SLC12A1* and analyzed their expression in our datasets. We found that *STK39* (SPAK), *OXSR1* (OSR1), *KCNJ1* (ROMK), and *SLC9A3*

(NHE3), key genes supporting and regulating TAL function in urine concentration(*158*, *159*), showed significantly lower expression in fruit bats (**Figure 3.2H-J**). *Stk39* null mice, as well as *Stk39* and *Oxsr1* double knockouts, have increased potassium in the urine, in line with the higher potassium levels observed in fruit bats(*160*). Mouse kidney tubule *Oxsr1* knockout, *Kcnj1* knockout, and *Slc9a3* conditional TAL knockout all show lower urine osmolality(*161–163*), fitting with their observed lower gene expression and the dilute urine found in fruit bats(*106*). Loss of function mutations in *SLC12A1* or *KCNJ1* in humans causes Bartter syndrome, which is characterized by hypokalemia, metabolic alkalosis (body pH elevation), polyuria (excess urination), salt wasting, hypercalciuria, and hypotension(*164–166*). Combined, our single-cell analyses find that both TAL abundance and gene expression changed substantially in fruit bats relative to insectivorous bats.

We also analyzed our scRNA-seq data for expression of kidney transporters *SLC22A6* (OAT1), *SLC22A12* (URAT1), *SLC2A9* (GLUT9), and *RHBG*, which were reported to be lost in OWFBs(*7, 15*) but intact in NWFBs (Jamaican fruit bat and Honduran yellow-shouldered fruit bat, *Sturnira hondurensis*)(*7*), and for *MYO6* (myosin VI), which is highly expressed in the OWFB kidney(*47*). We found that *SLC22A6*, which is involved in the excretion of organic anions, and *SLC22A12,* which is involved in renal reabsorption of urate, show significantly higher expression in fruit bat proximal tubules and proximal tubule-like cells compared to those of the insectivorous bat (**Supplementary Figure 3.4E**). *SLC22A6* has extremely broad substrate specificity, for both endogenous and exogenous substrates(*167*), so in NWFBs it could possibly still have a role in the excretion of toxins. Higher expression of *SLC22A12* might be necessary to maintain urate homeostasis in fruit bats, as urate comes from endogenous purines and from diet(*168*), and fruits are low in purines(*169*). *SLC2A9*, which also mediates urate reabsorption in the kidney, was found to be highly expressed in type A intercalated cells in the insectivorous bat

and showed no expression in the frugivorous bat in these cells (**Figure 3.2G**). Lack of *SLC2A9* expression in the collecting duct cells in fruit bats could support urate reabsorption from a high glucose diet, as high glucose concentration inhibits urate reabsorption by this transporter(*170*). *RHBG*, which encodes an ammonium transporter, was expressed in type A intercalated cells in insectivorous bats but largely absent in fruit bats (**Figure 3.2G**). Ammonium inhibits sodium reabsorption and potassium secretion(*171*), so reduced expression of *RHBG* would accommodate the low sodium-high potassium dietary intake of a frugivore. *MYO6* was upregulated in fruit bat podocytes (**Supplementary Figure 3.4E**). Adaptive evolution of *MYO6* in OWFBs was hypothesized to support protein preservation from a protein-scarce diet(*47*), as mice lacking *MYO6* had reduced endocytosis-mediated protein absorption in proximal tubules and also hypertension(*172*), which may similarly be supported by upregulation of *MYO6* in NWFB kidney. Together, these results suggest that kidney adaptations are not entirely synonymous across fruit bat lineages.

Bats have been evaluated as a model for diabetes due to their range in lifestyles, diet, and genetic factors(*3*). Features of diabetic kidneys include increased gluconeogenesis and glucose reabsorption(*173*). We scanned our datasets for gene expression signatures for these processes. We found gluconeogenesis genes *G6PC*, *FBP1*, *ALDOB*, and *PCK1(173)* to all be upregulated in fruit bat proximal tubules and proximal tubule-like cells, as well as the glucose reabsorption transporter *SLC5A2* (SGLT2)(*174*), along with other sodium/glucose cotransport family SLC5 genes *SLC5A10* (SGLT5), *SLC5A6*, and *SLC5A12* (**Figure 3.2J-K**). Increased potassium secretion is also a signature of diabetic nephropathy(*149*), and we observed many gene expression changes in fruit bat kidneys that promote potassium secretion (*KCNMA1*, *STK39*, *OSR1*, *RHBG*) (**Figure 3.2G,I**). Together, our gene expression data indicates that the fruit bat kidney shares many characteristics of a diabetic kidney.

Gene regulatory differences between insectivorous and frugivorous bat kidneys

To analyze our scATAC-seq datasets and identify gene regulatory differences between insectivorous and frugivorous bat kidneys, we used MACS2(*130*) to call 209,895 and 188,845 cell-type-specific peaks for these bat kidneys, respectively (**Supplemental File 3.5-3.6**). To associate open chromatin regions to genes and human phenotypes, we converted our bat peaks to human genome coordinates (*hg38*)(*131*), converting 141,779 insectivorous and 124,845 frugivorous bat peaks to *hg38*, respectively. We then used the Genomic Regions Enrichment of Annotations Tool (GREAT)(*68*) to characterize the target genes associated with these peaks. The fruit bat kidney was found to be highly enriched for human kidney phenotypes, including "glucose intolerance", "metabolic alkalosis", "abnormality of renin-angiotensin system", "abnormal circulating renin", and "hypokalemia" (**Supplementary Figure 3.4F**), fitting with our cell-type and scRNA-seq analyses that showed that the fruit bat kidney resembles a diabetic kidney. The insectivorous bat kidney was only enriched for the human kidney phenotype "enuresis nocturna" (involuntary urination) (**Supplementary Figure 3.4F**).

Next, to investigate differences in *cis*-regulation between frugivorous and insectivorous bats, we performed differential motif enrichment analyses on renal epithelial cells using AME(*175*). For collecting duct cells, we found two clusters of differentially enriched motifs, which separated frugivorous and insectivorous bats (**Figure 3.3A**). Insectivorous bats were found to be enriched for *NKX* and *NFAT* motifs, and frugivorous bats were enriched for *ONECUT* motifs. *NKX* TFs are broadly expressed and important for organ development, and *NFAT* TFs control proteinuria(*176*), fitting with the high protein diet of insectivores. *NFAT5* upregulates ubiquitin ligase under hypertonic conditions in the collecting duct to protect renal medullary cells from apoptosis(*177*), which is also fitting for the low water insectivores obtain from their diet, as compared to frugivores, and the upregulation of *NFAT5* in insectivore principal cells and type A

intercalated cells (**Supplementary Figure 3.4G**). *ONECUT1* is known to antagonize glucocorticoid-stimulated gene expression(*178*), regulate glucose metabolism(*179*), and mutations in this gene are associated with diabetes in humans(*180, 181*). Glucocorticoids stimulate expression of gluconeogenesis genes in the diabetic kidney(*149, 173*), fitting with the observed *ONECUT* motif enrichment which could allow tighter regulation of gluconeogenesis in response to fluctuating high glucose intake in frugivores. We also separated proximal tubules, proximal tubule-like cells, distal convoluted tubules, TAL and DTL for differential motif enrichment and found multiple TFBS clusters that differentiate frugivorous and insectivorous bats (**Figure 3.3A**). The *NFAT5* motif was again enriched in insectivore tubules, and we also saw upregulation of *NFAT5* in insectivore proximal tubules and distal convoluted tubules (**Supplementary Figure 3.4G**). We found the *KLF9* motif to be enriched in fruit bat renal tubules (**Figure 3.3A**), which was also enriched in diabetic proximal tubules(*173*). *KLF9* is a potential target for the mineralocorticoid receptor in mouse distal convoluted tubules(*182*), which mediates fluid and electrolyte balance(*183*) and undergoes non-ligand activation with elevated glucose(*184*). The *KLF9* motif was also enriched in fruit bat collecting duct cells (**Figure 3.3A**). In summary, kidney TFBS differential enrichment between insectivorous and frugivorous bats identified key TFs involved in diet, with frugivorous bats demonstrating diabetic-associated motif signatures.

We next sought to utilize our integrative single-cell datasets to survey diabetes-associated regions in bat kidneys. Examination of *KLK1*, which regulates blood pressure, showed a substantial increase in chromatin accessibility in the collecting duct cells as well as proximal tubules, TAL, and podocytes in the fruit bat (**Figure 3.3B**). *ATP6V1G3*, which is a biomarker for diabetic nephropathy(*185*), showed greater chromatin accessibility, particularly in its promoter, in fruit bat type A and type B intercalated cells (**Supplementary Figure 3.4H**). The insectivore

ATP6V1G3 promoter is predicted to bind *RREB1*, which represses RAAS through the angiotensin gene and is associated with end stage diabetic renal disease(*186*). Angiotensin is known to stimulate vacuolar H⁺-ATPase activity in intercalated cells(*187*), so the lack of the *RREB1* motif in the fruit bat *ATP6V1G3* promoter may allow greater activation of RAAS effects. In addition, *RREB1* expression is lower in fruit bat proximal tubules and proximal tubule-like cells (**Supplementary Figure 3.4G**). We also observed similar chromatin accessibility trends between human diabetic proximal tubules and controls(*173*) in bats (**Supplementary Figure 3.5A-D**). For example, *PCK1* demonstrated decreased accessibility near the end of the gene body in diabetic proximal tubules(*173*), which was also observed in the fruit bat kidney (**Supplementary Figure 3.5A**). The fruit bat sequence homologous to the insectivorous bat peak following the end of the *PCK1* gene body is predicted to bind transcriptional repressor *ZNF331(188)*, fitting with the decreased accessibility of this region in fruit bat proximal tubules. Together, our analyses indicate that fruit bats evolved many features similar to a human diabetic kidney.

We next compiled a list of noncoding SNPs associated with type 1 diabetes (T1D) and T2D from the NHGRI-EBI Genome-Wide Association Study (GWAS) catalog(*132*). We then filtered this list with SNiPA(*133*) for noncoding single nucleotide polymorphisms (SNPs) that are in linkage disequilibrium (LD) with nearby variants ($r^2 > 0.8$) (**Supplemental File 3.7**). We then determined the overlap of 3,460 noncoding SNPs with bat scATAC peaks in human genome coordinates, finding 381 overlaps (hypergeometric p-value < .0001) (**Supplemental File 3.8**), with the majority of these SNPs (~90%) associated with T2D, including hits near *INSR(189)*, *PIK3R1(190, 191)*, and *MAP3K1(192–194)*. More than half of the SNPs overlapped peaks in proximal tubule and tubule-like cells and thick ascending limbs, and over a third of the SNPs overlapped peaks in distal convoluted tubules, principal cells, and type A intercalated cells. Our

cell-type-specific regulatory profilings allowed us to investigate human disease variants that are difficult to study without cell-type-specific data, such as rs79687284 at the *PROX1* locus. This variant is highly associated with T2D and was predicted to have regulatory function as an enhancer, but it did not demonstrate transcriptional activity in luciferase reporter assays in a pancreatic beta cell line that expresses *PROX1(195)*. We observed an increase in chromatin accessibility at this region in insectivorous bats and corresponding gene upregulation in insectivore distal convoluted tubules (**Figure 3.3C**). *Prox1* heterozygous knockout (lower-expressing) mice have increased circulating insulin and glucagon and adult-onset obesity(*196*), and deletion of this gene in distal convoluted tubules causes hypomagnesemia(*197*), which is associated with progression of T2D(*198*, *199*), and fitting with the magnesium-rich diet of insectivores(*200*). In summary, our scATAC-seq datasets allowed us to identify candidate gene regulatory elements in bats that overlap with T1D and T2D GWAS SNPs.

Insectivorous and frugivorous bat pancreases have different cell-type composition
We next set out to analyze our multiome data for the bat pancreas. Similar to the kidney, we initially annotated all major cell-types using markers from previously reported scRNA-seq datasets from human and mouse pancreases(*119*, *124*, *142*). We detected all major known cell-types found in human and mouse pancreas: acinar cells (*CPA2*), ductal cells (*SLC4A4*), beta cells (*INS*), alpha cells (*GCG*), delta cells (*SST*), endothelial cells (*SLCO2A1*), active pancreatic stellate cells (aPSCs; *LAMA2*), quiescent pancreatic stellate cells (qPSCs; *COL3A1*), and immune cells (*ACTB*, *CD74*, *IL7R*, *MRC1*) (**Figure 3.4A-E**). We identified gamma cells and epsilon cells by other known markers (*CHGA*;(*201*) and *ACSL1(119)*, respectively), as many major markers were not shared across species. We observed high similarity between our cell-type annotations and automated annotations using human single-cell pancreas reference

databases (**Figure 3.4F-G, Supplementary Figure 3.5A-B**)(*110, 119*);(*120–125*). Consistent

with the adult human pancreas, acinar cells showed clear heterogeneity in the bat pancreas,

with secretory acinar cells (acinar-s) expressing higher levels of digestive enzyme genes (*CPA1,*

PLA2G1B, SYCN, KLK1, CLPS) and idling acinar cells (acinar-i) expressing lower levels of

digestive enzyme genes (**Supplementary Figure 3.6C**). We compared the annotations

determined separately for each species before integration and the annotations determined

jointly after integration and observed high similarity (**Figure 3.4H**).

Fruit bats are reported to have an expansion of endocrine tissue in the pancreas(*75, 103*).

Indeed, of the pancreatic cell-types, we found that the fruit bat pancreas is about 26%

endocrine, whereas the insectivorous bat pancreas is about 12% (**Supplementary Figure**

3.6D). Both beta and alpha cell clusters are largely correlated with the fruit bat by CNA

(**Supplementary Figure 3.6E**). As the smaller frugivorous bat sample size (N=3) precluded cell

composition significance between species with Wilcoxon rank-sum test, we also conducted

Chi-square tests of independence on Pearson residuals. We found, with high confidence, that

both beta and alpha cells are more abundant in the fruit bat (**Supplementary Figure 3.6F**).

Beta and alpha cells regulate blood glucose levels via insulin and glucagon, respectively, and

their increased numbers are in line with the need to respond to a high glucose diet and tight

blood glucose regulation in fruit bats, which have been shown to robustly regulate blood sugar

in intraperitoneal glucose tolerance tests(*3, 17, 19, 20*). We found acinar cells to be more

abundant in the insectivorous bat (**Supplementary Figure 3.6E**), with fruit bats having fewer

acinar cells than insectivorous bats with high confidence (**Supplementary Figure 3.6F**). Acinar

cells, which produce digestive enzymes for storage and secretion(*202*), comprise about 75% of

insectivorous bat pancreatic cells, which is largely consistent with that of the human adult

pancreas(*119*), compared to 52% in fruit bats (**Supplementary Figure 3.6D**). The substantial

reduction in acinar cells in fruit bats could accommodate the increase of endocrine cells. The fruit bat also has more ductal cells with high confidence (**Supplementary Figure 3.6F**), which secrete enzymes from acinar cells into the duodenum and secrete bicarbonate to neutralize stomach acidity(*203*), although the overall percentage of ductal cells in the fruit bat pancreas (11%) corresponds with that of humans (10%)(*203*). In summary, our results suggest that compared to the insectivorous bat, the expanded endocrine pancreas in the fruit bat is largely attributable to increased numbers of beta and alpha cells, with proportionately lower exocrine acinar cell numbers that likely compensate for the endocrine expansion.

To validate our single-cell composition results, we performed immunofluorescence on pancreas tissue sections from the same bat species (**Supplementary Table 3.4**). For beta cells, we used an antibody against insulin, encoded by *INS*, and found that there were significantly more insulin-expressing cells in fruit bats than in insectivorous bats (**Figure 3.5A**). For alpha cells, we used an antibody against glucagon, encoded by *GCG*, and found that there were also significantly more glucagon-expressing cells in fruit bats than in insectivorous bats (**Figure 3.5B**). Our immunofluorescence results thus validated our observed single-cell composition differences for endocrine cells.

Gene expression differences between insectivorous and frugivorous bat pancreases
To identify pancreatic gene expression differences associated with frugivory, we performed gene enrichment analysis on all cell-types using differentially expressed genes between frugivorous and insectivorous bats (**Supplemental File 3.9-3.10**). As acinar-i cells appeared largely separated by species according to CNA (**Figure 3.4E, Supplementary Figure 3.6E**), we did sub-clustering to analyze phenotypic differences within this cell-type (**Supplementary Figure 3.7A**), and found that cluster 0 (C0) was predominantly insectivorous bat-based and cluster 1

71

(C1) to be predominantly fruit bat-based. C0 was enriched for KEGG 2021 Human Pathway terms related to protein synthesis and secretion, fitting with the increased protein composition in insects. C1 was enriched for "diabetic cardiomyopathy" as well as "oxidative phosphorylation" (**Supplementary Figure 3.7B-C**), which is notably shared with human T2D acinar cells(*122*). Ductal cells, which produce alkaline-high pancreatic secretions and are regulated by calcium signaling(*204*), were enriched in fruit bats for "calcium signaling pathway" genes, including *PDE1A*, *EGFR*, and *ERBB4* (**Supplementary Figure 3.7D-F**). Fruit bats may stimulate more pancreatic secretions due to their low calcium diet to quickly digest carbohydrates and rapidly fuel their metabolism(*18*).

We next determined if genes suggested to be adaptive for frugivory in OWFBs were differentially expressed between our NWFBs and insectivorous bats. *FAM3B* and *FFAR3*, metabolic genes involved in insulin metabolism and signaling, were reported to be lost in OWFB genomes(*7*, *15*). We observed both genes to be expressed in the NWFB pancreas (**Supplementary Figure 3.7F**). Expression of *FAM3B* was sparse throughout the NWFB pancreas as well as the insectivore pancreas. *FFAR3*, an inhibitor of insulin secretion(*205*, *206*), was expressed in insectivorous bat beta cells and was nearly absent in fruit bats (**Supplementary Figure 3.7F**). Weak expression of *FFAR3* in NWFB pancreas is consistent with the hypothesis that loss of this gene in OWFBs is adaptive, as fruit bats secrete large amounts of insulin(*75*).

We then analyzed the expression of glucose transporters (GLUTs) in pancreatic cells. We first examined the main glucose transporter in beta cells(*207*), *SLC2A2* (GLUT2), which was previously shown to have a fruit bat-specific 11 bp deletion in its proximal promoter(*3*), and found that it was not differentially expressed between our insectivorous and frugivorous bats

(**Supplementary Figure 3.7F**). Other GLUTs identified in human pancreatic cells also did not

show differential expression between insectivorous and frugivorous bats, with the exception of

SLC2A13 (GLUT13)*,* a H+/myo-inositol transporter found in both human endocrine and exocrine

cells(*207*). *SLC2A13* showed significantly higher expression in fruit bat exocrine acinar-i and

ductal cells (**Supplementary Figure 3.7F**). Myo-inositol is found in fruits like cantaloupe, which

was consumed by our fruit bats, and is produced in tissues such as the kidneys(*208*). Increased

sugar intake and diabetes increases the need for myo-inositol(*208*). Increased expression of

SLC2A13 in fruit bat exocrine pancreas complements the upregulation of myo-inositol

oxygenase enzyme (*MIOX*), which catalyzes myo-inositol breakdown, in the liver of the same

fruit bat species we analyzed (Jamaican fruit bat) but not in four insectivorous bat species(*54*).

In line with these findings, we also observed upregulation of the myo-inositol transporter

SLC5A10(209) in the fruit bat kidney proximal tubules (**Figure 3.2K**).

Features of a diabetic pancreas include a reduced exocrine pancreas and loss of beta cells,

with differences in the presence of insulitis and islet amyloidosis between T1D and T2D(*210*).

We compared our gene expression data to reported single-cell datasets from the diabetic

human pancreas(*122*, *123*, *211*). 60 differentially expressed, protein-coding genes between

human T2D beta cells and normal beta cells, determined from three separate single-cell studies

with a combined 19 T2D donors and 33 non-diabetic donors(*212*), included 6 (out of 1,808)

genes that were also differentially expressed in our datasets between insectivorous and

frugivorous bat beta cells: *TAF1*, *PTPRD*, *NOXA1*, *MED13*, *AFF1*, and *MEST* (**Figure 3.5C**).

PTPRD is involved in insulin signaling and variants within this gene were found to be associated

with gestational diabetes risk(*213*, *214*). This gene was found to be upregulated in T2D beta

cells and in our fruit bat beta, alpha, and ductal cells. *AFF1*, which has circular RNAs that

control beta cell apoptosis(*215*), was also upregulated in both T2D and fruit bat beta and acinar

cells. Chronic hyperglycemia is known to upregulate NAPDH oxidase (NOX) genes(*216*), and *NOXA1* (NOX activator 1) was found to be upregulated in our fruit bat beta, alpha, and acinar-i cells. We also observed the monogenic diabetes risk gene *ABCC8*, a regulator of potassium channels and insulin release, whose loss of function causes hyperinsulinism (excessive insulin secretion) and polyuria(*217*), to be downregulated in fruit bat beta cells. In addition, the insulin receptor substrate-1 (*IRS1*), whose deletion in mice causes hyperinsulinism but not diabetes(*218*), was found to be downregulated in fruit bat beta cells and upregulated in fruit bat alpha and acinar-i cells. Taken together, the fruit bat pancreas exhibits gene expression changes that are known to elevate and tightly regulate insulin secretion and signaling in response to a high sugar diet, many of which correspond to human diabetic gene expression dysregulation.

Gene regulatory differences between insectivorous and frugivorous bat pancreases

To analyze open chromatin regions specific to cell-types in the pancreas, we called peaks with MACS2(*130*), finding 187,421 and 281,823 cell-type-specific peaks in insectivorous and frugivorous bat pancreases, respectively (**Supplemental File 3.11-3.12**). To associate open chromatin regions to genes and human and mouse phenotypes, we identified 111,619 and 184,360 peaks from the insectivorous and frugivorous bat genomes, respectively, in humans (*hg38*). Analysis of these human peaks using GREAT(*68*) found the insectivorous bat pancreas to be enriched for many mouse phenotype terms involving fat metabolism, including "abnormal lipolysis", "impaired lipolysis", "increased inguinal fat pad weight", "increased white fat cell number", and "abnormal white fat cell number" (**Supplementary Figure 3.7G**), which could be indicative of adaptation to a higher fat diet from insects. The fruit bat pancreas was most enriched for human pancreas phenotype "metabolic alkalosis" (**Supplementary Figure 3.7G**), indicative of the high bicarbonate diet of frugivores.

We then examined differentially enriched TFBSs in scATAC-seq peaks between insectivorous and frugivorous bats in endocrine and exocrine cell-types. We found that there are two clusters of TFBSs that show differential enrichment across alpha and beta cells in the two bat genomes, resulting in separation of dietary phenotypes (**Figure 3.5D**). Insectivorous bats were enriched for *TBX* and *MEF2* motifs, TFs with important roles in organ and embryonic development, respectively, and *HNF1B*, which controls endocrine precursor generation and when mutated is associated with maturity-onset diabetes of the young (MODY)(*219*). The enrichment for *HNF1B* motifs in insectivorous bats over fruit bats indicates differential regulation of beta cells in fruit bats. Fruit bats were enriched for *KLF* motifs. *KLF* TFs regulate functions and pathophysiology of the digestive system, many of which maintain beta cell function and control oxidative stress response genes(*220*). Two clusters of TFBSs were also annotated for exocrine cell-types, indicating again that there is a divergence in TFBS usage between these two bat species (**Figure 3.5D**). Insectivore exocrine cells were enriched for *MAFA* and *NEUROD1* motifs, TFs that regulate insulin gene expression and beta cell development, respectively(*221*). *NEUROD1* was not found to be differentially expressed between bat exocrine cells, but was found to be upregulated in fruit bat alpha, beta, delta, and gamma cells (**Supplementary Figure 3.7F**). *NEUROD1* can switch cell fates from endocrine to exocrine in response to NOTCH signaling(*222*), so the lack of enrichment for *NEUROD1* in fruit bat exocrine cells may promote endocrine precursor cells and prevent exocrine cell development, consistent with the differences in cell abundances. Fruit bat ductal cells and acinar-i cells showed a motif enrichment for *USF1*, which is involved in the regulation of genes in response to high glucose including insulin(*223*) and the regulation of diabetic kidney disease(*224*), as well as for *ARNT*, which has reduced expression in diabetic human islets and its removal in mouse beta cells results in abnormal glucose tolerance and insulin secretion(*225*). Both *USF1* and *ARNT* were upregulated in fruit

bat acinar-i cells (**Supplementary Figure 3.7F**). Taken together, fruit bats exhibit TFBS usage

indicative of unique glucose and insulin regulation.

Next, we examined specific diabetes-associated regions for gene regulatory differences

between insectivorous and frugivorous bats. We found more open chromatin in the promoter of

INS, which was more highly expressed in fruit bat beta cells (**Figure 3.5E, Supplementary**

Figure 3.7F,H). In the promoter of *GCG*, as well as in a region downstream of *GCG*, we

observed increased open chromatin in fruit bat alpha cells (**Figure 3.5F, Supplementary Figure**

3.7I), though *GCG* was not differentially expressed between these species (**Supplementary**

Figure 3.7F). The *INS* promoter in fruit bats is predicted to bind *NR2C2* and *USF2* (**Figure**

3.5E). *NR2C2* is known to have a role in beta cell regulation and mice lacking *NR2C2* have

hypoglycemia(*226*) and increased oxidative stress(*227*), while *USF2* expression is stimulated by

high glucose and is known to control the synthesis of insulin(*223*, *228*, *229*). There were many

high-scoring motif occurrences for the *INS* promoter in insectivorous bats, including *PRDM9*,

which has a differentially enriched motif in fruit bat endocrine cells (**Figure 3.5D**), *GLI2*, which

can serve as a transcriptional repressor in embryonic development(*230*), and transcriptional

repressor *ZNF263(231)* (**Figure 3.5E**). The *GCG* promoter in fruit bats is predicted to bind

RREB1 (**Figure 3.5F**). *RREB1* is known to potentiate the transcriptional activity of *NEUROD1*,

fitting with the upregulation of this TF in fruit bat endocrine cells (**Supplementary Figure 3.7F**).

The insectivore *GCG* promoter is predicted to bind *NRL* (**Supplementary Figure 3.7I**), a motif

that was also identified in human *GCG(232)* and differentially enriched in insectivore endocrine

cells (**Figure 3.5D**). The fruit bat peak downstream of *GCG* is predicted to bind *FOXA1* (**Figure**

3.5F), which regulates alpha cell differentiation, glucagon synthesis and secretion(*233*). The

homologous sequence in insectivorous bats is predicted to bind *POU2F2* (**Figure 3.5F**)*, a motif

that was also differentially enriched in insectivore endocrine cells (**Figure 3.5D**). The lack of

differential expression of *GCG* between bats with different diets may be attributed to other *cis-* or *trans*-acting regulators. Of note, however, our scATAC-seq analyses found fruit bats to have both *GCG* and *INS* proximal and distal *cis*-regulatory regions that show a more open chromatin state compared to insectivorous bats (**Figure 3.5E-F, Supplementary Figure 3.7H-I**).

We also intersected our T1D and T2D list of 3,460 noncoding SNPs, and overlapped them with bat pancreas scATAC peaks, and found 421 overlaps (hypergeometric p-value < .0001; >90% from T2D) (**Supplemental File 3.13**). More than half of the SNPs overlapped peaks in acinar cells, and over a third of the SNPs overlapped peaks in beta, ductal, and alpha cells. These overlaps included variants at loci highly associated with human diabetes, including *FTO(234, 235)*, *IGF2BP2(236, 237)*, *CMIP(238, 239)*, and *VPS13C(240, 241)*. In addition, we are able to identify peak overlap with novel human diabetes variants like rs1947178, which is located 159,430 bp downstream of the *TOX* transcription start site (TSS) and is predicted to increase chromatin accessibility and gene expression of *TOX*, a transcriptional regulator in T1D(*242*). We observed increased accessibility in fruit bats in this region, along with upregulation of *TOX* in fruit bat beta, alpha, acinar-i and acinar-s cells (**Supplementary Figure 3.8A**). As *TOX* was also associated with diabetic nephropathy(*243*), we examined this locus in our kidney datasets and also observed increased accessibility at rs1947178 in fruit bat principal cells, combined with upregulation of *TOX* in these cells (**Supplementary Figure 3.8B**). In summary, our bat integrative single-cell datasets allowed us to identify gene regulatory elements and variants that are associated with diabetes in humans.

Single-cell regulatory differences between bats are not recapitulated in reporter assays

To identify and functionally characterize cell type-specific gene regulatory regions between insectivorous and frugivorous bats, we selected candidate regulatory regions to compare using

luciferase reporter assays. For the kidney, we amplified homologous regions between bats spanning scATAC peaks from genomic DNA (see **Methods**) for promoters (PRO) of *ATP6V1G3*, *KCNMA1*, *KLK1*, and *SLC12A3* and for a candidate *cis*-regulatory element (CRE) downstream of the *KLK1* gene (**Figure 3.3B**). For the pancreas, we amplified homologous regions between bats spanning scATAC peaks from genomic DNA for promoters of *INS* and *GCG* and for a candidate CRE downstream of *GCG* (**Figure 3.5F**). Promoter sequences were amplified to capture as much of the scATAC peak without including transcription start sites (TSSs) to prevent reporter transcription interference. Promoters were cloned into a promoter assay vector (pGL4.11b; Promega) (Empty vector), which contains the luciferase reporter gene, and candidate CREs were cloned into an enhancer assay vector (pGL4.23[minP]; Promega), which contains a minimal promoter (minP) followed by the luciferase reporter gene (**Supplementary Table 3.3**). Bat kidney regions were tested for luciferase activity in human kidney cells (HEK293T, and bat pancreas regions were tested in human liver cells (HepG2), as there are no well-established beta cell lines, and HepG2 cells are glucose-responsive, a valid model for insulin signaling and production, and easily transfectable (*244*, *245*). Cells were grown in standard growth mediums (see **Methods**) and co-transfected with the *Renilla* luciferase reporter gene vector (pGL4.73[hRluc/SV40]; Promega) to normalize for transfection efficiency, and transfected cell extracts were surveyed for relative luciferase activity 48 hours after transfection.

In HEK293T, only the insectivore *KLK1* CRE drove significantly higher luciferase activity than the minP empty vector and the frugivore *KLK1* CRE (**Supplementary Figure 3.9A-B**). In HepG2, both the frugivore and insectivore *INS* PRO drove significantly higher luciferase activity than the empty vector, but luciferase activity was not significantly different between these promoters (**Supplementary Figure 3.9C**).

3.5: Concluding Remarks

Using integrative single-cell sequencing, we characterized the cell populations, transcriptomes, and regulomes of insectivorous and frugivorous bat kidneys and pancreases in a single-cell manner. We identified major cell-types in the kidneys and pancreases of these bats, dissected the transcriptional and regulatory differences between them, and validated several cell composition findings with immunofluorescence. For the frugivore kidney, we found a reduction in loop of Henle cells, combined with a loss of urine-concentrating transporters. We observed an expansion of collecting duct cells, combined with upregulation of sodium reabsorption-potassium secretion genes, and genomic enrichment for disease phenotypes (**Figure 3.6**). For the frugivore pancreas, we found an expansion of beta and alpha cells, accompanied by a reduction in acinar cells, and several genes and genomic regions associated with insulin secretion and signaling (**Figure 3.6**). Combined, our work provides a cellular and molecular blueprint of frugivorous adaptation in mammals and can inform potential therapeutic targets for human disease, particularly hypertension, hyperkalemia and diabetes.

To our knowledge, our results provide the first unbiased and comprehensive analysis of the cell-type compositions that distinguish insectivorous and frugivorous mammals. For the kidney, we find that the medullary and cortical differences observed between frugivorous and insectivorous kidneys(*22, 35, 145*) are due to specific nephron composition differences in DTL, TAL, connecting tubules, principal cells, type A intercalated cells, type B intercalated cells, and proximal tubules-like cells (**Figure 3.6**). As loop of Henle cells, DTL and TAL, are responsible for urine concentration and water recovery, and fruit bats get a substantial amount of water from fruit, fruit bats likely do not need as much of a structure for preserving water while excreting waste. A reduced renal medulla in response to water availability in diet or climate has also been

observed in birds(*246–248*). The greater abundance of connecting tubules and collecting duct cells in fruit bats highlights nephron restructuring in response to high potassium and bicarbonate and low sodium and acid, favoring a larger ASDN. We also found fewer proximal tubules-like cells in fruit bats, which is notably observed in human diabetic kidneys(*149*). For the pancreas, we found that the large amount of endocrine tissue observed in fruit bats(*75, 103*) is due to greater beta and alpha cell abundances, relative to insectivorous bats, and a compensatory reduction in exocrine tissue, specifically in acinar cells (**Figure 3.6**). These cell composition differences likely contribute to the unique ability of fruit bats to lower their blood sugar rapidly, even faster than insectivorous bats(*3, 17, 19, 20*). Taken together, our integrative single-cell analysis identified several novel cell composition differences and provides a cellular-level detailed catalog of previously observed morphological frugivorous kidney and pancreas features.

Prior to our study, identifying potential molecular adaptations to frugivory was mainly restricted to molecular evolutionary analyses of specific genes and comparative genomics(*3, 7, 15, 21, 46–52*). Here, we were able to use scRNA-seq to systematically identify gene expression differences in an unbiased manner. Our gene expression analyses identified numerous novel molecular adaptations that could be associated and vital for fruit specialization in mammals. In the kidney, the dilute urine observed in fruit bats(*106*) is likely attributed to decreased expression of key urine-concentrating transporters, *STK39*, *OXSR1*, *KCNJ1*, *SLC9A3*, within the TAL (**Figure 3.6**). We also found that the fruit bat kidney exhibits gene expression changes to stimulate sodium reabsorption and potassium excretion, such as *KCNMA1*, *WNK1*, and *WNK4*, resembling an activated RAAS, which supports low sodium, high potassium dietary specialization (**Figure 3.6**). Some genes that were lost in OWFB genomes, like kidney transporters *SLC22A6* and *SLC22A12*, have been hypothesized to be adaptive for frugivory(*15*),

and our data allowed us to determine whether these hypotheses are lineage-specific (loss of

SLC22A6 and *SLC22A12* is specific to OWFBs). In the pancreas, the increased sensitivity to

insulin and glucose observed in fruit bats(*75, 103*) involves many genes that are associated with

insulin secretion and signaling, like *PTPRD* and *ABCC8* (**Figure 3.6**). Moreover, we discovered

a unique connection in myo-inositol transport and metabolism from our data and previous gene

expression analyses in the same fruit bat species(*54*) (**Figure 3.6**).

Our scATAC-seq datasets allowed us to identify gene regulatory elements and TFBSs that could

be involved in adaptation to frugivory. For the kidney, we found regulatory elements for many

differentially expressed and diabetes-associated genes, including *KLK1*, which demonstrated

cell-type-specific expression in bats with both promoter and enhancer accessibility, and *PCK1*,

which was also shown to have differential chromatin accessibility downstream of the gene body

in human diabetic proximal tubules(*173*). We also found insectivore renal epithelial cells to be

enriched for *NFAT* motifs like *NFAT5*, which regulates proteinuria and osmotic pressure(*176,

177*), fitting the high protein consumption and relatively low water intake of insectivores,

compared with that of frugivores. Frugivores were enriched for diabetes-associated motifs, such

as *ONECUT*, and were highly enriched for human kidney phenotypes and for RAAS in

particular. For the pancreas, we found *INS* and *GCG* promoters to be highly accessible in

frugivore beta and alpha cells, respectively. As frugivorous bats are known to have remarkable

blood sugar regulation(*3, 17, 19, 20*), promoter accessibility of *INS* and *GCG* may play a role.

Insectivorous bats were also found to be enriched for motifs of many canonical pancreatic TFs

like *HNF1B* and for many fat-related mouse phenotypes, corresponding to their higher fat diets,

while frugivores were enriched for motifs of TFs related to diabetes like *KLF*s. As metabolites

like lipids and carbohydrates regulate transcription through macronutrient-sensing TFs(*249*), the

regulatory datasets we generated in this study offer the opportunity to investigate regulatory element evolution in response to dietary specialization.

The limitations of our study include our sample size, genome assembly and annotation, and the lack of functional genomic datasets for bats, precluding our ability to have known markers in bat kidneys and pancreases. Nonetheless, known markers in kidneys and pancreases of humans and mice were sufficient to identify major cell-types in the bat kidneys and pancreases. However, using human and mouse markers prevented us from identifying novel cell-types in bats. The smaller frugivore sample size for pancreas prevented us from detecting significant differences in cell composition, but we were still able to identify differences between bats with high confidence. Future studies may utilize spatial transcriptomics to compare cellular architecture and structures. We were also limited in analysis tools for non-model organism genomes. High-quality genomes are needed for deeper analyses, as they can provide better gene annotation, transcriptional isoforms and improve connection of gene regulatory elements to their target genes. Furthermore, novel genome annotations are needed to increase the number of shared features that can be recognized between species and to examine more genes in relation to humans and human disease. This was apparent in our results with the insectivorous bat genome having more annotated genes shared with humans than the fruit bat genome. Additional bat genomes, functional genomic databases, and more single-cell datasets from other insectivorous and frugivorous mammals will further determine whether our findings are specific to the species we investigated here, to NWFBs, and/or to other frugivorous mammals. Future studies with similar tissues could further investigate the differences between fasted and fed states within each bat species at various time points and understand developmental mechanisms for cell composition differences between the kidneys and pancreases of frugivorous and insectivorous bats. In addition, our study only focused on two

organ types (kidney and pancreas) and thus could not uncover tissue adaptations in other organs that might be associated with frugivory. Lastly, we were unable to identify cell type-specific *cis*-regulatory regions that drive differential transcriptional activity between bats using luciferase reporter assays in human cells, with the exception of *KLK1* CRE, which the frugivore CRE was predicted to drive higher transcriptional activity than the insectivore CRE. These findings are consistent with findings in **Chapter 2**, and thus, reiterate the importance of cellular context (human cells vs. bat cells) and environmental conditions (ie. glucose stimulation).

Bats have been viewed as a model for diabetes research due to endocrine tissue differences between bat species and their unique blood sugar regulation. To the best of our knowledge, our study provides the first detailed analysis of the mammalian frugivorous kidney. The frugivore kidney exhibits many diabetic features, including decreased proximal tubules, upregulation of gluconeogenesis, glucose reabsorption, and potassium secretion genes, and corresponding diabetes-associated motifs, including *KLF9* which was also enriched in T2D renal tubules(*173*) (**Figure 3.6**). Many of these kidney traits correspond with what we found in the frugivorous pancreas, such as increased transport and need for myo-inositol and motif and expression enrichment of diabetes-associated genes (**Figure 3.6**). We also provide the first detailed analyses of bat pancreases, in which we documented differential expression of many signature genes in human T2D beta cells and differential enrichment of diabetes-associated motifs in endocrine and exocrine cells (**Figure 3.6**). Together, our integrative single-cell study indicates that fruit bats evolved many diabetic-like features to deal with their diets but evolved protective mechanisms that prevent disease, such as upregulation of *KLK1* in type A intercalated cells in the kidney, which protects against diabetic tissue damage(*154*), and downregulation of *IRS1* in beta cells in the pancreas, which causes hyperinsulinism but not diabetes(*218*). Importantly,

these data can be utilized to understand regulation of and adaptation to conditions that can cause disease in humans. We demonstrated how GWAS variants can be used with our single-cell data to reveal cell-type-specific regulation of diabetes-associated loci, such as *PROX1*, whereby rs79687284 was not previously examined in distal convoluted tubules or in the kidney. Our study also provides a unique perspective for human disease therapeutic development by revealing cellular and molecular adaptations to high sugar, potassium, and bicarbonate levels and to low amounts of protein, sodium, and calcium. *Cis*-regulation therapy (CRT)(*100*), for example, can take advantage of the genes and regulatory elements identified here for metabolic disease treatment. Moreover, our cell-type-specific data provides insight for designing localized treatments in heterogeneous tissues.

3.6: Figures & Tables
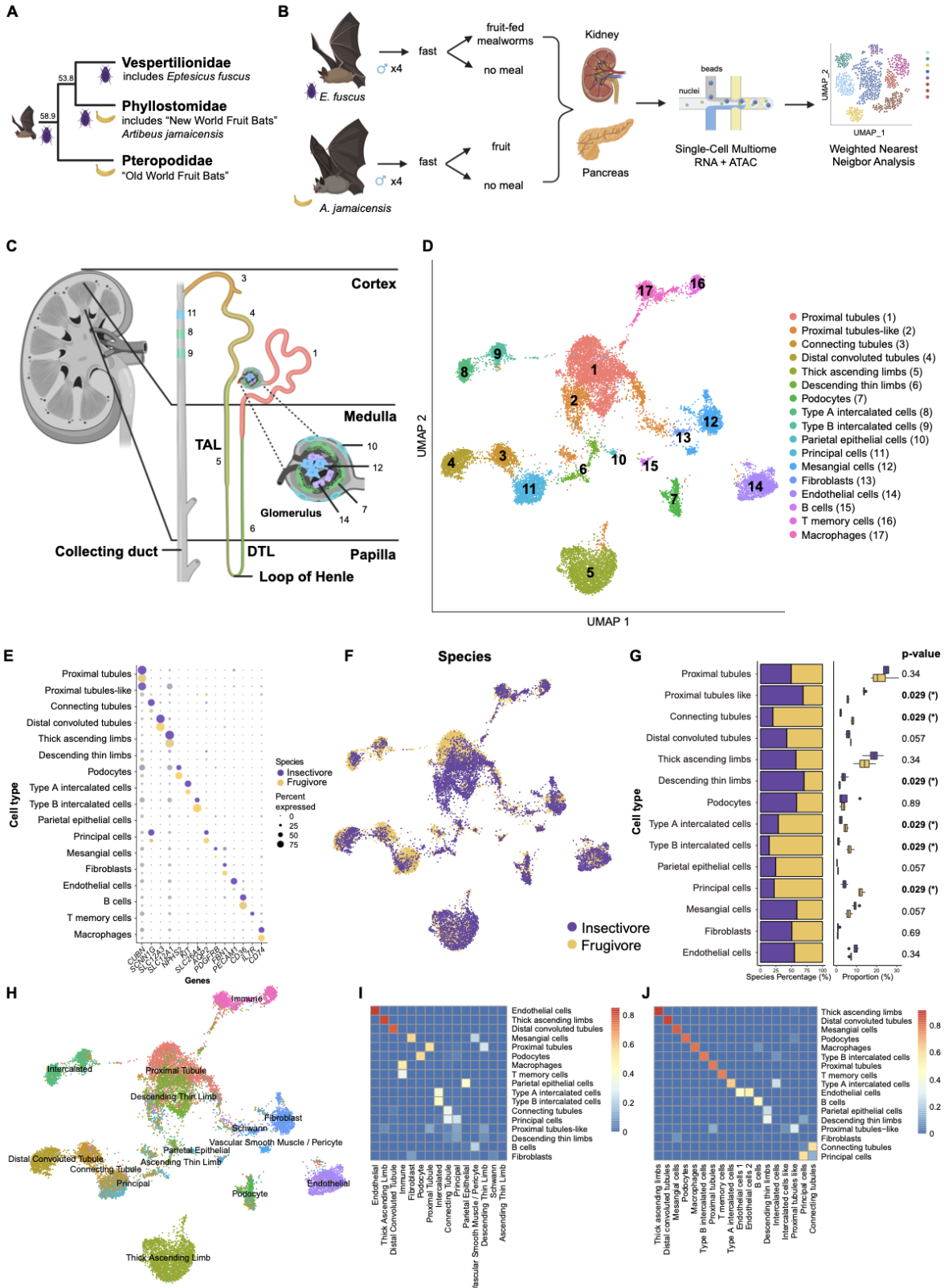
Figure 3.1: Joint scRNA and scATAC profiling of bat kidney.
(**A**) Representative phylogenetic tree showing the evolution of frugivory in bats (order: Chiroptera). The beetle denotes lineage with insectivorous diet, and banana denotes lineage with frugivorous diet. Numbers are in millions of years(*135*). (**B**) Experimental design for joint scRNA and scATAC profiling of bat tissues (see Materials and Methods). (**C**) Diagram of kidney with zoom-in on the nephron. Colors correspond to cell-type colors in **D**. (**D**) Uniform Manifold Approximation and Projection (UMAP) of bat kidney cell-types based on scRNA-seq profiles. (**E**) Dot plot of marker gene expression across all bat kidney cell-types. Color intensity indicates the average expression level across all cells within a cell-type (purple or yellow is high; gray is low). (**F**) UMAP of bat kidney cell-types by species/dietary phenotype. (**G**) Plot of species percentage across all renal cells (left) with corresponding Wilcoxon rank-sum test for differential cell-type abundance (right) with significant p-values shown in bold font; * < 0.05. (**H**) UMAP of bat kidney cell-types annotated with human adult kidney single-cell reference data from Azimuth(*110*). (**I**) Jaccard overlap of annotations in **H** (horizontal) with our integrated species annotations in **D** (vertical). (**J**) Jaccard overlap of all annotations found in each species before integration (horizontal) with after integration in **D** (vertical).
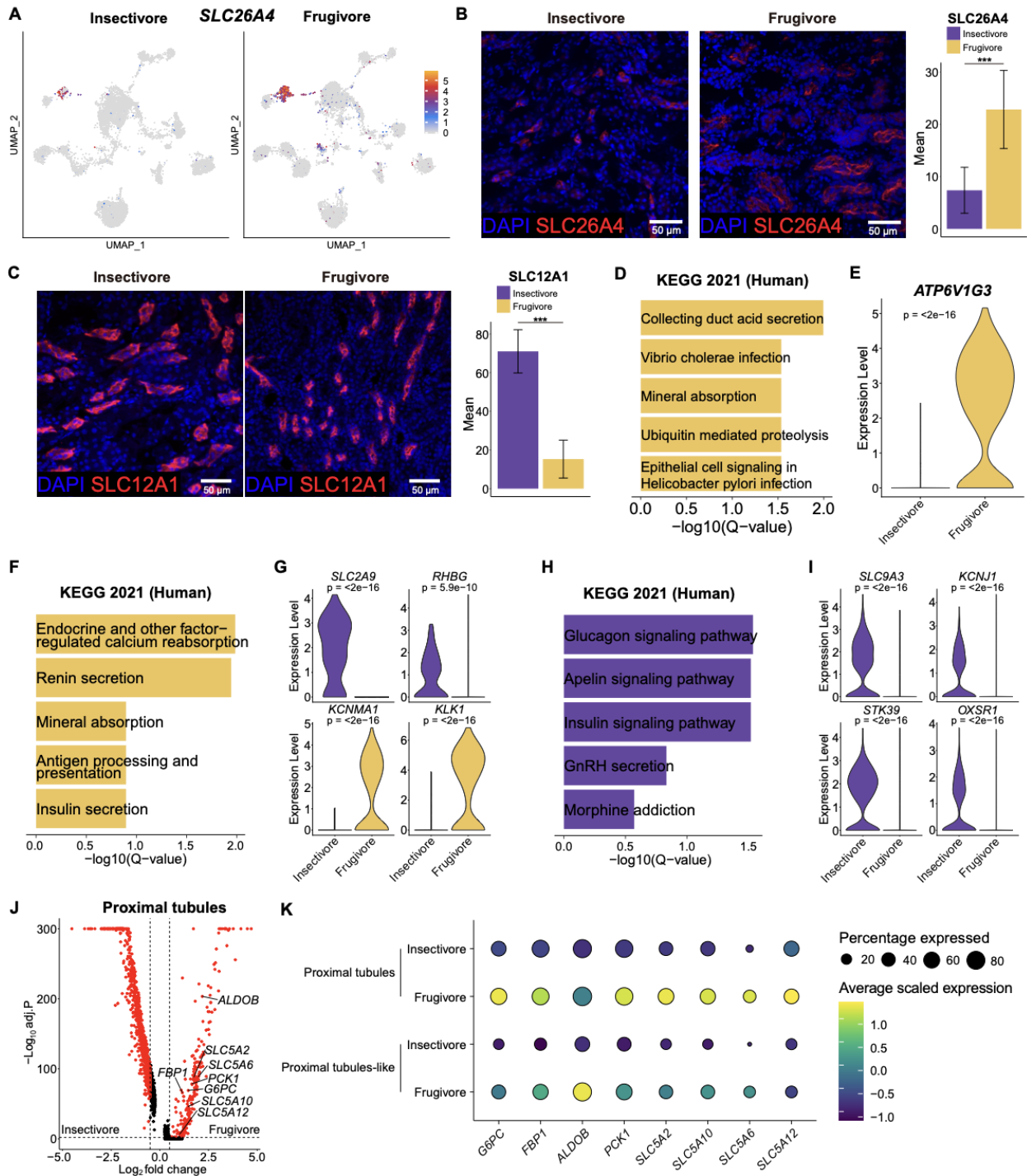
Figure 3.2: Immunofluorescence and differential gene expression analyses identify traits facilitating frugivory in the bat kidney.
(**A**) UMAPs of type B intercalated cell marker gene *SLC26A4* expression in each species. (**B**) (Left) Representative images of *SLC26A4* immunofluorescence (red) in bat kidneys. Nuclei are stained with DAPI (blue). (Right) Quantification of *SLC26A4* immunofluorescence normalized to nuclei in bat kidneys. Results represent arbitrary units of fluorescence (AU) mean ± standard error of the mean (SEM) derived from 3 insectivorous and 3 frugivorous bats (n = 3/phenotype,

n = 10 images/individual [see Materials and Methods]). Mixed effects model ***p-value < .0001. (**C**) (Left) Representative images of *SLC12A1* immunofluorescence (red) in bat kidneys. Nuclei are stained with DAPI (blue). (Right) Quantification of *SLC12A1* immunofluorescence normalized to nuclei in bat kidneys. Results represent mean (AU) ± standard error of the mean (SEM) derived from 3 insectivorous and 3 frugivorous bats (n = 3/phenotype, n = 10 images/individual [see Materials and Methods]). Mixed effects model ***p-value < .0001. (**D**) Bar plots showing Kyoto Encyclopedia of Genes and Genomes (KEGG) Human 2021 pathways enriched in fruit bat type B intercalated cells. (**E**) Violin plot of *ATP6V1G3* expression in type B intercalated cells. (**F**) Bar plots showing KEGG Human 2021 pathways enriched in fruit bat type A intercalated cells. (**G**) Violin plots of *SLC2A9*, *RHBG*, *KCNMA1*, and *KLK1* in type A intercalated cells. (**H**) Bar plots showing KEGG Human 2021 pathways enriched in insectivore thick ascending limbs. (**I**) Violin plots of *SLC9A3*, *OXSR1*, *STK39*, and *KCNJ1* expression in TAL. (**J**) Volcano plot showing differentially expressed genes between species in proximal tubules cells. (**K**) Dot plots showing the expression of gluconeogenesis and various SLC5 genes in bat proximal tubules and proximal tubules-like cells.
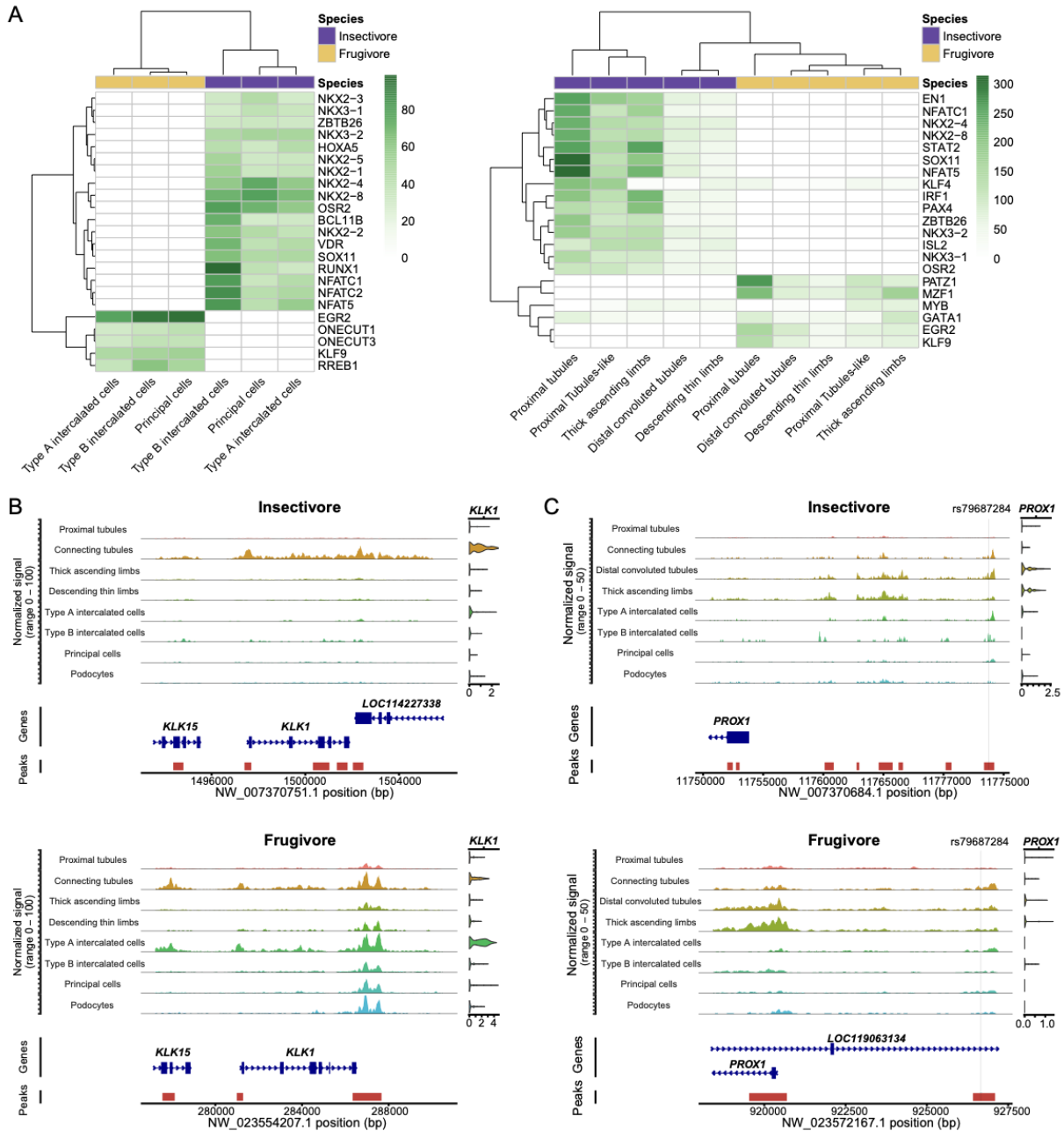
Figure 3.3: scATAC analyses reveal TFBSs and chromatin accessibility in bat kidneys.
(**A**) (Left) Heatmap of TF motifs enriched in bat kidney collecting duct cells. (Right) Heatmap of TF motifs enriched in bat renal tubule and limb cells. Both heatmaps display -log(p-value). (**B**) scATAC-seq coverage plots of *KLK1* in bats. SCTransform-normalized expression plot visualized on the right by cell-type. (**C**) scATAC-seq coverage plots of rs79687284 in the *PROX1* locus in bats. SCTransform-normalized expression plot visualized on the right by cell-type.
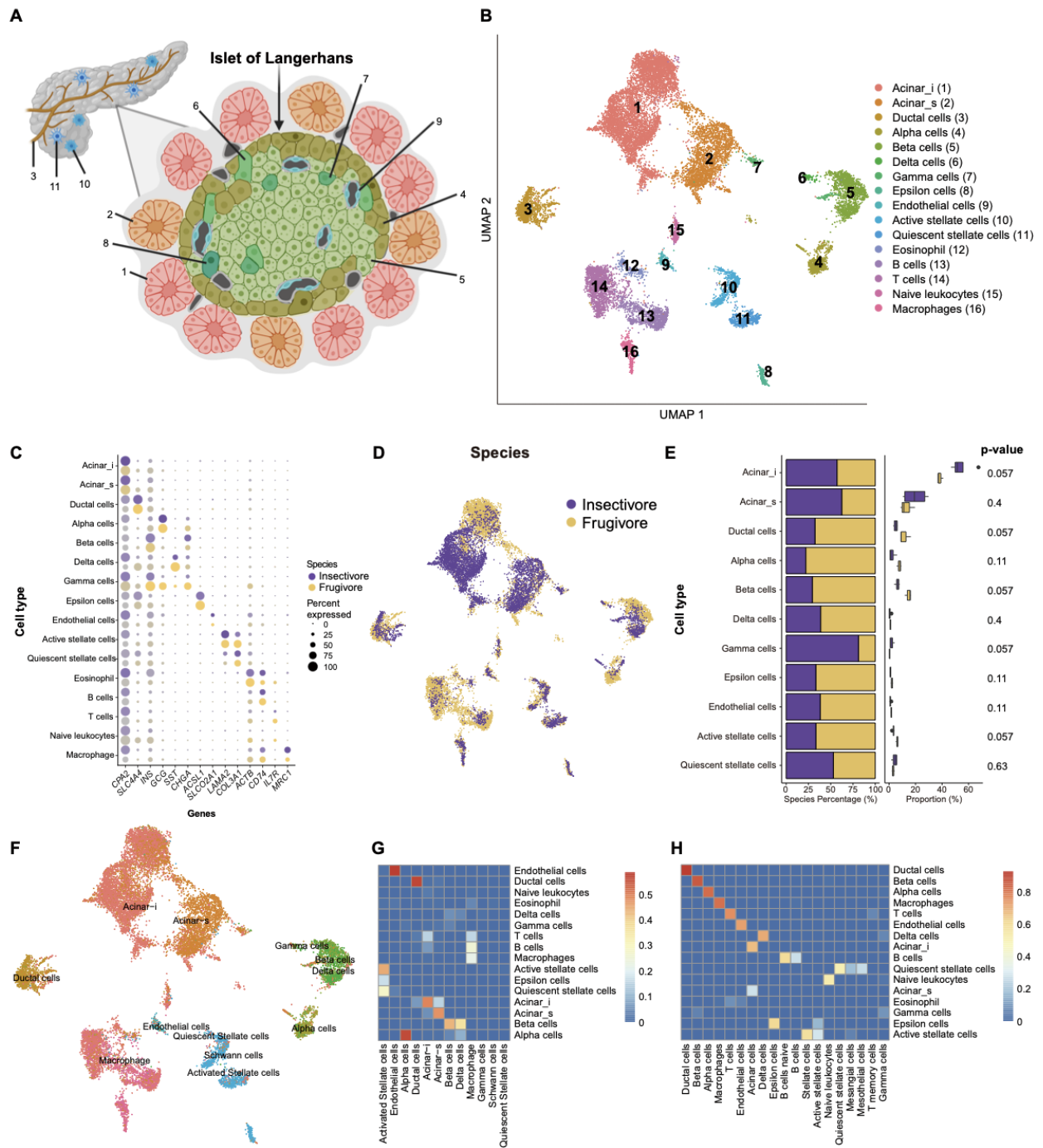
Figure 3.4: Joint scRNA and scATAC profiling of the bat pancreas.
(**A**) Diagram of pancreas with zoom-in on the Islet of Langerhans. Colors correspond to cell-type colors in **B**. (**B**) UMAP of bat pancreas cell-types based on scRNA-seq data. (**C**) Dot plot of marker gene expression across all bat pancreas cell-types. Color intensity indicates the average expression level across all cells within a cell-type (purple or yellow is high, gray is low). (**D**) UMAP of bat pancreas cell-types by species/dietary phenotype. (**E**) Plot of species percentage

across all pancreatic cells (left) with corresponding Wilcoxon rank-sum test for differential cell-type abundance (right). (**F**) UMAP of bat pancreas cell-types automatically annotated with human pancreas single-cell reference data(*119*). (**G**) Jaccard overlaps of auto-annotations in **G** (horizontal) with our integrated species annotations in **B** (vertical). (**H**) Jaccard overlap of all annotations found in each species before integration (horizontal) with after integration in **B** (vertical).
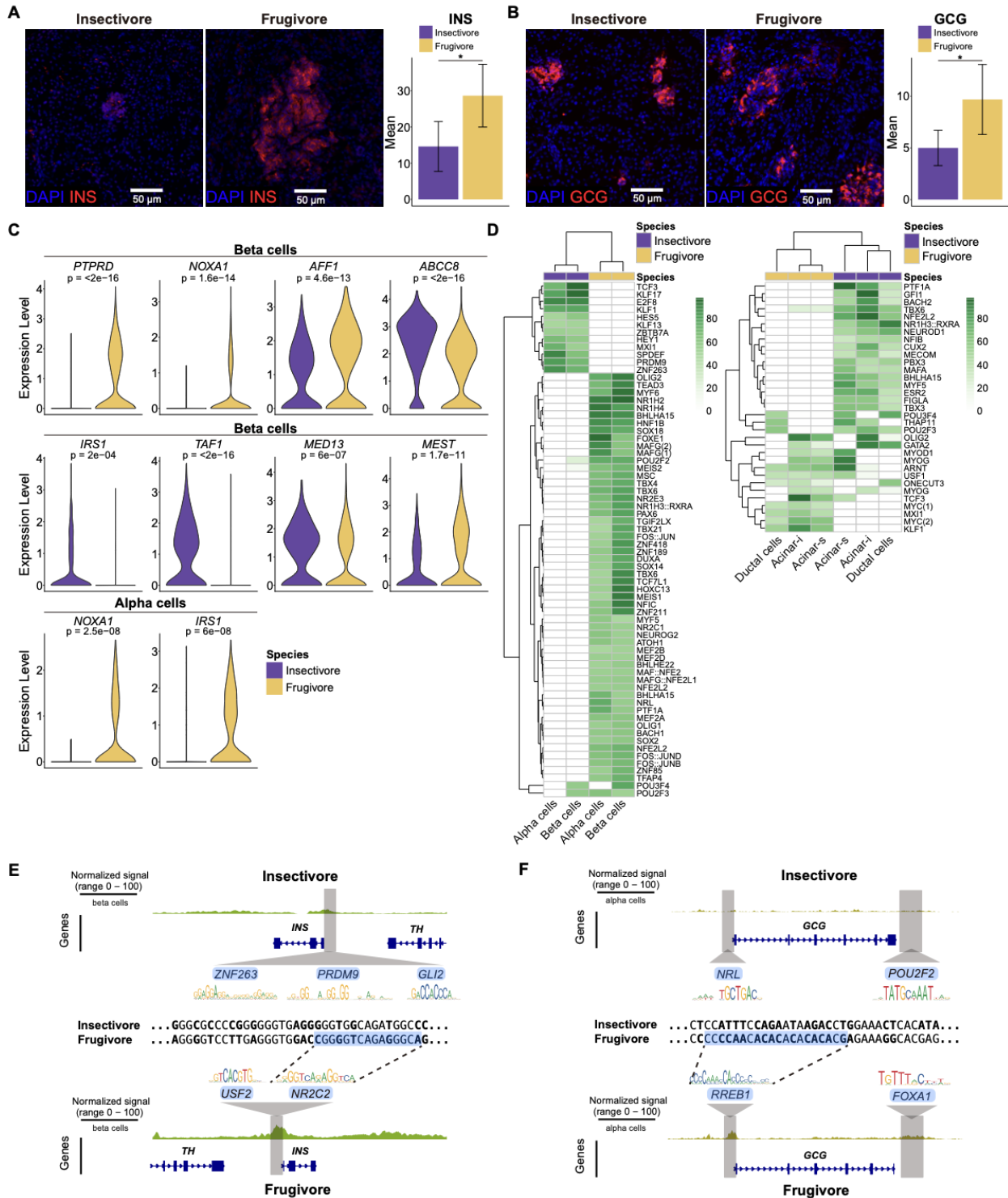
Figure 3.5: scRNA and scATAC analyses depict both exocrine and endocrine dietary differences between insectivorous and frugivorous bat pancreases.

(**A**) (Left) Representative images of *INS* immunofluorescence (red) in bat pancreases. Nuclei are stained with DAPI (blue). (Right) Quantification of *INS* immunofluorescence normalized to

nuclei in bat pancreases. Results represent arbitrary units of fluorescence (AU) mean ± standard error of the mean (SEM) derived from 3 insectivorous and 3 frugivorous bats (n = 3/phenotype, n = 10 images/individual [see Materials and Methods]). Mixed effects model *p*-value = .002. (**B**) (Left) Representative images of *GCG* immunofluorescence (red) in bat pancreases. Nuclei are stained with DAPI (blue). (Right) Quantification of *GCG* immunofluorescence normalized to nuclei in bat pancreases. Results represent arbitrary units of fluorescence (AU) mean ± standard error of the mean (SEM) derived from 3 insectivorous and 3 frugivorous bats (n = 3/phenotype, n = 10 images/individual [see Materials and Methods]). Mixed effects model *p*-value = .008. (**C**) Violin plots of diabetes-associated genes in bat endocrine cells. (**D**) (Left) Heatmap of TF motifs enriched in bat pancreatic endocrine cells. (Right) Heatmap of TF motifs enriched in bat pancreatic exocrine cells. Both heatmaps display -log(p-value). (**E**) scATAC-seq coverage plots of *INS* in bat pancreases with predicted TFBS highlighted in gray. Aligned sequence zoom-in on *NR2C2* motif with differing nucleotides depicted in bold. (**F**) scATAC-seq coverage plots of *GCG* in bat pancreases with predicted TFBS highlighted in gray. Aligned sequence zoom-in on *RREB1* motif with differing nucleotides depicted in bold.
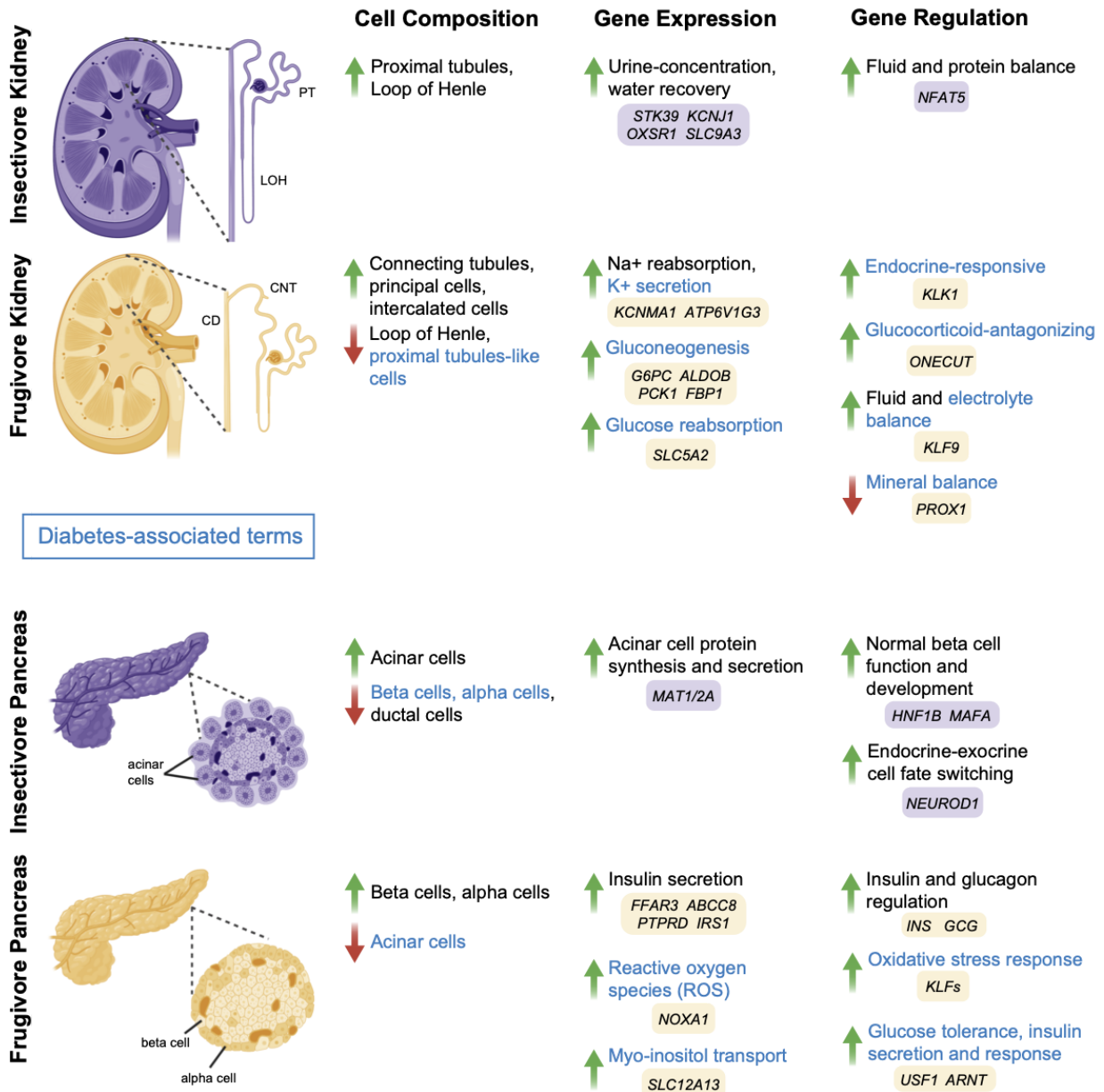
Figure 3.6: Summary of cell composition, gene expression and gene regulation differences between the bat insectivorous and frugivorous kidney and pancreas and how it relates to human diabetes. PT = proximal tubules, LOH = loop of Henle, CD = collecting duct, CNT = connecting tubules.
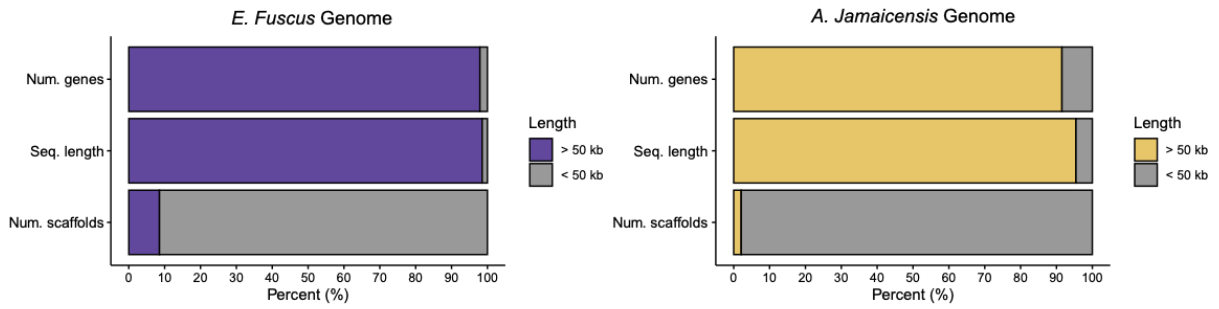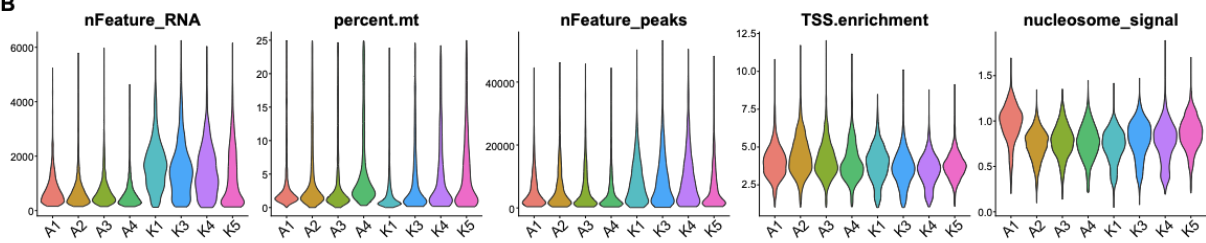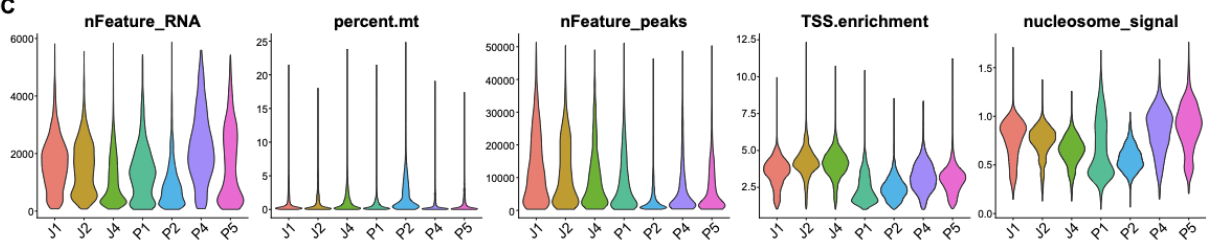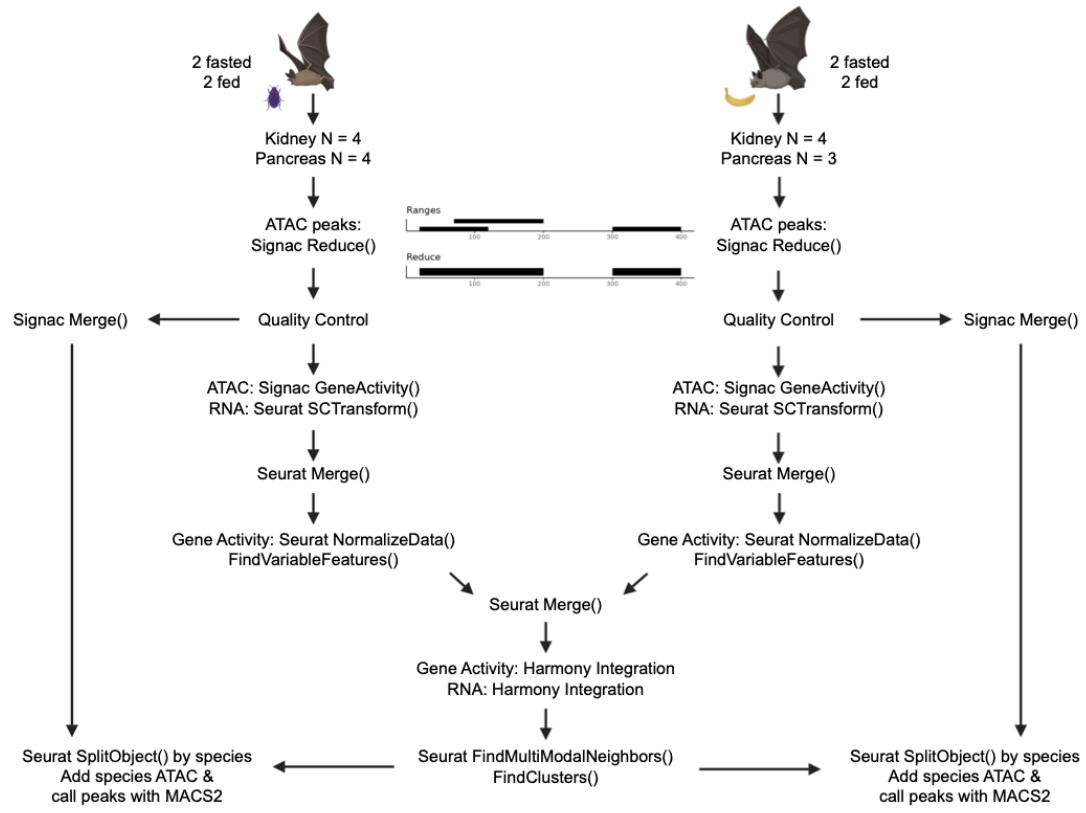
Supplementary Figure 3.1. Joint scRNA and scATAC design and processing in bat tissues.

(**A**) Bar chart showing that the number of genes and the genome sequence length captured by the number of scaffolds > 50 kb is > 90% for both bat genomes used in this study. (**B**) Violin plots of QC metrics Num. RNA features (nFeature_RNA), mitochondrial percentage (percent.mt), Num. ATAC features (nFeature_peaks), transcription start site enrichment (TSS.enrichment), and nucleosome signal for single-cell multiome on bat kidneys. A1-2 = fasted fruit bats, A3-4 = treated fruit bats, K1-3 = treated insect bats, K4-5 = fasted insect bats. (**C**) Violin plots of QC metrics nFeature_RNA, percent.mt, nFeature_peaks, TSS.enrichment, and nucleosome signal for single-cell multiome on bat pancreases. J1-2 = fasted fruit bats, J4= treated fruit bat, P1-2 = treated insect bats, P4-5 = fasted insect bats. (**D**) Schematic for cross species integration of multi-omic data(*111*).

A

Proximal tubules · Proximal tubules-like · Connecting tubules · Distal convoluted tubules · Thick ascending limbs · Descending thin limbs · Podocytes · Type A intercalated cells · Type B intercalated cells · Parietal epithelial cells · Principal cells · Mesangial cells · Fibroblasts · Endothelial cells · B cells · T memory cells · Macrophages

sign: neg, pos

1 : Frugivore Fast vs Fed
2 : Insectivore Fast vs Fed
3 : Frugivore A1+A3 vs A2+A4
4 : Frugivore A1+A4 vs A2+A3
5 : Insectivore K1+K4 vs K3+K5
6 : Insectivore K1+K5 vs K3+K4
7 : Frugivore vs Insectivore

B

Acinar_i · Acinar_s · Ductal cells · Alpha cells · Beta cells · Delta cells · Gamma cells · Epsilon cells · Endothelial cells · Active stellate cells · Quiescent stellate cells · Eosinophil · B cells · T cells · Naive leukocytes · Macrophages

sign: neg, pos

1 : Frugivore Fast vs Fed
2 : Insectivore Fast vs Fed
3 : Frugivore J1 vs J2+J4
4 : Insectivore P1+P4 vs P2+P5
5 : Insectivore P1+P5 vs P2+P4
6 : Frugivore vs Insectivore

C

Kidney condition association

global p=0.166 · Filtered for FDR<0.10

Correlation: 0.50, 0.25, 0.00, −0.25, −0.50

Pancreas condition association

global p=0.673 · Filtered for FDR<0.10

Correlation: 0.2, 0.1, 0.0, −0.1, −0.2

97

Supplementary Figure 3.2. Differential expressed gene (DEG) counts by condition and by species in bat kidney and pancreas.

(**A**) Bar charts of DEG counts in bat kidneys (legend notation: pos. vs neg.). A1-2 = fasted fruit bats, A3-4 = treated fruit bats, K1-3 = treated insect bats, K4-5 = fasted insect bats. (**B**) Bar charts of DEG counts in bat pancreases (legend notation: pos. vs neg.). J1-2 = fasted fruit bats, J4= treated fruit bat, P1-2 = treated insect bats, P4-5 = fasted insect bats. (**C**) (Left) CNA of condition associations to cell-types in bat kidney. (Right) CNA of condition associations to cell-types in bat pancreas.

A

Macrophage

Intercalated cells

Proximal convoluted tubules
Proximal straight tubules

Stroma 2
Stroma 1

Distal convoluted tubules

Principal cells

Podocytes

Endothelial cells

Loop of Henle

B

Endothelial cells
Thick ascending limbs
Podocytes
Macrophages
Mesangial cells
Proximal tubules
Type A intercalated cells
Type B intercalated cells
T memory cells
Proximal tubules-like
Descending thin limbs
Parietal epithelial cells
B cells
Fibroblasts
Principal cells
Connecting tubules
Distal convoluted tubules

Endothelial cells
Loop of Henle
Podocytes
Macrophage
Stroma 1
Proximal straight tubules
Intercalated cells
Stroma 2
Distal convoluted tubules
Principal cells
Proximal convoluted tubules

C

MIOX p = <2e−16

SLC34A1 p = <2e−16

LRP2 p = <2e−16

Expression Level

Proximal tubules   Proximal tubules-like

D

Insectivore

9 (1.5%) 11 (4.6%)
8 (2.4%)
7 (5.9%)
6 (4.5%)
1 (30.5%)
5 (24.1%)
4 (7.1%)
3 (2.6%)
2 (16.8%)

Frugivore

11 (14%)
9 (7.9%)
8 (5.1%)
7 (3.7%)
6 (1.7%)
5 (15.8%)
4 (8.4%)
3 (9.2%)
2 (6.8%)
1 (27.3%)

Celltype
Proximal tubules (1)
Proximal tubules-like (2)
Connecting tubules (3)
Distal convoluted tubules (4)
Thick ascending limbs (5)
Descending thin limbs (6)
Podocytes (7)
Type A intercalated cells (8)
Type B intercalated cells (9)
Principal cells (11)

E Species association

global p=0.033

Filtered for FDR<0.10

Correlation
0.5
0.0
−0.5

UMAP_2

UMAP_1

F

Insectivore   Frugivore

Proximal tubules
Proximal tubules-like
Connecting tubules
Distal convoluted tubules
Thick ascending limbs
Descending thin limbs
Podocytes
Type A intercalated cells
Type B intercalated cells
Parietal epithelial cells
Principal cells
Mesangial cells
Fibroblasts
Endothelial cells
B cells
T memory cells
Macrophages

Pearson residuals:
13
4
2
0
−2
−4
−13

p-value = < 2.22e−16

G Condition

Proximal tubules
Proximal tubules-like
Connecting tubules
Distal convoluted tubules
Thick ascending limbs
Descending thin limbs
Podocytes
Type A intercalated cells
Type B intercalated cells
Parietal epithelial cells
Principal cells
Mesangial cells
Fibroblasts
Endothelial cells
B cells
T memory cells
Macrophages

Fasted
Fed

Percentage (%)

Supplementary Figure 3.3. Single-cell composition analysis of bat kidney.

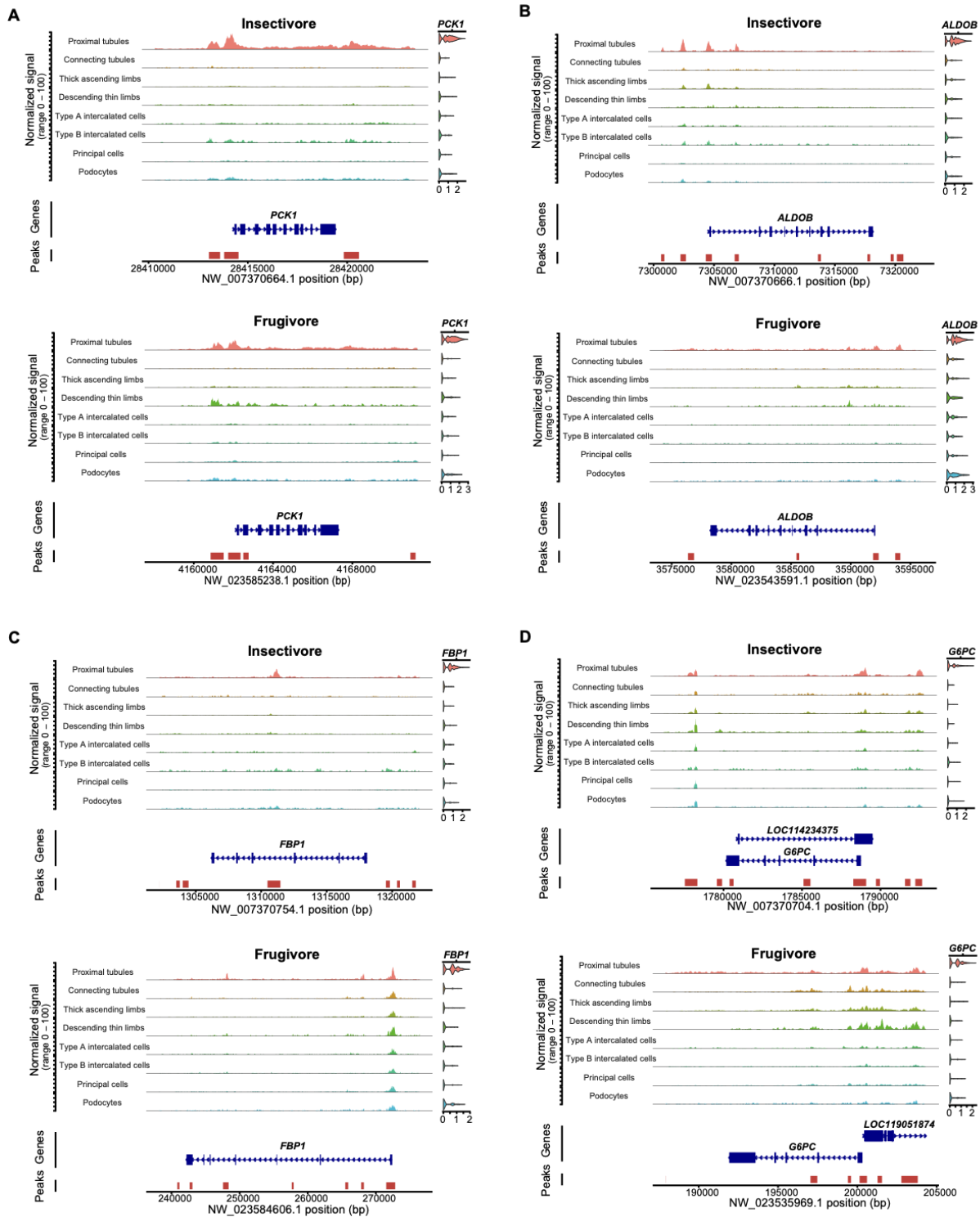(**A**) UMAP of bat kidney cell-types automatically annotated with mouse kidney single-cell reference data from Azimuth(*110*). (**B**) Jaccard overlaps of auto-annotations in **A** (horizontal) with our integrated species annotations in **Fig. 1D** (vertical). (**C**) Violin plots of proximal tubule marker gene expression in bat proximal tubules and proximal tubules-like cells. (**D**) Pie charts of cell-type percentages across renal epithelial cells. (**E**) CNA of species associations to cell-types. (**F**) Pearson residuals visualized by mosaic plot of all cell-types identified in bat kidneys. (**G**) UMAP of bat kidney cell-types by condition. The bar chart shows the proportion of condition for each cell-type.
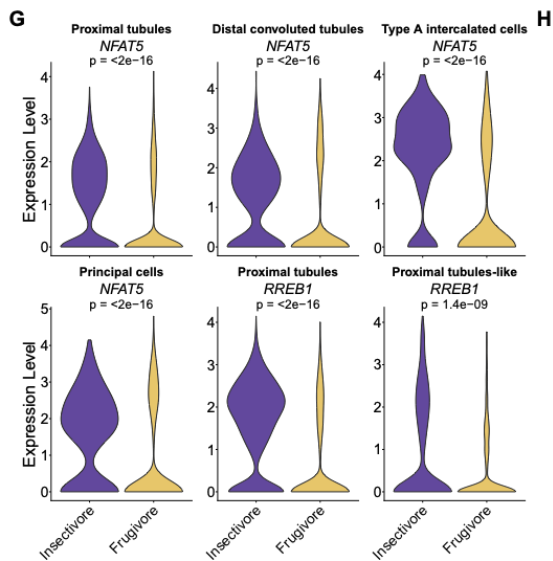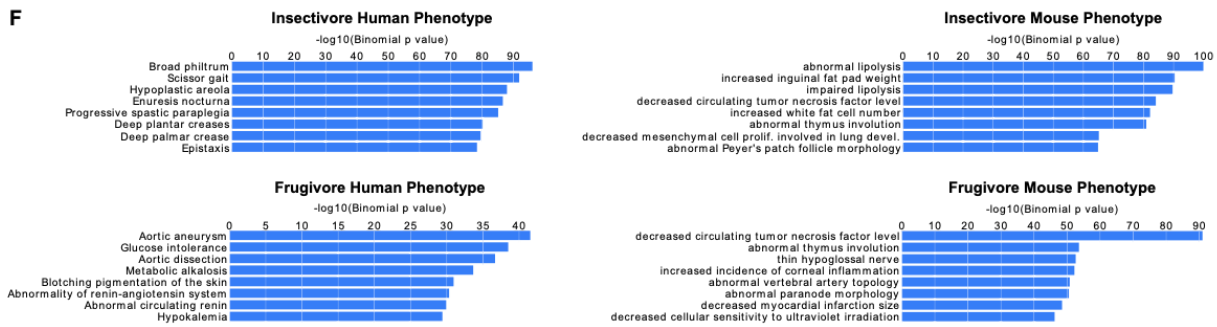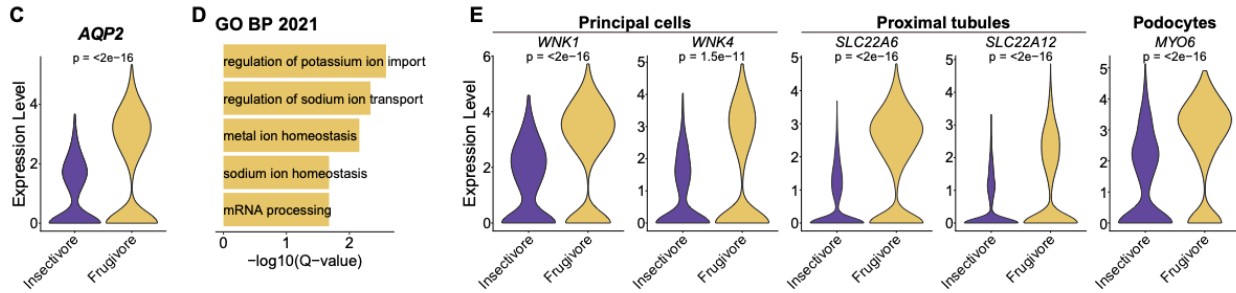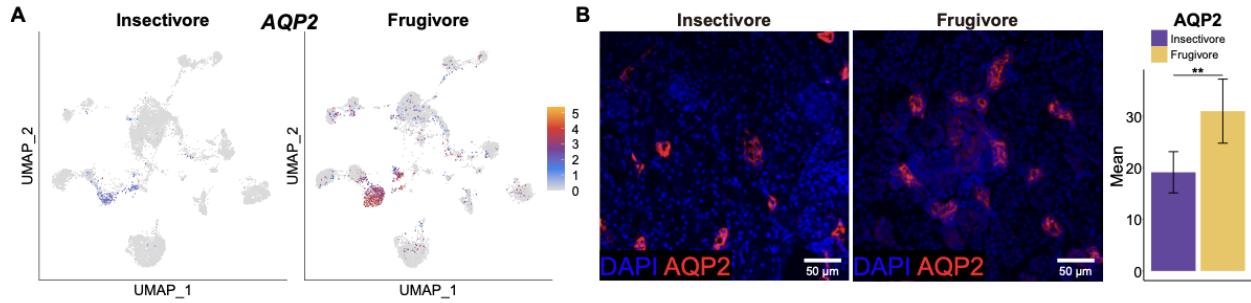
**Supplementary Figure 3.4. scRNA-seq and scATAC-seq analysis of bat kidney.**

(**A**) UMAPs of principal cell marker gene *AQP2* expression in each species. (**B**) (Left) Representative images of *SLC26A4* immunofluorescence (red) in bat kidneys. Nuclei are stained with DAPI (blue). (Right) Quantification of *AQP2* immunofluorescence normalized to nuclei in bat kidneys. Results represent arbitrary units of fluorescence (AU) mean ± standard error of the mean (SEM) derived from 3 insectivorous and 3 frugivorous bats (n = 3/phenotype, n = 10 images/individual [see Materials and Methods]). Mixed effects model **p-value = .001. (**C**) Violin plot of *AQP2* expression in principal cells. (**D**) Bar plots showing GO Biological Process 2021 terms enriched in frugivore principal cells. (**E**) Violin plots of differentially expressed genes in bat kidney cells. (**F**) Bar plots showing GREAT human and mouse phenotypes enriched in insectivore and frugivore kidneys. (**G**) Violin plots of differentially expressed TFs in bat renal epithelial cells. (**H**) scATAC-seq coverage plots of *ATP6V1G3* in bat kidneys. SCTransform-normalized expression plot visualized on the right by cell-type.
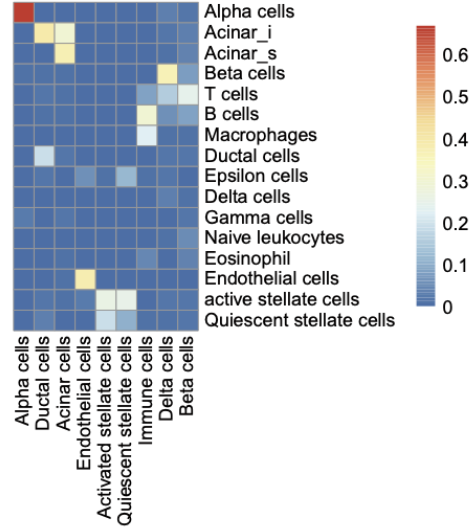
Supplementary Figure 3.5. scATAC-seq coverage plots of diabetes-associated genes in bat kidneys.

(**A-D**) scATAC-seq coverage plots of *PCK1* (**A**), *ALDOB* (**B**), *FBP1* (**C**) and *G6PC* (**D**) in bat kidneys. SCTransform-normalized expression plot visualized on the right by cell-type.
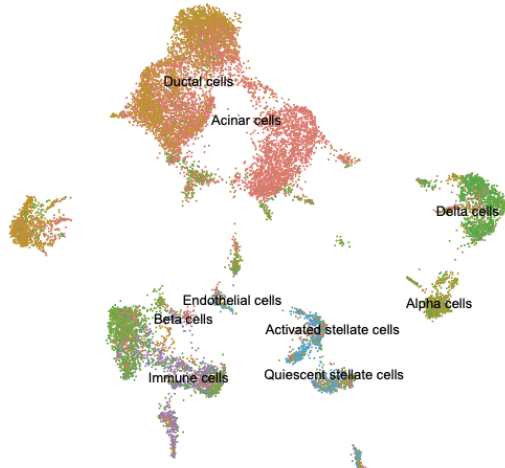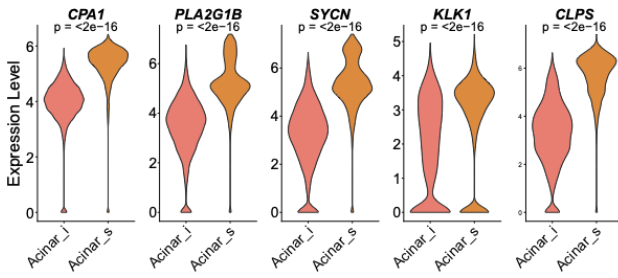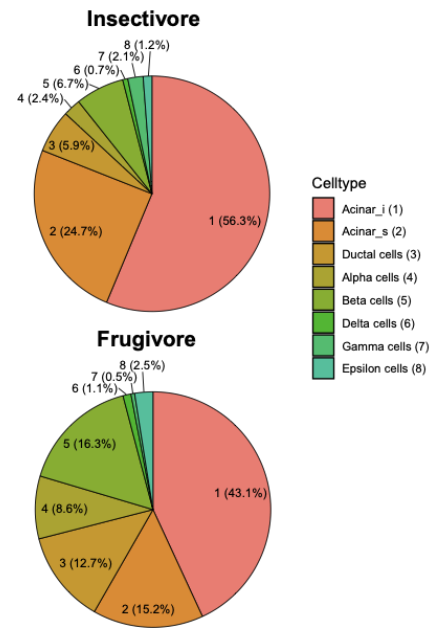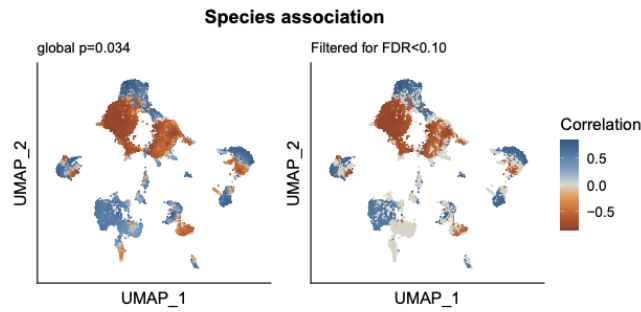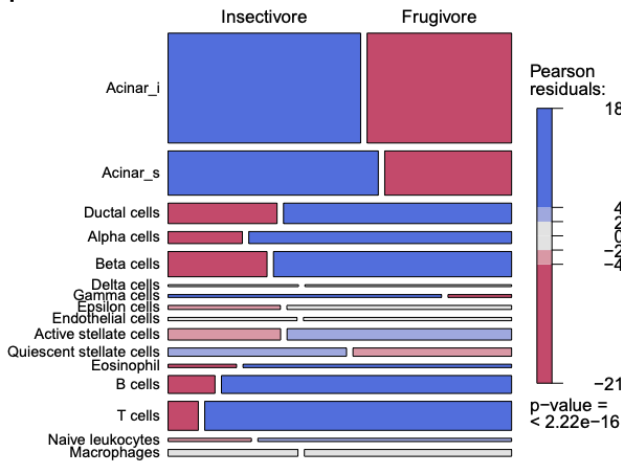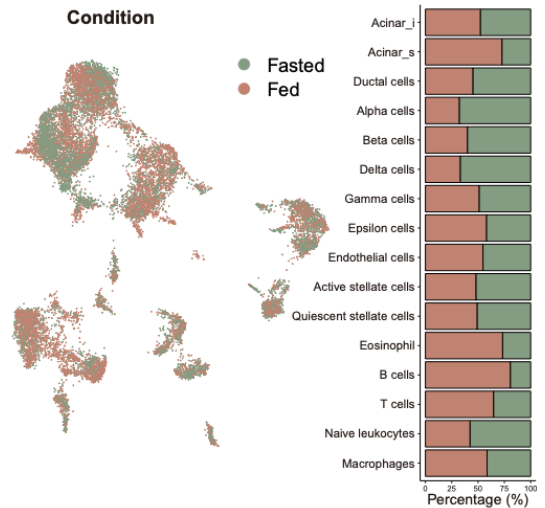
Supplementary Figure 3.6. Single-cell composition analysis of bat pancreas.

(**A**) UMAP of bat pancreas cell-types automatically annotated with human pancreas single-cell reference data from Azimuth(*110*). (**B**) Jaccard overlaps of auto-annotations in **A** (horizontal) with our integrated species annotations in **Fig. 4B** (vertical). (**C**) Violin plots of proximal tubule marker gene expression in bat proximal tubules and proximal tubules-like cells. (**D**) Pie charts of cell-type percentages across renal epithelial cells. (**E**) CNA of species associations to cell-types. (**F**) Pearson residuals visualized by mosaic plot of all cell-types identified in bat pancreases. (**G**) UMAP of bat pancreases cell-types by condition. Bar chart shows the proportion per condition for each cell-type.
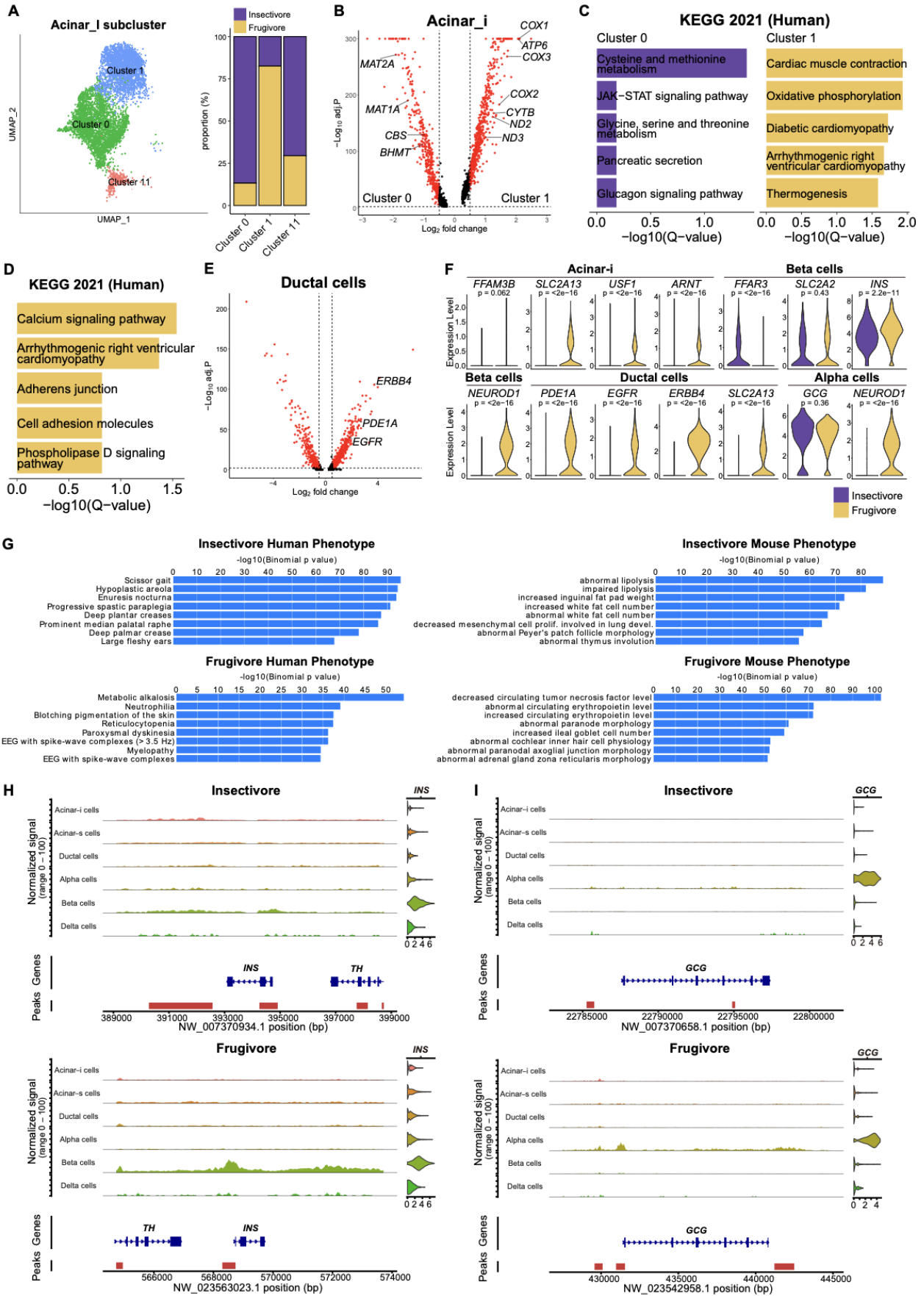
Supplementary Figure 3.7. scRNA-seq and scATAC-seq analysis of bat pancreas.

(**A**) (Left) UMAP of acinar-i subclusters. (Right) Proportion of species in each acinar-i subcluster. (**B**) Volcano plot showing differentially expressed genes between acinar-i subcluster 0 (insectivore-enriched) and subcluster 1 (frugivore-enriched). (**C**) Bar plots showing KEGG Human 2021 pathways enriched in acinar-i subcluster 0 (insectivore-enriched) and subcluster 1. (**D**) Bar plots showing KEGG Human 2021 pathways enriched in frugivore ductal cells. (**E**) Volcano plot showing differentially expressed genes between species in ductal cells. (**F**) Violin plots of differentially expressed genes, except *GCG*, in bat pancreas cells. (**G**) Bar plots showing GREAT human and mouse phenotypes enriched in insectivore and frugivore pancreases. (**H**) scATAC-seq coverage plots of *INS* in bat pancreases. SCTransform-normalized expression plot visualized on the right by cell-type. (**I**) scATAC-seq coverage plots of *GCG* in bat pancreases. SCTransform-normalized expression plot visualized on the right by cell-type.
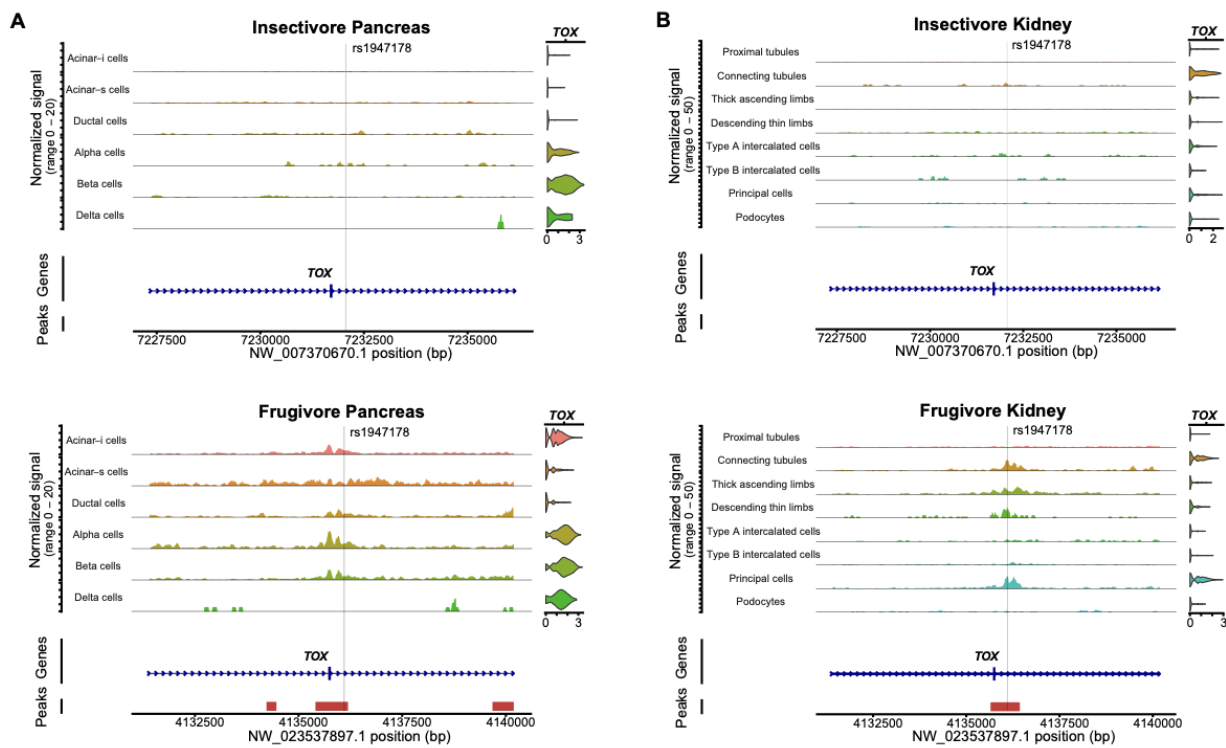

Supplementary Figure 3.8. scATAC-seq coverage plots of diabetes-associated SNP rs1947178 in bats (151,857 bp and 171,812 bp downstream of *TOX* TSS in insectivorous and frugivorous bats, respectively).

(**A-B**) scATAC-seq coverage plots of rs1947178 at *TOX* locus in bat pancreas (**A**) and kidney (**B**). SCTransform-normalized expression plot visualized on the right by cell-type.
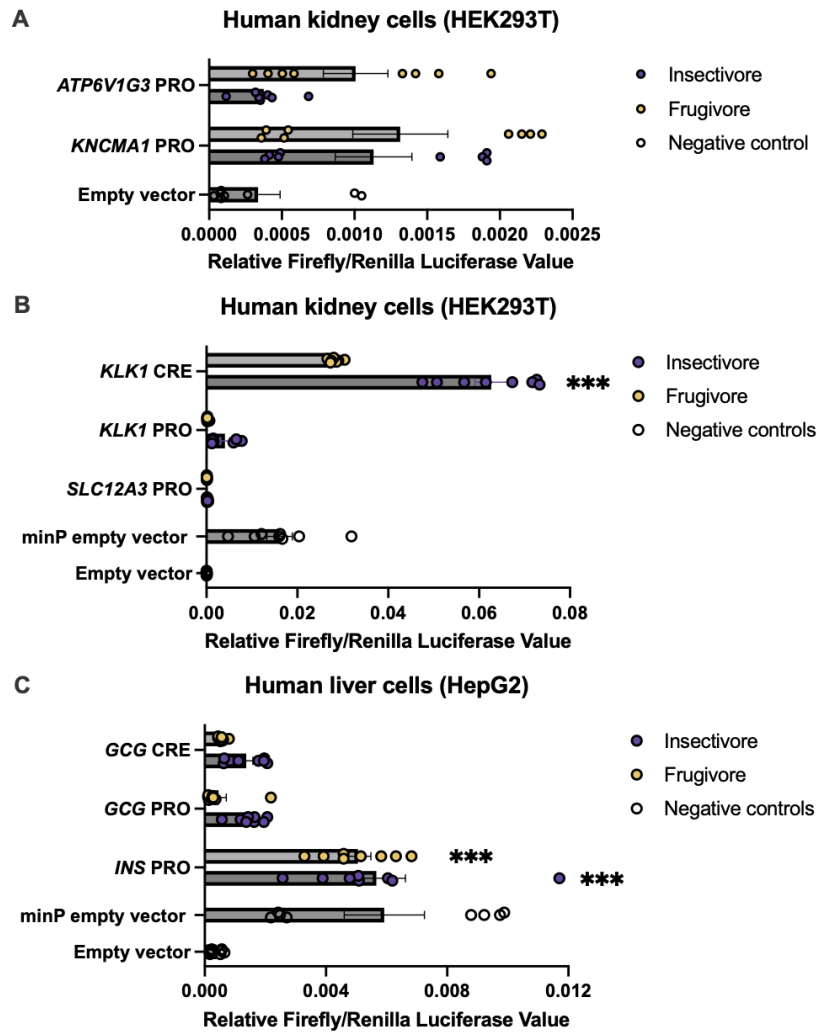
Supplementary Figure 3.9. Luciferase assays in human cells with bat promoters (PRO) and candidate *cis*-regulatory elements (CREs).

Bar charts showing relative firefly/renilla luciferase means ± SEM of bat regulatory regions in HEK293T (**A-B**) or HepG2 (**C**). Student's t-test vs. respective negative control ***p<.001. Minimal promoter is indicated by "minP."

Supplementary Table 3.1: Sequencing QC.
A1-2 = fasted fruit bats, A3-4 = treated fruit bats, K1-3 = treated insect bats, K4-5 = fasted insect bats, J1-2 = fasted fruit bats, J4= treated fruit bat, P1-2 = treated insect bats, P4-5 = fasted insect bats.

Sample	Total RNA read pairs	Mean RNA reads per cell	Total ATAC read pairs	Mean ATAC fragments per cell
A1	657490451	216921.9568	705761251	232847.6579
A2	543890027	126251.1669	740382314	171862.1899
A3	819522891	245734.0003	651414009	195326.5394
A4	774977825	289278.7701	730688551	272746.7529
E1	672912456	240411.7385	633695035	226400.5127
E3	759040415	385887.3488	787887849	400553.0498
E4	777840969	497022.9834	549534015	351139.9457
E5	639507539	144717.7051	807937847	182832.733
J1	642821902	305088.7053	726793113	344942.1514
J2	559726040	200762.5681	719196923	257961.5936
J4	578768902	123827.3218	949710083	203190.0049
P1	927901678	548405.247	635548081	375619.4332
P2	787957604	294014.0313	613644645	228971.8825
P4	770059917	637994.9602	502313147	416166.6504
P5	866020327	541601.2051	523034574	327101.0469
Mean	718562596.2	319861.3139	685169429.1	279177.4763
Std Dev.	114842825.9	165476.5497	117573079	82871.35383

Supplementary Table 3.2: Immunofluorescence antibodies and epitope matching.

Gene	Thermo Scientific Antibody Catalog #	Insect bat NCBI protein ID	Insect bat NCBI Protein Blast % Identity	Fruit bat NCBI protein ID	Fruit bat NCBI Protein Blast % Identity
SLC12A1	18970-1-AP	XP_008141339.1	87.01	XP_037012367.1	86.36
AQP2	PA5-78808	XP_008138879.1	100	XP_036992308.1	100
SLC26A4	PA5-115911	XP_008148295.1	85.19	XP_037006378.1	96.3
INS	15848-1-AP	XP_008158493.1	81.82	XP_037012244.1	86.36
GCG	15954-1-AP	XP_027996892.1	78.33	XP_036996293.1	93.33

Supplementary Table 3.3: Luciferase assay primers.

Region	Species (assembly)	Coordinates	Genomic DNA Primer Forward	Genomic DNA Primer Reverse
KCNMA1 PRO	*E. fuscus* (GCF_000308155.1)	NW_007370681.1:1736921-1738260	GAAAACCACCTGCTGCTGTC	GAACACCCATCAACAGCCAT
KCNMA1 PRO	*A. jamaicensis* (GCA_014825515.1)	NW_023580292.1:1519163-1520512	TCTACTGCCTAACGGACCTG	GCCATGTTGCTGctcctg
ATP6V1G3 PRO	*E. fuscus* (GCF_000308155.1)	NW_007370988.1:457744-458715	TTCTTTGGAACGTGGGCATG	GGGCTCTTACAGACAGACGT
ATP6V1G3 PRO	*A. jamaicensis* (GCA_014825515.1)	NW_023556122.1:3337626-3338370	AGTGCCCGTCAAGTCAAGT	TTTCACGGTCCCATCCTAGG
SLC12A3 PRO	*E. fuscus* (GCF_000308155.1)	NW_007370728.1:805729-807654	ACGCAGGGTGGCTTTATAGG	TACAGTCCCACCAGCATGTT
SLC12A3 PRO	*A. jamaicensis* (GCA_014825515.1)	NW_023544564.1:1738367-1740304	GGTGACTTTATAGGCTGGGC	GCGGGAAAGGCTCATAGAAG
KLK1 PRO	*E. fuscus* (GCF_000308155.1)	NW_007370751.1:1497150-1497621	GGGGCAGGGGTCATTAGTta	GAGCCTTTAAAAgccccgg
KLK1 PRO	A. jamaicensis (GCA_014825515.1)	NW_023554207.1:280824-281250	CAAGCAATGCAGGGAGAACA	TGAACCGGCAGCGGAGGA
KLK1 CRE	*E. fuscus* (GCF_000308155.1)	NW_007370751.1:1501496-1503222	GCTAGAACTCCACCCACTGT	GTAGTCCACTGCCAGTCACT
KLK1 CRE	*A. jamaicensis* (GCA_014825515.1)	NW_023554207.1:286057-287969	GGAGAGGTGACATCTGCGAT	CCCCTGGGTCTGTCTTTGAA
INS PRO	*E. fuscus* (GCF_000308155.1)	NW_007370934.1:394531-395103	CCCTACCCTGTGCTCACC	CGGAGCAGAGATGTCCCTC
INS PRO	*A. jamaicensis* (GCA_014825515.1)	NW_023563023.1:568133-568766	GAATCCTGGGTCCACAATGC	GACTGACCTGCTTGCTTCG
GCG PRO	*E. fuscus* (GCF_000308155.1)	NW_007370658.1:22785136-22787724	GCACCCCAAGAATCAGTGTT	AGCTCTGTACTTCTGCACCA
GCG PRO	*A. jamaicensis* (GCA_014825515.1)	NW_023542958.1:429464-431785	TGGTGGGCTAGAGATGAAGC	CATTTCGCAGCTTCATTGTGT
GCG CRE	*E. fuscus* (GCF_000308155.1)	NW_007370658.1:22796927-22798885	TAACGTCACCTACCAGCCAC	GACATGATTGCCTAGTTTCAAGC
GCG CRE	*A. jamaicensis* (GCA_014825515.1)	NW_023542958.1:440782-442545	TATGCCACTGTTCAGGACCA	TCAAGCAATTGTGAATTTGTGaagt

References

1. J. C. Motta Junior, K. Martins, "The frugivorous diet of the maned wolf, *Chrysocyon brachyurus* , in Brazil: ecology and conservation" in *Seed dispersal and frugivory: ecology, evolution and conservation. Third International Symposium-Workshop on Frugivores and Seed Dispersal, São Pedro, Brazil, 6-11 August 2000* (CABI Publishing, UK, 2002), pp. 291–303.

2. A. Peredo, D. Martínez, J. Rodríguez-Pérez, D. García, Mammalian seed dispersal in Cantabrian woodland pastures: Network structure and response to forest loss. *Basic Appl. Ecol.* **14**, 378–386 (2013).

3. F. Meng, L. Zhu, W. Huang, D. M. Irwin, S. Zhang, Bats: Body mass index, forearm mass index, blood glucose levels and SLC2A2 genes for diabetes. *Sci. Rep.* **6**, 29960 (2016).

4. A. R. DeCasien, S. A. Williams, J. P. Higham, Primate brain size is predicted by diet but not sociality. *Nat Ecol Evol*. **1**, 112 (2017).

5. S. Koike, T. Masaki, Characteristics of fruits consumed by mammalian frugivores in Japanese temperate forest. *Ecol. Res.* **34**, 246–254 (2019).

6. P. Perelman, W. E. Johnson, C. Roos, H. N. Seuánez, J. E. Horvath, M. A. M. Moreira, B. Kessing, J. Pontius, M. Roelke, Y. Rumpler, M. P. C. Schneider, A. Silva, S. J. O'Brien, J. Pecon-Slattery, A molecular phylogeny of living primates. *PLoS Genet.* **7**, e1001342 (2011).

7. K. Wang, S. Tian, J. Galindo-González, L. M. Dávalos, Y. Zhang, H. Zhao, Molecular adaptation and convergent evolution of frugivory in Old World and neotropical fruit bats. *Mol. Ecol.* **29**, 4366–4381 (2020).

8. G. F. Gunnell, N. B. Simmons, Fossil Evidence and the Origin of Bats. *J. Mamm. Evol.* **12**, 209–246 (2005).

9. C. A. Emerling, F. Delsuc, M. W. Nachman, Chitinase genes (*CHIA*s) provide genomic footprints of a post-Cretaceous dietary radiation in placental mammals. *Science Advances*. **4**, eaar6478 (2018).

10. D. Rojas, M. A. Borrero-Ospina, Ó. E. Murillo-García, Coevolution of brain and palate during the diversification of specialized frugivorous bats. *Biol. J. Linn. Soc. Lond.* **136**, 346–353 (2022).

11. R. A. Barton, Visual specialization and brain evolution in primates. *Proceedings of the Royal Society of London. Series B: Biological Sciences*. **265**, 1933–1937 (1998).

12. J. J. C. Ramakers, D. K. N. Dechmann, R. A. Page, M. T. O'Mara, Frugivorous bats prefer information from novel social partners. *Anim. Behav.* **116**, 83–87 (2016).

13. E. Bachorec, I. Horáček, P. Hulva, A. Konečný, R. K. Lučan, P. Jedlička, W. M. Shohdi, Š. Řeřucha, M. Abi-Said, T. Bartonička, Spatial networks differ when food supply changes: Foraging strategy of Egyptian fruit bats. *PLoS One*. **15**, e0229110 (2020).

14. A. Sadier, K. T. Davies, L. R. Yohe, K. Yun, P. Donat, B. P. Hedrick, E. R. Dumont, L. M. Dávalos, S. J. Rossiter, K. E. Sears, Multifactorial processes underlie parallel opsin loss in neotropical bats. *Elife*. **7** (2018), doi:10.7554/eLife.37412.

15. V. Sharma, N. Hecker, J. G. Roscito, L. Foerster, B. E. Langer, M. Hiller, A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. *Nat. Commun.* **9**, 1215 (2018).

16. J. T. Kratzer, M. A. Lanaspa, M. N. Murphy, C. Cicerchi, C. L. Graves, P. A. Tipton, E. A. Ortlund, R. J. Johnson, E. A. Gaucher, Evolutionary history and metabolic insights of ancient mammalian uricases. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 3763–3768 (2014).

17. D. J. Keegan, Aspects of the assimilation of sugars by Rousettus aegyptiacus. *Comp. Biochem. Physiol. A Physiol.* **58**, 349–352 (1977).

18. O. Amitai, S. Holtze, S. Barkan, E. Amichai, C. Korine, B. Pinshow, C. C. Voigt, Fruit bats (Pteropodidae) fuel their metabolism rapidly and directly with exogenous sugars. *J. Exp. Biol.* **213**, 2693–2699 (2010).

19. M. B. Freitas, J. F. Queiroz, C. I. Dias Gomes, C. B. Collares-Buzato, H. C. Barbosa, A. C. Boschero, C. A. Gonçalves, E. C. Pinheiro, Reduced insulin secretion and glucose intolerance are involved in the fasting susceptibility of common vampire bats. *Gen. Comp. Endocrinol.* **183**, 1–6 (2013).

20. X. Peng, X. He, Y. Sun, J. Liang, H. Xie, Difference in glucose tolerance between phytophagous and insectivorous bats. *Journal of Comparative Physiology. B, Biochemical, Systemic, and Environmental Physiology; Heidelberg*. **189**, 751–756 (2019).

21. B. Shen, X. Han, J. Zhang, S. J. Rossiter, S. Zhang, Adaptive evolution in the glucose transporter 4 gene Slc2a4 in Old World fruit bats (family: Pteropodidae). *PLoS One*. **7**, e33197 (2012).

22. J. E. Schondube, L. G. Herrera-M, C. Martínez del Rio, Diet and the evolution of digestion and renal function in phyllostomid bats. *Zoology* . **104**, 59–73 (2001).

23. A. N. Makanya, J. N. Maina, T. M. Mayhew, S. A. Tschanz, P. H. Burri, A stereological comparison of villous and microvillous surfaces in small intestines of frugivorous and entomophagous bats: species, inter-individual and craniocaudal differences. *J. Exp. Biol.* **200**, 2415–2423 (1997).

24. D. J. Chivers, C. M. Hladik, Morphology of the gastrointestinal tract in primates:

comparisons with other mammals in relation to diet. *J. Morphol.* **166**, 337–386 (1980).

25. R. Gadelha-Alves, A. M. D. S. Rozensztranch, O. Rocha- Barbosa, Comparative Intestinal Histomorphology of Five Species of Phyllostomid Bats (Phyllostomidae, Microchiroptera): Ecomorphological Relations with Alimentary Habits. *Int. J. Morphol.* **26***, 591–602 (2008).

26. J. E. Lambert, "Evolutionary Biology of Ape and Monkey Feeding and Nutrition" in *Handbook of Paleoanthropology: Vol I:Principles, Methods and Approaches Vol II:Primate Evolution and Human Origins Vol III:Phylogeny of Hominids*, W. Henke, I. Tattersall, Eds. (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013), pp. 1–27.

27. J. E. Lambert, "Evolutionary Biology of Ape and Monkey Feeding and Nutrition" in *Handbook of Paleoanthropology*, W. Henke, I. Tattersall, Eds. (Springer Berlin Heidelberg, Berlin, Heidelberg, 2015), pp. 1631–1660.

28. E. R. Dumont, Cranial shape in fruit, nectar, and exudate feeders: implications for interpreting the fossil record. *Am. J. Phys. Anthropol.* **102**, 187–202 (1997).

29. M. R. Marchán-Rivadeneira, C. J. Phillips, R. E. Strauss, J. Antonio Guerrero, C. A. Mancina, R. J. Baker, Cranial differentiation of fruit-eating bats (GenusArtibeus) based on size-standardized data. *Acta Chiropt.* **12**, 143–154 (2010).

30. E. R. Dumont, K. Samadevam, I. Grosse, O. M. Warsi, B. Baird, L. M. Davalos, Selection for mechanical advantage underlies multiple cranial optima in new world leaf-nosed bats. *Evolution*. **68**, 1436–1449 (2014).

31. J. H. Arbour, A. A. Curtis, S. E. Santana, Signatures of echolocation and dietary ecology in the adaptive evolution of skull shape in bats. *Nat. Commun.* **10**, 1–13 (2019).

32. L. V. García-Herrera, L. A. Ramírez-Fráncel, G. Guevara, G. Reinoso-Flórez, A.

Sánchez-Hernández, B. K. Lim, S. Losada-Prado, Foraging strategies, craniodental traits, and interaction in the bite force of Neotropical frugivorous bats (Phyllostomidae: Stenodermatinae). *Ecol. Evol.* **11**, 13756–13772 (2021).

33. D. Massoud, M. M. A. Abumandour, Anatomical features of the tongue of two chiropterans endemic in the Egyptian fauna; the Egyptian fruit bat (Rousettus aegyptiacus) and insectivorous bat (Pipistrellus kuhlii). *Acta Histochem.* **122**, 151503 (2020).

34. J. F. Pastor, M. N. Muchlinski, J. M. Potau, A. Casado, Y. García-Mesa, J. A. Vega, R. Cabo, The Tongue in Three Species of Lemurs: Flower and Nectar Feeding Adaptations. *Animals (Basel)*. **11** (2021), doi:10.3390/ani11102811.

35. G. Casotti, L. Gerardo Herrera M, J. J. Flores M, C. A. Mancina, E. J. Braun, Relationships between renal morphology and diet in 26 species of new world bats (suborder microchiroptera). *Zoology* . **109**, 196–207 (2006).

36. D. Osorio, M. Vorobyev, Colour vision as an adaptation to frugivory in primates. *Proc. Biol. Sci.* **263**, 593–599 (1996).

37. A. D. Melin, D. W. Kline, C. M. Hickey, L. M. Fedigan, Food search through the eyes of a monkey: a functional substitution approach for assessing the ecology of primate color vision. *Vision Res.* **86**, 87–96 (2013).

38. A. D. Melin, K. L. Chiou, E. R. Walco, M. L. Bergstrom, S. Kawamura, L. M. Fedigan, Trichromacy increases fruit intake rates of wild capuchins (Cebus capucinus imitator). *Proc. Natl. Acad. Sci. U. S. A.* **114**, 10402–10407 (2017).

39. R. E. Onstein, D. N. Vink, J. Veen, C. D. Barratt, S. G. A. Flantua, S. A. Wich, W. D. Kissling, Palm fruit colours are linked to the broad-scale distribution and diversification of

primate colour vision systems. *Proc. Biol. Sci.* **287**, 20192731 (2020).

40. K. T. J. Davies, L. R. Yohe, J. Almonte, M. K. R. Sánchez, E. M. Rengifo, E. R. Dumont, K. E. Sears, L. M. Dávalos, S. J. Rossiter, Foraging shifts and visual preadaptation in ecologically diverse bats. *Mol. Ecol.* **29**, 1839–1859 (2020).

41. N. J. Dominy, Fruits, Fingers, and Fermentation: The Sensory Cues Available to Foraging Primates1. *Integr. Comp. Biol.* **44**, 295–303 (2004).

42. C. Hiramatsu, A. D. Melin, F. Aureli, C. M. Schaffner, M. Vorobyev, S. Kawamura, Interplay of olfaction and vision in fruit foraging of spider monkeys. *Anim. Behav.* **77**, 1421–1426.

43. O. Nevo, D. Razafimandimby, J. A. J. Jeffrey, S. Schulz, M. Ayasse, Fruit scent as an evolved signal to primate seed dispersal. *Sci Adv*. **4**, eaat4871 (2018).

44. K. Kshitish Acharya, A. Roy, A. Krishna, Relative role of olfactory cues and certain non-olfactory factors in foraging of fruit-eating bats. *Behav. Processes*. **44**, 59–64 (1998).

45. F. Sánchez, C. Korine, M. Steeghs, L.-J. Laarhoven, S. M. Cristescu, F. J. M. Harren, R. Dudley, B. Pinshow, Ethanol and methanol as possible odor cues for Egyptian fruit bats (Rousettus aegyptiacus). *J. Chem. Ecol.* **32**, 1289–1300 (2006).

46. Y. Liu, H. Xu, X. Yuan, S. J. Rossiter, S. Zhang, Multiple adaptive losses of alanine-glyoxylate aminotransferase mitochondrial targeting in fruit-eating bats. *Mol. Biol. Evol.* **29**, 1507–1511 (2012).

47. B. Shen, X. Han, G. Jones, S. J. Rossiter, S. Zhang, Adaptive evolution of the myo6 gene in old world fruit bats (family: pteropodidae). *PLoS One*. **8**, e62307 (2013).

48. L. Fang, B. Shen, D. M. Irwin, S. Zhang, Parallel evolution of the glycogen synthase 1 (muscle) gene Gys1 between Old World and New World fruit bats (Order: Chiroptera).

Biochem. Genet. **52**, 443–458 (2014).

49. Y. Qian, T. Fang, B. Shen, S. Zhang, The glycogen synthase 2 gene (Gys2) displays parallel evolution between Old World and New World fruit bats. *J. Mol. Evol.* **78**, 66–74 (2014).

50. B. Shen, T. Fang, T. Yang, G. Jones, D. M. Irwin, S. Zhang, Relaxed evolution in the tyrosine aminotransferase gene tat in old world fruit bats (Chiroptera: Pteropodidae). *PLoS One*. **9**, e97483 (2014).

51. L. Zhu, Q. Yin, D. M. Irwin, S. Zhang, Phosphoenolpyruvate carboxykinase 1 gene (Pck1) displays parallel evolution between Old World and New World fruit bats. *PLoS One*. **10**, e0118666 (2015).

52. Q. Yin, L. Zhu, D. Liu, D. M. Irwin, S. Zhang, Y.-H. Pan, Molecular Evolution of the Nuclear Factor (Erythroid-Derived 2)-Like 2 Gene Nrf2 in Old World Fruit Bats (Chiroptera: Pteropodidae). *PLoS One*. **11**, e0146274 (2016).

53. J. H. T. Potter, K. T. J. Davies, L. R. Yohe, M. K. R. Sanchez, E. M. Rengifo, M. Struebig, K. Warren, G. Tsagkogeorga, B. K. Lim, M. Dos Reis, L. M. Dávalos, S. J. Rossiter, Dietary Diversification and Specialization in Neotropical Bats Facilitated by Early Molecular Evolution. *Mol. Biol. Evol.* **38**, 3864–3883 (2021).

54. D. D. Moreno-Santillán, C. Machain-Williams, G. Hernández-Montes, J. Ortega, De Novo Transcriptome Assembly and Functional Annotation in Five Species of Bats. *Sci. Rep.* **9**, 6222 (2019).

55. F. Luca, G. H. Perry, A. Di Rienzo, Evolutionary adaptations to dietary changes. *Annu. Rev. Nutr.* **30**, 291–314 (2010).

56. A. Levchenko, A. Kanapin, A. Samsonova, R. R. Gainetdinov, Human Accelerated Regions and Other Human-Specific Sequence Variations in the Context of Evolution and Their Relevance for Brain Development. *Genome Biol. Evol.* **10**, 166–188 (2018).

57. F. Inoue, N. Ahituv, Decoding enhancers using massively parallel reporter assays. *Genomics*. **106**, 159–164 (2015).

58. Z. Hu, T. B. Sackton, S. V. Edwards, J. S. Liu, Bayesian Detection of Convergent Rate Changes of Conserved Noncoding Elements on Phylogenetic Trees. *Mol. Biol. Evol.* **36**, 1086–1100 (2019).

59. N. Hecker, M. Hiller, A genome alignment of 120 mammals highlights ultraconserved element variability and placenta-associated enhancers. *Gigascience*. **9**, giz159 (2020).

60. E. C. Teeling, S. C. Vernes, L. M. Dávalos, D. A. Ray, M. T. P. Gilbert, E. Myers, Bat1K Consortium, Bat Biology, Genomes, and the Bat1K Project: To Generate Chromosome-Level Genomes for All Living Bat Species. *Annu Rev Anim Biosci*. **6**, 23–46 (2018).

61. R. S. Harris, thesis, The Pennsylvania State University (2007).

62. A. S. Hinrichs, D. Karolchik, R. Baertsch, G. P. Barber, G. Bejerano, H. Clawson, M. Diekhans, T. S. Furey, R. A. Harte, F. Hsu, J. Hillman-Jackson, R. M. Kuhn, J. S. Pedersen, A. Pohl, B. J. Raney, K. R. Rosenbloom, A. Siepel, K. E. Smith, C. W. Sugnet, A. Sultan-Qurraie, D. J. Thomas, H. Trumbower, R. J. Weber, M. Weirauch, A. S. Zweig, D. Haussler, W. J. Kent, The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590–8 (2006).

63. M. Blanchette, W. J. Kent, C. Riemer, L. Elnitski, A. F. A. Smit, K. M. Roskin, R. Baertsch,

K. Rosenbloom, H. Clawson, E. D. Green, D. Haussler, W. Miller, Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**, 708–715 (2004).

64. A. Siepel, G. Bejerano, J. S. Pedersen, A. S. Hinrichs, M. Hou, K. Rosenbloom, H. Clawson, J. Spieth, L. W. Hillier, S. Richards, G. M. Weinstock, R. K. Wilson, R. A. Gibbs, W. J. Kent, W. Miller, D. Haussler, Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).

65. B. M. Booker, T. Friedrich, M. K. Mason, J. E. VanderMeer, J. Zhao, W. L. Eckalbar, M. Logan, N. Illing, K. S. Pollard, N. Ahituv, Bat Accelerated Regions Identify a Bat Forelimb Specific Enhancer in the HoxD Locus. *PLoS Genet.* **12**, e1005738 (2016).

66. ADW: Home, (available at https://animaldiversity.org/).

67. *Hu-etal-2019-data: Data, result, and simulation files from Hu et al. 2019* (Github; https://github.com/phyloacc/Hu-etal-2019-data).

68. C. Y. McLean, D. Bristor, M. Hiller, S. L. Clarke, B. T. Schaar, C. B. Lowe, A. M. Wenger, G. Bejerano, GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).

69. T. L. Bailey, J. Johnson, C. E. Grant, W. S. Noble, The MEME Suite. *Nucleic Acids Res.* **43**, W39–49 (2015).

70. J. A. Castro-Mondragon, R. Riudavets-Puig, I. Rauluseviciute, R. B. Lemma, L. Turchi, R. Blanc-Mathieu, J. Lucas, P. Boddie, A. Khan, N. Manosalva Pérez, O. Fornes, T. Y. Leung, A. Aguirre, F. Hammal, D. Schmelter, D. Baranasic, B. Ballester, A. Sandelin, B. Lenhard, K. Vandepoele, W. W. Wasserman, F. Parcy, A. Mathelier, JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **50**,

D165–D173 (2022).

71. L. A. Pennacchio, N. Ahituv, A. M. Moses, S. Prabhakar, M. A. Nobrega, M. Shoukry, S. Minovitsky, I. Dubchak, A. Holt, K. D. Lewis, I. Plajzer-Frick, J. Akiyama, S. De Val, V. Afzal, B. L. Black, O. Couronne, M. B. Eisen, A. Visel, E. M. Rubin, In vivo enhancer analysis of human conserved non-coding sequences. *Nature*. **444**, 499–502 (2006).

72. T. Sakuma, S. Nakade, Y. Sakane, K.-I. T. Suzuki, T. Yamamoto, MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems. *Nat. Protoc.* **11**, 118–133 (2016).

73. A. Ushiki, Y. Zhang, C. Xiong, J. Zhao, I. Georgakopoulos-Soares, L. Kane, K. Jamieson, M. J. Bamshad, D. A. Nickerson, University of Washington Center for Mendelian Genomics, Y. Shen, L. A. Lettice, E. L. Silveira-Lucas, F. Petit, N. Ahituv, Deletion of CTCF sites in the SHH locus alters enhancer-promoter interactions and leads to acheiropodia. *Nat. Commun.* **12**, 2282 (2021).

74. G. Jones, E. C. Teeling, The evolution of echolocation in bats. *Trends Ecol. Evol.* **21**, 149–156 (2006).

75. A. O. P. Protzek, A. Rafacho, B. A. Viscelli, J. R. Bosqueiro, A. P. Cappelli, F. M. M. Paula, A. C. Boschero, E. C. Pinheiro, Insulin and glucose sensitivity, insulin secretion and β-cell distribution in endocrine pancreas of the fruit bat Artibeus lituratus. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **157**, 142–148 (2010).

76. J. R. Friedman, G. K. Voeltz, The ER in 3D: a multifunctional dynamic membrane network. *Trends Cell Biol.* **21**, 709–717 (2011).

77. L. M. Westrate, J. E. Lee, W. A. Prinz, G. K. Voeltz, Form follows function: the importance

of endoplasmic reticulum shape. *Annu. Rev. Biochem.* **84**, 791–811 (2015).

78. K. P. West, I. Darnton-Hill, "Vitamin A Deficiency" in *Nutrition and Health in Developing Countries*, R. D. Semba, M. W. Bloem, P. Piot, Eds. (Humana Press, Totowa, NJ, 2008), pp. 377–433.

79. B. Hentsch, I. Lyons, R. Li, L. Hartley, T. J. Lints, J. M. Adams, R. P. Harvey, Hlx homeo box gene is essential for an inductive tissue interaction that drives expansion of embryonic liver and gut. *Genes Dev.* **10**, 70–79 (1996).

80. M. J. Geske, X. Zhang, K. K. Patel, D. M. Ornitz, T. S. Stappenbeck, Fgf9 signaling regulates small intestinal elongation and mesenchymal development. *Development*. **135**, 2959–2968 (2008).

81. H. Barak, S.-H. Huh, S. Chen, C. Jeanpierre, J. Martinovic, M. Parisot, C. Bole-Feysot, P. Nitschké, R. Salomon, C. Antignac, D. M. Ornitz, R. Kopan, FGF9 and FGF20 maintain the stemness of nephron progenitors in mice and man. *Dev. Cell*. **22**, 1191–1207 (2012).

82. S. Harel, E. Y. Tu, S. Weisberg, M. Esquilin, S. M. Chambers, B. Liu, C. T. Carson, L. Studer, B. Reizis, M. J. Tomishima, ZFX controls the self-renewal of human embryonic stem cells. *PLoS One*. **7**, e42302 (2012).

83. X. Yan, L. Yan, Z. Su, Q. Zhu, S. Liu, Z. Jin, Y. Wang, Zinc-finger protein X-linked is a novel predictor of prognosis in patients with colorectal cancer. *Int. J. Clin. Exp. Pathol.* **7**, 3150–3157 (2014).

84. C. Li, H. Li, T. Zhang, J. Li, F. Ma, M. Li, Z. Sui, J. Chang, ZFX is a Strong Predictor of Poor Prognosis in Renal Cell Carcinoma. *Med. Sci. Monit.* **21**, 3380–3385 (2015).

85. A. C. Beck, E. Cho, J. R. White, L. Paemka, T. Li, V. W. Gu, D. T. Thompson, K. E. Koch, C.

Franke, M. Gosse, V. T. Wu, S. R. Landers, A. J. Pamatmat, M. V. Kulak, R. J. Weigel, AP-2α Regulates S-Phase and Is a Marker for Sensitivity to PI3K Inhibitor Buparlisib in Colon Cancer. *Mol. Cancer Res.* **19**, 1156–1167 (2021).

86. J. Xie, D. Jin, J. Xu, F. Yang, J. Jin, Hsa_hsa_circ_0081069 promotes the progression of colorectal cancer through sponging miR-665 and regulating E2F3 expression. *J. Clin. Lab. Anal.* **36**, e24710 (2022).

87. Z. Xu, H. Qu, Y. Ren, Z. Gong, H. J. Ri, F. Zhang, S. Shao, X. Chen, X. Chen, Systematic Analysis of E2F Expression and Its Relation in Colorectal Cancer Prognosis. *Int. J. Gen. Med.* **15**, 4849–4870 (2022).

88. P. J. Mitchell, P. M. Timmons, J. M. Hébert, P. W. Rigby, R. Tjian, Transcription factor AP-2 is expressed in neural crest cell lineages during mouse embryogenesis. *Genes Dev.* **5**, 105–119 (1991).

89. B. E. Chambers, G. F. Gerlach, E. G. Clark, K. H. Chen, A. E. Levesque, I. Leshchiner, W. Goessling, R. A. Wingert, Tfap2a is a novel gatekeeper of nephron differentiation during kidney development. *Development*. **146** (2019), doi:10.1242/dev.172387.

90. J. O. Lamontagne, H. Zhang, A. M. Zeid, K. Strittmatter, A. D. Rocha, T. Williams, S. Zhang, A. G. Marneros, Transcription factors AP-2α and AP-2β regulate distinct segments of the distal nephron in the mammalian kidney. *Nat. Commun.* **13**, 2226 (2022).

91. W.-I. Choi, J.-H. Yoon, J.-Y. Song, B.-N. Jeon, J.-M. Park, D.-I. Koh, Y.-H. Ahn, K.-S. Kim, I.-K. Lee, M.-W. Hur, Zbtb7c is a critical gluconeogenic transcription factor that induces glucose-6-phosphatase and phosphoenylpyruvate carboxykinase 1 genes expression during mice fasting. *Biochim. Biophys. Acta Gene Regul. Mech.* **1862**, 643–656 (2019).

92. L. Hartmann, S. Dutta, S. Opatz, S. Vosberg, K. Reiter, G. Leubolt, K. H. Metzeler, T. Herold, S. A. Bamopoulos, K. Bräundl, E. Zellmeier, B. Ksienzyk, N. P. Konstandin, S. Schneider, K.-P. Hopfner, A. Graf, S. Krebs, H. Blum, J. M. Middeke, F. Stölzel, C. Thiede, S. Wolf, S. K. Bohlander, C. Preiss, L. Chen-Wichmann, C. Wichmann, M. C. Sauerland, T. Büchner, W. E. Berdel, B. J. Wörmann, J. Braess, W. Hiddemann, K. Spiekermann, P. A. Greif, ZBTB7A mutations in acute myeloid leukaemia with t(8;21) translocation. *Nat. Commun.* **7**, 11733 (2016).

93. I. Piragyte, T. Clapes, A. Polyzou, R. I. Klein Geltink, S. Lefkopoulos, N. Yin, P. Cauchy, J. D. Curtis, L. Klaeylé, X. Langa, C. C. A. Beckmann, M. W. Wlodarski, P. Müller, D. Van Essen, A. Rambold, F. G. Kapp, M. Mione, J. M. Buescher, E. L. Pearce, A. Polyzos, E. Trompouki, A metabolic interplay coordinated by HLX regulates myeloid differentiation and AML through partly overlapping pathways. *Nature Communications*. **9** (2018), , doi:10.1038/s41467-018-05311-4.

94. A. D. Rouillard, G. W. Gundersen, N. F. Fernandez, Z. Wang, C. D. Monteiro, M. G. McDermott, A. Ma'ayan, The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database* . **2016** (2016), doi:10.1093/database/baw100.

95. Y. Luo, B. C. Hitz, I. Gabdank, J. A. Hilton, M. S. Kagda, B. Lam, Z. Myers, P. Sud, J. Jou, K. Lin, U. K. Baymuradov, K. Graham, C. Litton, S. R. Miyasato, J. S. Strattan, O. Jolanki, J.-W. Lee, F. Y. Tanaka, P. Adenekan, E. O'Neill, J. M. Cherry, New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res.* **48**, D882–D889 (2020).

96. R. Kothary, S. Clapoff, A. Brown, R. Campbell, A. Peterson, J. Rossant, A transgene containing lacZ inserted into the dystonia locus is expressed in neural tube. *Nature*. **335**,

435–437 (1988).

97. V. Govindarajan, P. A. Overbeek, FGF9 can induce endochondral ossification in cranial mesenchyme. *BMC Dev. Biol.* **6**, 7 (2006).

98. I. H. Hung, K. Yu, K. J. Lavine, D. M. Ornitz, FGF9 regulates early hypertrophic chondrocyte differentiation and skeletal vascularization in the developing stylopod. *Dev. Biol.* **307**, 300–313 (2007).

99. F. Zhao, L. Zhang, M. Zhang, J. Huang, J. Zhang, Y. Chang, FGF9 Alleviates the Fatty Liver Phenotype by Regulating Hepatic Lipid Metabolism. *Front. Pharmacol.* **13**, 850128 (2022).

100. N. Matharu, N. Ahituv, Modulating gene regulation to treat genetic disorders. *Nat. Rev. Drug Discov.* **19**, 757–775 (2020).

101. J. E. Campbell, C. B. Newgard, Mechanisms controlling pancreatic islet cell function in insulin secretion. *Nat. Rev. Mol. Cell Biol.* **22**, 142–158 (2021).

102. M. Karpińska, M. Czauderna, Pancreas-Its Functions, Disorders, and Physiological Impact on the Mammals' Organism. *Front. Physiol.* **13**, 807632 (2022).

103. A. J. Michelmore, D. J. Keegan, B. Kramer, Immunocytochemical Identification of Endocrine Cells in the Pancreas of the Fruit Bat,Rousettus aegyptiacus. *General and Comparative Endocrinology*. **110** (1998), pp. 319–325.

104. M. S. Balzer, T. Rohacs, K. Susztak, How Many Cell Types Are in the Kidney and What Do They Do? *Annu. Rev. Physiol.* **84**, 507–531 (2022).

105. I. H. de Boer, K. M. Utzschneider, The kidney's role in systemic metabolism-still much to learn. *Nephrol. Dial. Transplant.* **32**, 588–590 (2017).

106. Z. Arad, C. Korine, Effect of water restriction on energy and water balance and osmoregulation of the fruit bat Rousettus aegyptiacus. *J. Comp. Physiol. B*. **163**, 401–405 (1993).

107. J.-J. Jin, W.-B. Yu, J.-B. Yang, Y. Song, C. W. dePamphilis, T.-S. Yi, D.-Z. Li, GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **21**, 241 (2020).

108. M. Bernt, A. Donath, F. Jühling, F. Externbrink, C. Florentz, G. Fritzsch, J. Pütz, M. Middendorf, P. F. Stadler, MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* **69**, 313–319 (2013).

109. D. M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).

110. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck 3rd, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data. *Cell*. **184**, 3573–3587.e29 (2021).

111. T. Stuart, A. Srivastava, S. Madad, C. A. Lareau, R. Satija, Single-cell chromatin state analysis with Signac. *Nat. Methods*. **18**, 1333–1341 (2021).

112. C. Hafemeister, R. Satija, Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).

113. I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, F. Zhang, K. Wei, Y. Baglaenko, M. Brenner, P.-R. Loh, S. Raychaudhuri, Fast, sensitive and accurate integration of single-cell

data with Harmony. *Nat. Methods*. **16**, 1289–1296 (2019).

114. H. Wickham, ggplot2: Elegant Graphics for Data Analysis (2016), (available at https://ggplot2.tidyverse.org).

115. Y. A. Reshef, L. Rumker, J. B. Kang, A. Nathan, I. Korsunsky, S. Asgari, M. B. Murray, D. B. Moody, S. Raychaudhuri, Co-varying neighborhood analysis identifies cell populations associated with phenotypes of interest from single-cell transcriptomics. *Nat. Biotechnol.* **40**, 355–363 (2022).

116. D. Meyer, A. Zeileis, K. Hornik, The Strucplot Framework: Visualizing Multi-way Contingency Tables with vcd. *J. Stat. Softw.* **17**, 1–48 (2007).

117. B. B. Lake, R. Menon, S. Winfree, Q. Hu, R. M. Ferreira, K. Kalhor, D. Barwinska, E. A. Otto, M. Ferkowicz, D. Diep, N. Plongthongkum, A. Knoten, S. Urata, A. S. Naik, S. Eddy, B. Zhang, Y. Wu, D. Salamon, J. C. Williams, X. Wang, K. S. Balderrama, P. Hoover, E. Murray, A. Vijayan, F. Chen, S. S. Waikar, S. Rosas, F. P. Wilson, P. M. Palevsky, K. Kiryluk, J. R. Sedor, R. D. Toto, C. Parikh, E. H. Kim, E. Z. Macosko, P. V. Kharchenko, J. P. Gaut, J. B. Hodgin, M. T. Eadon, P. C. Dagher, T. M. El-Achkar, K. Zhang, M. Kretzler, S. Jain, for the KPMP consortium, An atlas of healthy and injured cell states and niches in the human kidney. *bioRxiv* (2021), p. 2021.07.28.454201.

118. Z. Miao, M. S. Balzer, Z. Ma, H. Liu, J. Wu, R. Shrestha, T. Aranyi, A. Kwan, A. Kondo, M. Pontoglio, J. Kim, M. Li, K. H. Kaestner, K. Susztak, Single cell regulatory landscape of the mouse kidney highlights cellular differentiation programs and disease targets. *Nat. Commun.* **12**, 2277 (2021).

119. L. Tosti, Y. Hang, O. Debnath, S. Tiesmeyer, T. Trefzer, K. Steiger, F. W. Ten, S. Lukassen, S. Ballke, A. A. Kühl, S. Spieckermann, R. Bottino, N. Ishaque, W. Weichert, S. K. Kim, R.

Eils, C. Conrad, Single-Nucleus and In Situ RNA–Sequencing Reveal Cell Topographies in the Human Pancreas. *Gastroenterology*. **160**, 1330–1344.e11 (2021).

120. D. Grün, M. J. Muraro, J.-C. Boisset, K. Wiebrands, A. Lyubimova, G. Dharmadhikari, M. van den Born, J. van Es, E. Jansen, H. Clevers, E. J. P. de Koning, A. van Oudenaarden, De Novo Prediction of Stem Cell Identity using Single-Cell Transcriptome Data. *Cell Stem Cell*. **19**, 266–277 (2016).

121. M. J. Muraro, G. Dharmadhikari, D. Grün, N. Groen, T. Dielen, E. Jansen, L. van Gurp, M. A. Engelse, F. Carlotti, E. J. P. de Koning, A. van Oudenaarden, A Single-Cell Transcriptome Atlas of the Human Pancreas. *Cell Syst*. **3**, 385–394.e3 (2016).

122. Å. Segerstolpe, A. Palasantza, P. Eliasson, E.-M. Andersson, A.-C. Andréasson, X. Sun, S. Picelli, A. Sabirsh, M. Clausen, M. K. Bjursell, D. M. Smith, M. Kasper, C. Ämmälä, R. Sandberg, Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes. *Cell Metab.* **24**, 593–607 (2016).

123. N. Lawlor, J. George, M. Bolisetty, R. Kursawe, L. Sun, V. Sivakamasundari, I. Kycia, P. Robson, M. L. Stitzel, Single-cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. *Genome Res.* **27**, 208–222 (2017).

124. M. Baron, A. Veres, S. L. Wolock, A. L. Faust, R. Gaujoux, A. Vetere, J. H. Ryu, B. K. Wagner, S. S. Shen-Orr, A. M. Klein, D. A. Melton, I. Yanai, A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Systems*. **3**, 346–360.e4 (2016).

125. Y. Xin, C. Adler, J. Kim, Y. Ding, M. Ni, Y. Wei, L. Macdonald, H. Okamoto, Single-cell RNA Sequencing and Analysis of Human Pancreatic Islets. *J. Vis. Exp.* (2019),

doi:10.3791/59866.

126. R. Kolde, *pheatmap: Pretty heatmaps* (Github; https://github.com/raivokolde/pheatmap).

127. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods*. **9**, 671–675 (2012).

128. D. R. Stirling, M. J. Swain-Bowden, A. M. Lucas, A. E. Carpenter, B. A. Cimini, A. Goodman, CellProfiler 4: improvements in speed, utility and usability. *BMC Bioinformatics*. **22**, 433 (2021).

129. M. V. Kuleshov, M. R. Jones, A. D. Rouillard, N. F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S. L. Jenkins, K. M. Jagodnik, A. Lachmann, M. G. McDermott, C. D. Monteiro, G. W. Gundersen, A. Ma'ayan, Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–7 (2016).

130. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

131. V. A. Schneider, T. Graves-Lindsay, K. Howe, N. Bouk, H.-C. Chen, P. A. Kitts, T. D. Murphy, K. D. Pruitt, F. Thibaud-Nissen, D. Albracht, R. S. Fulton, M. Kremitzki, V. Magrini, C. Markovic, S. McGrath, K. M. Steinberg, K. Auger, W. Chow, J. Collins, G. Harden, T. Hubbard, S. Pelan, J. T. Simpson, G. Threadgold, J. Torrance, J. M. Wood, L. Clarke, S. Koren, M. Boitano, P. Peluso, H. Li, C.-S. Chin, A. M. Phillippy, R. Durbin, R. K. Wilson, P. Flicek, E. E. Eichler, D. M. Church, Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res.* **27**, 849–864 (2017).

132. A. Buniello, J. A. L. MacArthur, M. Cerezo, L. W. Harris, J. Hayhurst, C. Malangone, A. McMahon, J. Morales, E. Mountjoy, E. Sollis, D. Suveges, O. Vrousgou, P. L. Whetzel, R. Amode, J. A. Guillen, H. S. Riat, S. J. Trevanion, P. Hall, H. Junkins, P. Flicek, T. Burdett, L. A. Hindorff, F. Cunningham, H. Parkinson, The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).

133. M. Arnold, J. Raffler, A. Pfeufer, K. Suhre, G. Kastenmüller, SNiPA: an interactive, genetic variant-centered annotation browser. *Bioinformatics*. **31**, 1334–1336 (2015).

134. S. T. Sherry, M. H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski, K. Sirotkin, dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).

135. I. Agnarsson, C. M. Zambrana-Torrelio, N. P. Flores-Saldana, L. J. May-Collado, A time-calibrated species-level phylogeny of bats (Chiroptera, Mammalia). *PLoS Curr.* **3**, RRN1212 (2011).

136. M. Laska, Food transit times and carbohydrate use in three phyllostomid bat species. *Zeitschrift für Säugetierkunde*. **55**, 49–54 (1990).

137. L. Montefiori, L. Hernandez, Z. Zhang, Y. Gilad, C. Ober, G. Crawford, M. Nobrega, N. Jo Sakabe, Reducing mitochondrial reads in ATAC-seq using CRISPR/Cas9. *Sci. Rep.* **7**, 2451 (2017).

138. S. Álvarez-Carretero, A. U. Tamuri, M. Battini, F. F. Nascimento, E. Carlisle, R. J. Asher, Z. Yang, P. C. J. Donoghue, M. Dos Reis, A species-level timeline of mammal evolution integrating phylogenomic data. *Nature*. **602**, 263–267 (2022).

139. R. A. Haeusler, K. Hartil, B. Vaitheesvaran, I. Arrieta-Cruz, C. M. Knight, J. R. Cook, H. L.

Kammoun, M. A. Febbraio, R. Gutierrez-Juarez, I. J. Kurland, D. Accili, Integrated control of hepatic lipogenesis versus glucose production requires FoxO transcription factors. *Nat. Commun.* **5**, 5190 (2014).

140. C. Q. Hoang, M. A. Hale, A. C. Azevedo-Pouly, H. P. Elsässer, T. G. Deering, S. G. Willet, F. C. Pan, M. A. Magnuson, C. V. E. Wright, G. H. Swift, R. J. MacDonald, Transcriptional Maintenance of Pancreatic Acinar Identity, Differentiation, and Homeostasis by PTF1A. *Mol. Cell. Biol.* **36**, 3033–3047 (2016).

141. M. A. Fenech, C. M. Sullivan, L. T. Ferreira, R. Mehmood, W. A. MacDonald, P. B. Stathopulos, C. L. Pin, Atp2c2 Is Transcribed From a Unique Transcriptional Start Site in Mouse Pancreatic Acinar Cells. *J. Cell. Physiol.* **231**, 2768–2778 (2016).

142. O. Franzén, L.-M. Gan, J. L. M. Björkegren, PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database* . **2019** (2019), doi:10.1093/database/baz046.

143. Y. Muto, P. C. Wilson, N. Ledru, H. Wu, H. Dimke, S. S. Waikar, B. D. Humphreys, Single cell transcriptional and chromatin accessibility profiling redefine cellular heterogeneity in the adult human kidney. *Nat. Commun.* **12**, 2190 (2021).

144. L. Chen, J. Z. Clark, J. W. Nelson, B. Kaissling, D. H. Ellison, M. A. Knepper, Renal-Tubule Epithelial Cell Nomenclature for Single-Cell RNA-Sequencing Studies. *J. Am. Soc. Nephrol.* **30**, 1358–1364 (2019).

145. E. P. Paksuz, Renal adaptation in relation to insectivorous feeding habit in the greater mouse-eared bat, Myotis myotis (Chiroptera: Vespertilionidae). *Anat. Rec.* (2022), doi:10.1002/ar.24946.

146. A. Roy, M. M. Al-bataineh, N. M. Pastor-Soler, Collecting duct intercalated cell function and regulation. *Clin. J. Am. Soc. Nephrol.* **10**, 305–324 (2015).

147. S. A. Lanham-New, The balance of bone health: tipping the scales in favor of potassium-rich, bicarbonate-rich foods. *J. Nutr.* **138**, 172S–177S (2008).

148. A. Staruschenko, Regulation of transport in the connecting tubule and cortical collecting duct. *Compr. Physiol.* **2**, 1541–1584 (2012).

149. P. C. Wilson, H. Wu, Y. Kirita, K. Uchimura, N. Ledru, H. G. Rennke, P. A. Welling, S. S. Waikar, B. D. Humphreys, The single-cell transcriptomic landscape of early human diabetic nephropathy. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 19619–19625 (2019).

150. N. Uehara-Watanabe, N. Okuno-Ozeki, A. Minamida, I. Nakamura, T. Nakata, K. Nakai, A. Yagi-Tomita, T. Ida, K. Ikeda, T. Kitani, N. Yamashita, M. Kamezaki, Y. Kirita, S. Matoba, K. Tamagaki, T. Kusaba, Direct evidence of proximal tubular proliferation in early diabetic nephropathy. *Sci. Rep.* **12**, 778 (2022).

151. T. Furusho, S. Uchida, E. Sohara, The WNK signaling pathway and salt-sensitive hypertension. *Hypertens. Res.* **43**, 733–743 (2020).

152. A. R. Rodan, A. Jenny, WNK Kinases in Development and Disease. *Curr. Top. Dev. Biol.* **123**, 1–47 (2017).

153. W. H. Yiu, D. W. L. Wong, L. Y. Y. Chan, J. C. K. Leung, K. W. Chan, H. Y. Lan, K. N. Lai, S. C. W. Tang, Tissue kallikrein mediates pro-inflammatory pathways and activation of protease-activated receptor-4 in proximal tubular epithelial cells. *PLoS One*. **9**, e88894 (2014).

154. P. Madeddu, C. Emanueli, S. El-Dahr, Mechanisms of disease: the tissue kallikrein-kinin

system in hypertension and vascular remodeling. *Nat. Clin. Pract. Nephrol.* **3**, 208–221 (2007).

155. G. Colla, H.-J. Kim, M. C. Kyriacou, Y. Rouphael, Nitrate in fruits and vegetables. *Sci. Hortic. .* **237**, 221–238 (2018).

156. S. El Moghrabi, P. Houillier, N. Picard, F. Sohet, B. Wootla, M. Bloch-Faure, F. Leviel, L. Cheval, S. Frische, P. Meneton, D. Eladari, R. Chambrey, Tissue kallikrein permits early renal adaptation to potassium load. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 13526–13531 (2010).

157. D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K. P. Tsafou, M. Kuhn, P. Bork, L. J. Jensen, C. von Mering, STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–52 (2015).

158. D. B. Mount, Thick ascending limb of the loop of Henle. *Clin. J. Am. Soc. Nephrol.* **9**, 1974–1986 (2014).

159. G. Schiano, B. Glaudemans, E. Olinger, N. Goelz, M. Müller, D. Loffing-Cueni, G. Deschenes, J. Loffing, O. Devuyst, The Urinary Excretion of Uromodulin is Regulated by the Potassium Channel ROMK. *Sci. Rep.* **9**, 19517 (2019).

160. E. H. Studier, D. E. Wilson, Natural urine concentrations and composition in neotropical bats. *Comp. Biochem. Physiol. A Physiol.* **75**, 509–515 (1983).

161. J. N. Lorenz, N. R. Baird, L. M. Judd, W. T. Noonan, A. Andringa, T. Doetschman, P. A. Manning, L. H. Liu, M. L. Miller, G. E. Shull, Impaired Renal NaCl Absorption in Mice Lacking the ROMK Potassium Channel, a Model for Type II Bartter's Syndrome*. *J. Biol.*

Chem. **277**, 37871–37880 (2002).

162. S.-H. Lin, I.-S. Yu, S.-T. Jiang, S.-W. Lin, P. Chu, A. Chen, H.-K. Sytwu, E. Sohara, S. Uchida, S. Sasaki, S.-S. Yang, Impaired phosphorylation of Na(+)-K(+)-2Cl(-) cotransporter by oxidative stress-responsive kinase-1 deficiency manifests hypotension and Bartter-like syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 17538–17543 (2011).

163. J. Xue, L. Thomas, J. D. Rieg, T. Rieg, Generation and characterization of thick ascending limb‑specific NHE3 knockout mice. *FASEB J.* **34**, 1–1 (2020).

164. D. B. Simon, F. E. Karet, J. M. Hamdan, A. DiPietro, S. A. Sanjad, R. P. Lifton, Bartter's syndrome, hypokalaemic alkalosis with hypercalciuria, is caused by mutations in the Na-K-2Cl cotransporter NKCC2. *Nat. Genet.* **13**, 183–188 (1996).

165. D. B. Simon, F. E. Karet, J. Rodriguez-Soriano, J. H. Hamdan, A. DiPietro, H. Trachtman, S. A. Sanjad, R. P. Lifton, Genetic heterogeneity of Bartter's syndrome revealed by mutations in the K+ channel, ROMK. *Nat. Genet.* **14**, 152–156 (1996).

166. N. Jeck, C. Derst, E. Wischmeyer, H. Ott, S. Weber, C. Rudin, H. W. Seyberth, J. Daut, A. Karschin, M. Konrad, Functional heterogeneity of ROMK mutations linked to hyperprostaglandin E syndrome. *Kidney Int.* **59**, 1803–1811 (2001).

167. N. Otani, M. Ouchi, K. Hayashi, P. Jutabha, N. Anzai, Roles of organic anion transporters (OATs) in renal proximal tubules and their localization. *Anat. Sci. Int.* **92**, 200–206 (2017).

168. R. T. Keenan, The biology of urate. *Semin. Arthritis Rheum.* **50**, S2–S10 (2020).

169. B. Jakše, B. Jakše, M. Pajek, J. Pajek, Uric Acid and Plant-Based Nutrition. *Nutrients*. **11** (2019), doi:10.3390/nu11081736.

170. Y. Chino, Y. Samukawa, S. Sakai, Y. Nakai, J.-I. Yamaguchi, T. Nakanishi, I. Tamai, SGLT2

inhibitor lowers serum uric acid through alteration of uric acid transport activity in renal tubule by increased glycosuria. *Biopharm. Drug Dispos.* **35**, 391–404 (2014).

171. I. D. Weiner, Roles of renal ammonia metabolism other than in acid-base homeostasis. *Pediatr. Nephrol.* **32**, 933–942 (2017).

172. N. Gotoh, Q. Yan, Z. Du, D. Biemesderfer, M. Kashgarian, M. S. Mooseker, T. Wang, Altered renal proximal tubular endocytosis and histology in mice lacking myosin-VI. *Cytoskeleton* . **67**, 178–192 (2010).

173. P. C. Wilson, Y. Muto, H. Wu, A. Karihaloo, S. S. Waikar, B. D. Humphreys, Multimodal single cell sequencing implicates chromatin accessibility and genetic background in diabetic kidney disease progression. *Nat. Commun.* **13**, 5253 (2022).

174. V. Vallon, Molecular determinants of renal glucose reabsorption. Focus on "Glucose transport by human renal Na+/D-glucose cotransporters SGLT1 and SGLT2." *Am. J. Physiol. Cell Physiol.* **300** (2011), pp. C6–8.

175. P. Machanick, T. L. Bailey, MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics*. **27**, 1696–1697 (2011).

176. Y. Wang, G. Jarad, P. Tripathi, M. Pan, J. Cunningham, D. R. Martin, H. Liapis, J. H. Miner, F. Chen, Activation of NFAT signaling in podocytes causes glomerulosclerosis. *J. Am. Soc. Nephrol.* **21**, 1657–1666 (2010).

177. Y. Maeoka, Y. Wu, T. Okamoto, S. Kanemoto, X. P. Guo, A. Saito, R. Asada, K. Matsuhisa, T. Masaki, K. Imaizumi, M. Kaneko, NFAT5 up-regulates expression of the kidney-specific ubiquitin ligase gene Rnf183 under hypertonic conditions in inner-medullary collecting duct cells. *J. Biol. Chem.* **294**, 101–115 (2019).

178. C. E. Pierreux, J. Stafford, D. Demonte, D. K. Scott, J. Vandenhaute, R. M. O'Brien, D. K. Granner, G. G. Rousseau, F. P. Lemaigre, Antiglucocorticoid activity of hepatocyte nuclear factor-6. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 8961–8966 (1999).

179. K. Yamamoto, T.-A. Matsuoka, S. Kawashima, S. Takebe, F. Kubo, T. Miyatsuka, H. Kaneto, I. Shimomura, A novel function of Onecut1 protein as a negative regulator of MafA gene expression. *J. Biol. Chem.* **288**, 21648–21658 (2013).

180. A. Philippi, S. Heller, I. G. Costa, V. Senée, M. Breunig, Z. Li, G. Kwon, R. Russell, A. Illing, Q. Lin, M. Hohwieler, A. Degavre, P. Zalloua, S. Liebau, M. Schuster, J. Krumm, X. Zhang, R. Geusz, J. R. Benthuysen, A. Wang, J. Chiou, K. Gaulton, H. Neubauer, E. Simon, T. Klein, M. Wagner, G. Nair, C. Besse, C. Dandine-Roulland, R. Olaso, J.-F. Deleuze, B. Kuster, M. Hebrok, T. Seufferlein, M. Sander, B. O. Boehm, F. Oswald, M. Nicolino, C. Julier, A. Kleger, Mutations and variants of ONECUT1 in diabetes. *Nat. Med.* **27**, 1928–1940 (2021).

181. S. Prudente, F. Andreozzi, L. Mercuri, F. Alberico, A. Di Giamberardino, G. C. Mannino, O. Ludovico, P. Piscitelli, R. Di Paola, S. Morano, G. Penno, M. Carella, S. De Cosmo, V. Trischitta, F. Barbetti, Contribution of ONECUT1 variants to different forms of non-autoimmune diabetes mellitus in Italian patients. *Acta Diabetol.* **59**, 1113–1116 (2022).

182. K. Ueda, K. Fujiki, K. Shirahige, C. E. Gomez-Sanchez, T. Fujita, M. Nangaku, M. Nagase, Genome-wide analysis of murine renal distal convoluted tubular cells for the target genes of mineralocorticoid receptor. *Biochem. Biophys. Res. Commun.* **445**, 132–137 (2014).

183. E. Gomez-Sanchez, C. E. Gomez-Sanchez, The multifaceted mineralocorticoid receptor. *Compr. Physiol.* **4**, 965–994 (2014).

184. M. Epstein, C. P. Kovesdy, C. M. Clase, M. M. Sood, R. Pecoits-Filho, Aldosterone,

Mineralocorticoid Receptor Activation, and CKD: A Review of Evolving Treatment Paradigms. *Am. J. Kidney Dis.* **80**, 658–666 (2022).

185. H. Joshi, B. Vastrad, N. Joshi, C. Vastrad, Integrated bioinformatics analysis reveals novel key biomarkers in diabetic nephropathy. *SAGE Open Med*. **10**, 20503121221137005 (2022).

186. J. A. Bonomo, M. Guan, M. C. Y. Ng, N. D. Palmer, P. J. Hicks, J. M. Keaton, J. P. Lea, C. D. Langefeld, B. I. Freedman, D. W. Bowden, The ras responsive transcription factor RREB1 is a novel candidate gene for type 2 diabetes associated end-stage kidney disease. *Hum. Mol. Genet.* **23**, 6441–6447 (2014).

187. F. Rothenberger, A. Velic, P. A. Stehberger, J. Kovacikova, C. A. Wagner, Angiotensin II stimulates vacuolar H+ -ATPase activity in renal acid-secretory intercalated cells from the outer medullary collecting duct. *J. Am. Soc. Nephrol.* **18**, 2085–2093 (2007).

188. Y. Wang, T. He, J. G. Herman, E. Linghu, Y. Yang, F. Fuks, F. Zhou, L. Song, M. Guo, Methylation of ZNF331 is an independent prognostic marker of colorectal cancer and promotes colorectal cancer growth. *Clin. Epigenetics*. **9**, 115 (2017).

189. X. Du, Y. Zhang, Q. Zhao, W. Qin, G. Ma, J. Fu, Q. Zhang, Effects of INSR genetic polymorphism on hippocampal volume and episodic memory in chinese type 2 diabetes. *Acta Diabetol.* **58**, 1471–1480 (2021).

190. Y. Liu, S. Ran, Y. Lin, Y.-X. Zhang, X.-L. Yang, X.-T. Wei, Z.-X. Jiang, X. He, H. Zhang, G.-J. Feng, H. Shen, Q. Tian, H.-W. Deng, L. Zhang, Y.-F. Pei, Four pleiotropic loci associated with fat mass and lean mass. *Int. J. Obes.* . **44**, 2113–2123 (2020).

191. L. O. Huang, A. Rauch, E. Mazzaferro, M. Preuss, S. Carobbio, C. S. Bayrak, N. Chami, Z.

Wang, U. M. Schick, N. Yang, Y. Itan, A. Vidal-Puig, M. den Hoed, S. Mandrup, T. O. Kilpeläinen, R. J. F. Loos, Genome-wide discovery of genetic loci that uncouple excess adiposity from its comorbidities. *Nat Metab*. **3**, 228–243 (2021).

192. S. Torkamandi, M. Bastami, H. Ghaedi, F. Moghadam, R. Mirfakhraie, M. D. Omrani, MAP3K1 May be a Promising Susceptibility Gene for Type 2 Diabetes Mellitus in an Iranian Population. *Int J Mol Cell Med*. **5**, 134–140 (2016).

193. A. Al Shamsi, N. Al Hassani, M. Hamchou, R. Almazrouei, A. Mhanni, A novel missense heterozygous mutation in MAP3K1 gene causes 46, XY disorder of sex development: case report and literature review. *Mol. Genet. Genomic Med.* **8**, e1514 (2020).

194. V. Todorovová, J. Hubacek, L. Dlouha, V. Adamkova, J. Pitha, M. Satny, R. Ceska, M. Vrablik, APOA5, GCKR, LRP1 AND MAP3K1 polymorphisms and the risk of acute coronary syndrome in the Czech population. *Atherosclerosis*. **331**, e215 (2021).

195. A. Wesolowska-Andersen, G. Zhuo Yu, V. Nylander, F. Abaitua, M. Thurner, J. M. Torres, A. Mahajan, A. L. Gloyn, M. I. McCarthy, Deep learning models predict regulatory variants in pancreatic islets and refine type 2 diabetes association signals. *Elife*. **9** (2020), doi:10.7554/eLife.51503.

196. N. L. Harvey, R. S. Srinivasan, M. E. Dillard, N. C. Johnson, M. H. Witte, K. Boyd, M. W. Sleeman, G. Oliver, Lymphatic vascular defects promoted by Prox1 haploinsufficiency cause adult-onset obesity. *Nat. Genet.* **37**, 1072–1081 (2005).

197. C. Schnoz, S. Moser, D. V. Kratschmar, A. Odermatt, D. Loffing-Cueni, J. Loffing, Deletion of the transcription factor Prox-1 specifically in the renal distal convoluted tubule causes hypomagnesemia via reduced expression of TRPM6 and NCC. *Pflugers Arch.* **473**, 79–93 (2021).

198. L. M. M. Gommers, J. G. J. Hoenderop, R. J. M. Bindels, J. H. F. de Baaij, Hypomagnesemia in Type 2 Diabetes: A Vicious Circle? *Diabetes*. **65**, 3–13 (2016).

199. L. J. Oost, J. I. P. van Heck, C. J. Tack, J. H. F. de Baaij, The association between hypomagnesemia and poor glycaemic control in type 1 diabetes is limited to insulin resistant individuals. *Sci. Rep.* **12**, 1–7 (2022).

200. L. Kouřimská, A. Adámková, Nutritional and sensory quality of edible insects. *NFS Journal*. **4**, 22–26 (2016).

201. M. Perez-Frances, L. van Gurp, M. V. Abate, V. Cigliola, K. Furuyama, E. Bru-Tari, D. Oropeza, T. Carreaux, Y. Fujitani, F. Thorel, P. L. Herrera, Pancreatic Ppy-expressing γ-cells display mixed phenotypic traits and the adaptive plasticity to engage insulin production. *Nat. Commun.* **12**, 4458 (2021).

202. M. M. Cooley, E. K. Jones, F. S. Gorelick, G. E. Groblewski, Pancreatic acinar cell protein synthesis, intracellular transport, and export. *Pancreapedia: Exocrine Pancreas Knowledge Base* (2020), , doi:10.3998/panc.2020.15.

203. A. Grapin-Botton, Ductal cells of the pancreas. *Int. J. Biochem. Cell Biol.* **37**, 504–510 (2005).

204. J. Maléth, P. Hegyi, Calcium signaling in pancreatic ductal epithelial cells: an old friend and a nasty enemy. *Cell Calcium*. **55**, 337–345 (2014).

205. M. Priyadarshini, B. T. Layden, FFAR3 modulates insulin secretion and global gene expression in mouse islets. *Islets*. **7**, e1045182 (2015).

206. C. Tang, K. Ahmed, A. Gille, S. Lu, H.-J. Gröne, S. Tunaru, S. Offermanns, Loss of FFA2 and FFA3 increases insulin secretion and improves glucose tolerance in type 2 diabetes.

Nat. Med. **21**, 173–177 (2015).

207. C. Berger, D. Zdzieblo, Glucose transporters in pancreatic islets. *Pflugers Arch.* **472**, 1249–1272 (2020).

208. J. J. DiNicolantonio, J. H O'Keefe, Myo-inositol for insulin resistance, metabolic syndrome, polycystic ovary syndrome and gestational diabetes. *Open Heart*. **9** (2022), doi:10.1136/openhrt-2022-001989.

209. M. J. Coady, B. Wallendorff, D. G. Gagnon, J.-Y. Lapointe, Identification of a novel Na+/myo-inositol cotransporter. *J. Biol. Chem.* **277**, 35219–35224 (2002).

210. M. A. Atkinson, M. Campbell-Thompson, I. Kusmartseva, K. H. Kaestner, Organisation of the human pancreas in health and in diabetes. *Diabetologia*. **63**, 1966–1973 (2020).

211. Y. Xin, J. Kim, H. Okamoto, M. Ni, Y. Wei, C. Adler, A. J. Murphy, G. D. Yancopoulos, C. Lin, J. Gromada, RNA Sequencing of Single Human Islet Cells Reveals Type 2 Diabetes Genes. *Cell Metab.* **24**, 608–615 (2016).

212. E. Bosi, L. Marselli, C. De Luca, M. Suleiman, M. Tesi, M. Ibberson, D. L. Eizirik, M. Cnop, P. Marchetti, Integration of single-cell datasets reveals novel transcriptomic signatures of β-cells in human type 2 diabetes. *NAR Genom Bioinform*. **2**, lqaa097 (2020).

213. Y.-T. Chen, W.-D. Lin, W.-L. Liao, Y.-J. Lin, J.-G. Chang, F.-J. Tsai, PTPRD silencing by DNA hypermethylation decreases insulin receptor signaling and leads to type 2 diabetes. *Oncotarget*. **6**, 12997–13005 (2015).

214. Y. Kang, H. Huang, H. Li, W. Sun, C. Zhang, Functional genetic variants in the 3'UTR of PTPRD associated with the risk of gestational diabetes mellitus. *Exp. Ther. Med.* **21**, 562 (2021).

215. L. Stoll, J. Sobel, A. Rodriguez-Trejo, C. Guay, K. Lee, M. T. Venø, J. Kjems, D. R. Laybutt, R. Regazzi, Circular RNAs as novel regulators of β-cell functions in normal and disease conditions. *Mol Metab*. **9**, 69–83 (2018).

216. S. Elumalai, U. Karunakaran, J.-S. Moon, K.-C. Won, NADPH Oxidase (NOX) Targeting in Diabetes: A Special Emphasis on Pancreatic β-Cell Dysfunction. *Cells*. **10** (2021), doi:10.3390/cells10071573.

217. T. Koufakis, A. Sertedaki, E.-B. Tatsi, C.-M. Trakatelli, S. N. Karras, E. Manthou, C. Kanaka-Gantenbein, K. Kotsa, First Report of Diabetes Phenotype due to a Loss-of-Function *ABCC8* Mutation Previously Known to Cause Congenital Hyperinsulinism. *Case Rep. Genet.* **2019**, 3654618 (2019).

218. E. Araki, M. A. Lipes, M. E. Patti, J. C. Brüning, B. Haag 3rd, R. S. Johnson, C. R. Kahn, Alternative pathway of insulin signalling in mice with targeted disruption of the IRS-1 gene. *Nature*. **372**, 186–190 (1994).

219. M. G. De Vas, J. L. Kopp, C. Heliot, M. Sander, S. Cereghini, C. Haumaitre, Hnf1b controls pancreas morphogenesis and the generation of Ngn3+ endocrine progenitors. *Development*. **142**, 871–882 (2015).

220. C.-K. Kim, P. He, A. B. Bialkowska, V. W. Yang, SP and KLF Transcription Factors in Digestive Physiology and Diseases. *Gastroenterology*. **152**, 1845–1875 (2017).

221. H. Kaneto, T.-A. Matsuoka, Role of pancreatic transcription factors in maintenance of mature β-cell function. *Int. J. Mol. Sci.* **16**, 6281–6297 (2015).

222. H. J. Li, A. Kapoor, M. Giel-Moloney, G. Rindi, A. B. Leiter, Notch signaling differentially regulates the cell fate of early endocrine precursor cells and their maturing descendants in

the mouse pancreas and intestine. *Dev. Biol.* **371**, 156–169 (2012).

223. S. Spohrer, R. Groß, L. Nalbach, L. Schwind, H. Stumpf, M. D. Menger, E. Ampofo, M. Montenarh, C. Götz, Functional interplay between the transcription factors USF1 and PDX-1 and protein kinase CK2 in pancreatic β-cells. *Sci. Rep.* **7**, 16367 (2017).

224. A. P. Sanchez, J. Zhao, Y. You, A.-E. Declèves, M. Diamond-Stanic, K. Sharma, Role of the USF1 transcription factor in diabetic kidney disease. *Am. J. Physiol. Renal Physiol.* **301**, F271–9 (2011).

225. J. E. Gunton, R. N. Kulkarni, S. Yim, T. Okada, W. J. Hawthorne, Y.-H. Tseng, R. S. Roberson, C. Ricordi, P. J. O'Connell, F. J. Gonzalez, C. R. Kahn, Loss of ARNT/HIF1beta mediates altered gene expression and pancreatic-islet dysfunction in human type 2 diabetes. *Cell*. **122**, 337–349 (2005).

226. H. S. Kang, K. Okamoto, Y.-S. Kim, Y. Takeda, C. D. Bortner, H. Dang, T. Wada, W. Xie, X.-P. Yang, G. Liao, A. M. Jetten, Nuclear orphan receptor TAK1/TR4-deficient mice are protected against obesity-linked inflammation, hepatic steatosis, and insulin resistance. *Diabetes*. **60**, 177–188 (2011).

227. Y.-F. Lee, S. Liu, N.-C. Liu, R.-S. Wang, L.-M. Chen, W.-J. Lin, H.-J. Ting, H.-C. Ho, G. Li, E. J. Puzas, Q. Wu, C. Chang, Premature aging with impaired oxidative stress defense in mice lacking TR4. *Am. J. Physiol. Endocrinol. Metab.* **301**, E91–8 (2011).

228. S. Wang, J. Skorczewski, X. Feng, L. Mei, J. E. Murphy-Ullrich, Glucose Up-regulates Thrombospondin 1 Gene Transcription and Transforming Growth Factor-β Activity through Antagonism of cGMP-dependent Protein Kinase Repression via Upstream Stimulatory Factor 2*. *J. Biol. Chem.* **279**, 34311–34322 (2004).

229. L. Shi, S. Liu, D. Nikolic, S. Wang, High glucose levels upregulate upstream stimulatory factor 2 gene transcription in mesangial cells. *J. Cell. Biochem.* **103**, 1952–1961 (2008).

230. R. K. Lex, W. Zhou, Z. Ji, K. N. Falkenstein, K. E. Schuler, K. E. Windsor, J. D. Kim, H. Ji, S. A. Vokes, GLI transcriptional repression is inert prior to Hedgehog pathway activation. *Nat. Commun.* **13**, 808 (2022).

231. R. J. Weiss, P. N. Spahn, A. G. Toledo, A. W. T. Chiang, B. P. Kellman, J. Li, C. Benner, C. K. Glass, P. L. S. M. Gordts, N. E. Lewis, J. D. Esko, ZNF263 is a transcriptional regulator of heparin and heparan sulfate biosynthesis. *Proceedings of the National Academy of Sciences*. **117**, 9311–9317 (2020).

232. C. Su, L. Gao, C. L. May, J. A. Pippin, K. Boehm, M. Lee, C. Liu, M. C. Pahl, M. L. Golson, A. Naji, the HPAP Consortium, S. F. A. Grant, A. D. Wells, K. H. Kaestner, The three-dimensional chromatin structure of the major human pancreatic cell types reveals lineage-specific regulatory architecture of T2D risk. *bioRxiv* (2022), p. 2021.11.30.470653.

233. M. Heddad Masson, C. Poisson, A. Guérardel, A. Mamin, J. Philippe, Y. Gosmain, Foxa1 and Foxa2 regulate α-cell differentiation, glucagon biosynthesis, and secretion. *Endocrinology*. **155**, 3781–3792 (2014).

234. S. L. Park, I. Cheng, S. A. Pendergrass, A. M. Kucharska-Newton, U. Lim, J. L. Ambite, C. P. Caberto, K. R. Monroe, F. Schumacher, L. A. Hindorff, M. T. Oetjens, S. Wilson, R. J. Goodloe, S.-A. Love, B. E. Henderson, L. N. Kolonel, C. A. Haiman, D. C. Crawford, K. E. North, G. Heiss, M. D. Ritchie, L. R. Wilkens, L. Le Marchand, Association of the FTO obesity risk variant rs8050136 with percentage of energy intake from fat in multiple racial/ethnic populations: the PAGE study. *Am. J. Epidemiol.* **178**, 780–790 (2013).

235. T. Bego, A. Čaušević, T. Dujić, M. Malenica, Z. Velija-Asimi, B. Prnjavorac, J. Marc, J.

Nekvindová, V. Palička, S. Semiz, Association of FTO Gene Variant (rs8050136) with Type 2 Diabetes and Markers of Obesity, Glycaemic Control and Inflammation. *J. Med. Biochem.* **38**, 153–163 (2019).

236. H. Jia, L. Yu, Z. Jiang, Q. Ji, Association between IGF2BP2 rs4402960 polymorphism and risk of type 2 diabetes mellitus: a meta-analysis. *Arch. Med. Res.* **42**, 361–367 (2011).

237. D. El-Lebedy, I. Ashmawy, A. A. Ibrahim, Common Variants in IGF2BP2 Gene rs4402960 and rs1470579 Polymorphisms Associate with Type 2 Diabetes Mellitus in Egyptians: A Replication Study. *International Journal of Diabetes Research*. **4**, 43–48 (2015).

238. M.-Q. Mo, L. Pan, Q.-M. Lu, Q.-L. Li, Y.-H. Liao, The association of the CMIP rs16955379 polymorphism with dyslipidemia and the clinicopathological features of IgA nephropathy. *Int. J. Clin. Exp. Pathol.* **11**, 5008–5023 (2018).

239. L. Pan, Y.-H. Liao, M.-Q. Mo, Q.-H. Zhang, R.-X. Yin, CMIP SNPs and their haplotypes are associated with dyslipidaemia and clinicopathologic features of IgA nephropathy. *Biosci. Rep.* **40** (2020), doi:10.1042/BSR20202628.

240. Z. B. Mehta, N. Fine, T. J. Pullen, M. C. Cane, M. Hu, P. Chabosseau, G. Meur, A. Velayos-Baeza, A. P. Monaco, L. Marselli, P. Marchetti, G. A. Rutter, Changes in the expression of the type 2 diabetes-associated gene VPS13C in the β-cell are associated with glucose intolerance in humans and mice. *Am. J. Physiol. Endocrinol. Metab.* **311**, E488–507 (2016).

241. T. Kuo, M. J. Kraakman, M. Damle, R. Gill, M. A. Lazar, D. Accili, Identification of C2CD4A as a human diabetes susceptibility gene with a role in β cell insulin secretion. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 20033–20042 (2019).

242. R. D. Albanus, X. Tang, H. J. Taylor, N. Manickam, M. Erdos, N. Narisu, Y. Han, P. Orchard, A. Varshney, C. Liu, A. Naji, HPAP Consortium, F. S. Collins, S. Chen, S. C. J. Parker, Single-cell gene expression and chromatin accessibility profiling of human pancreatic islets at basal and stimulatory conditions nominates mechanisms of type 1 diabetes genetic risk. *bioRxiv* (2022), p. 2022.11.12.516291.

243. F. Wei, C. Cai, S. Feng, J. Lv, S. Li, B. Chang, H. Zhang, W. Shi, H. Han, C. Ling, P. Yu, Y. Chen, N. Sun, J. Tian, H. Jiao, F. Yang, M. Li, Y. Wang, L. Zou, L. Su, J. Li, R. Li, H. Qiu, J. Shi, S. Liu, M. Chang, J. Lin, L. Chen, W.-D. Li, TOX and CDKN2A/B Gene Polymorphisms Are Associated with Type 2 Diabetes in Han Chinese. *Sci. Rep.* **5**, 11900 (2015).

244. T. Zhang, H. H. Dong, Glucose-regulated insulin production in the liver improves glycemic control in type 1 diabetic mice. *Mol Metab*. **4**, 70–76 (2015).

245. S. Sefried, H.-U. Häring, C. Weigert, S. S. Eckstein, Suitability of hepatocyte cell lines HepG2, AML12 and THLE-2 for investigation of insulin signalling and hepatokine gene expression. *Open Biol.* **8** (2018), doi:10.1098/rsob.180147.

246. G. Casotti, K. C. Richardson, J. S. Bradley, Ecomorphological constraints imposed by the kidney component measurements in honeyeater birds inhabiting different environments. *J. Zool.* . **231**, 611–625 (1993).

247. G. Casotti, K. C. Richardson, A stereological analysis of kidney structure of honeyeater birds (Meliphagidae) inhabiting either arid or wet environments. *J. Anat.* **180 (Pt 2)**, 281–288 (1992).

248. C. A. Beuchat, M. R. Preest, E. J. Braun, Glomerular and medullary architecture in the kidney of Anna's Hummingbird. *J. Morphol.* **240**, 95–100 (1999).

249. I. Bravo-Ruiz, M. Á. Medina, B. Martínez-Poveda, From Food to Genes: Transcriptional Regulation of Metabolism by Lipids and Carbohydrates. *Nutrients*. **13** (2021), doi:10.3390/nu13051513.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

_____   3/7/2023_____
Author Signature   Date