Genomic and Epigenomic Characterizations of Major Pancreatic Islet Endocrine Cell Types

By

ALEX M. MAWLA
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Integrative Genetics & Genomics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Dr. Mark O. Huising, Chair

Dr. Ian Korf

Dr. Alex S. Nord

Dr. Gerald Quon

Committee in Charge

2021

Dedication

To my late father, who unfortunately could not stay in this world long enough to see his son become a doctor. You inspired me to end up in this field, so that I can make a difference for others. Dad – I know you would have been very proud of me. To my mom, who is a ball of light and love that shines brighter than any sun. Thank you both for all that you have done for me.

To Dr. Mark Huising and Dr. Talitha van der Meulen - your mentorship, and friendship, goes beyond what words can truly convey; I am entirely indebted for life.

To the Dargous and the Zubair family, thank you for your friendship, love, and support over the years.

To Kal Zubair, my best and closest friend, since we were six. Kal and I learned many lessons in our younger years, through building PVC potato cannons and other "scientific" minded endeavors we pursued. I have no siblings, but in my eyes, Kal is my brother.

To the Obeid family - my adopted grandparents, thank you for everything.

Acknowledgements

This work could not have been done without the support of my colleagues, family, friends, and mentors.

I want to first and foremost thank my advisor and friend, Dr. Mark O. Huising, for all his mentorship, support, patience, and advocacy from day one throughout these years. I was very lucky to be connected to Mark through a colleague, Christopher Cowing-Zitron. Mark provided me immeasurable opportunities, supported me through my graduate student journey, and was always a fantastic advocate.

I want to thank Dr. Talitha van der Meulen, who was an amazing professor and advocate for me, and who was an incredible mentor to learn from when working on research together. Separately, Talitha recruited me as a teaching assistant for what became 3 years of my graduate career; and this opportunity and experience made me fall in love with teaching and working with and supporting students.

I want to thank Dr. Ian F. Korf – I took MCB 182 as an undergraduate in Neurobiology, Physiology, and Behavior, and was lucky enough to have him as the professor, and later to have the opportunity to intern under his guidance. If it were not for Ian, I would not have applied to this program. My first experience as a teaching assistant was also under Ian, in winter quarter of 2017, for the same class (MCB 182) that I took as an undergraduate. I did not realize this then, but this laid the foundation for my love of teaching – something I never had considered before. This remains one of my favorite classes, and Ian's invention of the genomic sudoku is an extremely cool concept that anyone who likes genetics will probably enjoy.

I want to thank Dr. Tyler A. Lesh. As an undergraduate, I had the opportunity to work under his guidance in schizophrenia and bipolar research at the University of California, Davis Medical Center. Tyler was a really fantastic mentor and interning with him provided me the opportunity to present my first research project, at the University of California, Davis Undergraduate Research Conference in 2014. This experience showed me how fulfilling research can be, and inspired me to pursue a graduate degree. Through Tyler, I was connected to Dr. Alex S Nord, and this initiated my transition from my undergraduate focus to my current field of expertise.

I want to thank Dr. Alex S. Nord. When I mentioned I wanted to pursue a PhD in genomics and was interested in the field, Dr. Tyler A. Lesh pointed me in Alex's direction. I was very lucky to be able to connect with Alex, and later join his lab as an intern. Through my time there, I learned a lot about computational biology, and his mentorship and support aided me in pursuing graduate school. I also wrote my very first genomics pipeline in Alex's lab under his guidance; many more came after. P.S., thank you Alex for graciously sponsoring me to use the genome center and, later on, crick servers.

I want to thank Dr. Gerald Quon. You gave me an opportunity to rotate through your lab, and the Dream Challenge I did under you, and Ian's, guidance, is one of my best and most fun memories of graduate school. I learned a lot during this period. I also really enjoyed being a teaching assistant under Gerald and getting to attend a CURE seminar alongside him back in the summer 2019 when Dr. van der Meulen and I wanted to create a course to teach computational genomics and bench-lab biology to students while Gerald was in parallel perusing a CURE course he still teaches as well. The collegiate environment I have been lucky to be around has been nothing short of supportive and amazing.

I want to thank my colleagues and fellow graduate students Michelle Chan, Ryan Hart, Jessica L. Huang, Sharon Lee, Glyn M. Noguchi, Mohammad Pourhosseinzadeh, Max Shin, and Jing Zhang. It was a pleasure to work alongside you.

I want to thank some amazing undergraduate volunteers I met during the course of my graduate studies (two of which became junior specialists for our lab). I had the pleasure to first meet Andrew Kimelman when he was a student in the course I was a teaching assistant for. I later was able to recruit him to our lab, to which he was an amazing addition. I also had the pleasure of working with Sophia Chen, who was a student in the same course as Andrew, becoming a learning assistant in following quarters – I got to witness her teaching discussions, which came naturally to her, with minimal support from me. She is now also a fantastic member of our lab. I also had the great pleasure of knowing Marcus Flisher and Jareseley Guillen. Both joined our lab early on as undergraduates, and both ended up holding a junior specialist position in our lab and contributing immensely. I got to know Marcus very well in particular, as he expressed great interest in computational genomics, during this period. I had the

spirit animal into computational biology and genomics, and he has since used this knowledge to drive several projects of his own that have contributed to our lab greatly.

I want to thank all the colleagues, faculty members, and undergraduates I've had the opportunity and good fortune to know and work with.

Lastly, I would like to thank my family, and my friends. No accomplishment exists in a bubble. I would never be where I am today without your genuine love, friendship, and warm hearts. I am the richest person in the world, because I have you in my life. I will never be able to repay you what all have you have given me, but I will spend my life trying. Thank you.

To all those mentioned directly, or indirectly, in this, please know I am eternally grateful to everything you have done for me. Thank you, sincerely.

Abstract

During pancreatic development, insulin-producing beta cells, glucagon-producing alpha cells, and somatostatin-producing delta cells differentiate from a common endocrine progenitor. While these cells have been considered terminally differentiated, over the last several years it has become evident that pancreatic islet cellular identity is more plastic than previously appreciated. A cell's transcriptome is the final outcome that stems from the interactions of both genomic and epigenomic components. Dynamic changes in chromatin accessibility alongside the recruitment of histone marks that either facilitate or repress gene transcription in tandem with transcription factor co-localization come together to define and maintain cellular identity through gene expression. Understanding of each of these components sheds further light on understanding cellular identity, with profiling the transcriptome providing signature patterns that are responsible for the functional behavior of these cells. Building on this through the interrogation of chromatin accessibility similarities and differences provides further insight into the mechanisms that govern cellular identity when coupled with the appropriate histone modification and transcription factor data. Ultimately, genomic and epigenomic characterizations of these major islet endocrine cell types can provide an avenue to further understanding the complex regulatory mechanisms that define and maintain cellular identity. Furthermore, these characterizations provide a resource for answering outstanding questions in the field.

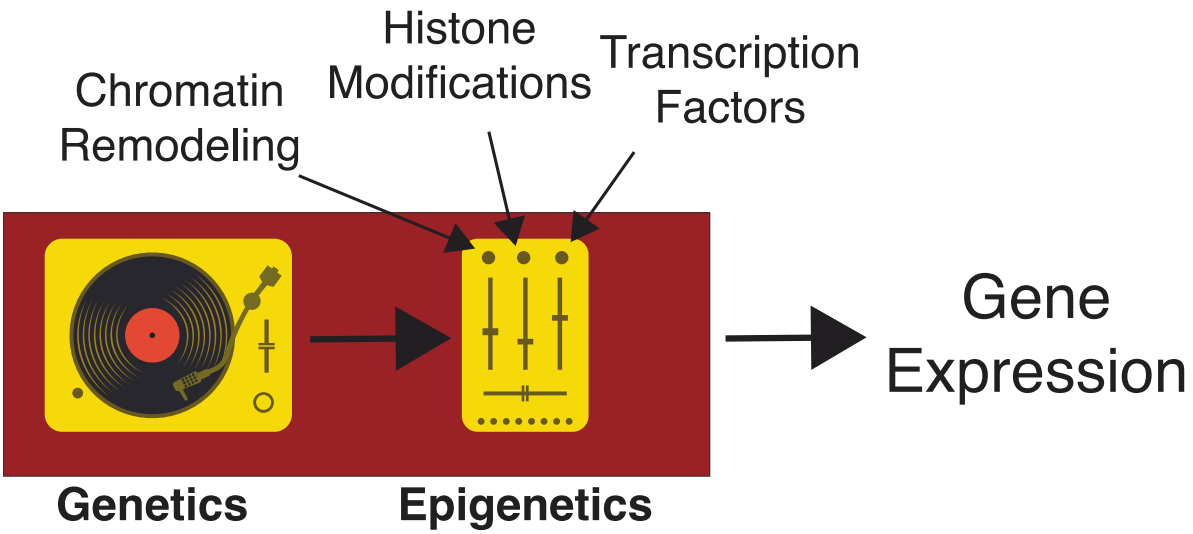


Table of Contents

I. Chapter 1: Introduction

The study of pancreatic islet biology can ultimately contribute to innovative treatment for Type 1 (T1D) and Type 2 (T2D) Diabetes, a disease that impacts roughly 1 in every 7 US adults [1]. Both diseases are driven by beta cell loss or deficiency, with autoimmune components driving T1D, and environment-gene interactions that damage or de-differentiate beta cells in T2D [2, 3]. Waddington's concept of the developmental epigenetic landscape proposes that dynamic gene regulation modulates progressive differentiation along a cell lineage with increasing irreversibility towards a stable cell fate [4]. However, as the breakthrough achievements of Yamanaka demonstrated, even stably differentiated cells retain the ability to adopt a different cell fate [5]. Although the field of pancreatic islet biology has made considerable progress in untangling the successive waves of transcription factors (TF) that guide beta cell identity during development, there is still an unexplored area of epigenetic and regulatory changes associated with beta cell maturation and maintenance [6]. While a handful of studies have examined different angles to beta cell maturation and regeneration—progenitor stem cell source [7], TF co-regulation in maintaining identity [8], and cell fate reprogramming [9]—no clinically established regeneration of human beta cell mass via proliferation has been achieved.

During pancreatic development, insulin-producing beta cells, glucagon producing alpha cells, and somatostatin-producing delta cells differentiate from a common endocrine progenitor. While these cells were previously deemed terminally differentiated, it is now evident that cell fate is more plastic than previously anticipated. Cell fate switching and maintenance between pancreatic islet cells has been established within both healthy and Type 2 Diabetic (T2D) mouse and human species, across development. While initially beta cell regeneration and maintenance was believed to occur primarily through self-reflection [10], alpha cells were observe to spontaneously transdifferentiate, switching fate to beta cells after beta cell ablation [11]. Further research in the field then identified that beta cell ablation was not required for this to occur, and that ~0.2 – 1.5% of beta cells – termed 'virgin beta cells' at any given time stem from alpha cells [12]. To a lesser extent, delta cell transdifferentiation has been identified upon beta

cell ablation, as well as a source of beta cells in juveniles [13]. Understanding the transcriptional nature and identities of these major endocrine cell types, along with that of the more rare and transient populations of virgin or transdifferentiated beta cells, can help provide an understanding of the functional differences and similarities. The ability to explore (some of) these cell types via their epigenomic profiles can help further this understanding.

The dynamic ability for chromatin accessibility to remodel between euchromatin and heterochromatin, along with nucleosome occupancy changes, states plays a vital role in epigenetic regulation behind cell fate switching and maintenance [14-18], and in in tissue- and cell-specific gene expression and regulatory activity [14, 19, 20]. Chromatin stability is required in order to effectively maintain healthy islet cell identity, with changes underlying dysregulation leading to disease [21-23]. Furthermore, chromatin accessibility is a requirement for other epigenetic factors to interact, such as transcription factor recruitment at enhancers to facilitate either the activation or repression of near and distal genes [24-26]. Previous studies have explored pancreatic islet cellular identity by evaluating epigenomic features such as methylation [27-29], histone modifications [9, 30-32], non-coding RNA [33-36], and enhancer regulatory regions [17, 23, 37-39]. While each of these factors contributes to defining and maintaining cell fate and identity, connecting chromatin accessibility differences to some of these epigenetic factors promises to provide further insight into outstanding questions within the field.

To better understand and characterize endocrine islet cell identity, we applied complementary chromatin accessibility and transcriptomes approach of FACS-purified mouse alpha, beta, and delta cells. In order to achieve this, we turned to triple-transgenic reporter mice - mIns1-H2b-mCherry beta cells crossed to mice with alpha or delta cells marked by YFP in a Cre-dependent fashion - previously used to evaluate these cells' transcriptional landscape, allowing us to directly link these two datasets [40, 41]. Through this, we provide a comprehensive characterization of the transcriptomic landscape of alpha, beta, and delta cells, alongside the similarities and differences that characterize immature 'virgin' beta cells. This theme is continued in the transcriptomic evaluation of alpha- or delta- origin transdifferentiated beta cells,

with ‘virgin’ beta cells believed to represent a transient intermediate in this shift. Finally, these transcriptomic findings are tied to epigenomic characterizations of alpha, beta, and delta cells evaluating chromatin accessibility via ATAC-Seq and linking chromatin availability with aggregated transcription factor and histone mark data that promises to elucidate the regulatory networks the drive and maintain cell identity.

To this end, the following research describes (1) the comprehensive transcriptomic evaluation of alpha, beta, and delta cells within pancreatic islets [40] (**Chapter 2**); (2) with a companion, easy to use R package developed to visualize calcium-signaling data, initially devised for **Chapter 2** but later utilized in several publications (**Chapter 3**); (3) the novel identification and characterization of the rare population of virgin and transdifferentiated beta cells through bulk RNA sequencing [12] (**Chapter 4**); (4) followed with an analysis on reproducibility in science, directly refuting a manuscript which claimed an FDA-approved drug could trigger alpha to beta cell transdifferentiation [42] (**Chapter 5**); (5) a meta-analysis and review of single-cell sequencing, done initially with the interest of isolating virgin beta cells in both mouse and human populations without the need for transgenic mouse lines or antibodies [43] (**Chapter 6**); and finally, (6) a characterization of the epigenomic landscape to the transcriptomic analysis of alpha, beta, and delta cells (**Chapter 7**), utilizing a novel R package designed to integrate multiple layers of -omics, and identity putative enhancer regions [44] (**Chapter 8**).

References

1. Menke A, Casagrande S, Geiss L, Cowie CC: **Prevalence of and Trends in Diabetes Among Adults in the United States, 1988-2012**. *JAMA* 2015, **314**(10):1021-1029.
2. Halban PA, Polonsky KS, Bowden DW, Hawkins MA, Ling C, Mather KJ, Powers AC, Rhodes CJ, Sussel L, Weir GC: **beta-cell failure in type 2 diabetes: postulated mechanisms and prospects for prevention and treatment**. *Diabetes care* 2014, **37**(6):1751-1758.
3. Rui J, Deng S, Arazi A, Perdigoto AL, Liu Z, Herold KC: **beta Cells that Resist Immunological Attack Develop during Progression of Autoimmune Diabetes in NOD Mice**. *Cell Metab* 2017, **25**(3):727-738.
4. Noble D: **Conrad Waddington and the origin of epigenetics**. *The Journal of experimental biology* 2015, **218**(Pt 6):816-818.
5. Takahashi K, Yamanaka S: **A decade of transcription factor-mediated reprogramming to pluripotency**. *Nat Rev Mol Cell Biol* 2016, **17**(3):183-193.
6. Bruin JE, Rezania A, Kieffer TJ: **Replacing and safeguarding pancreatic beta cells for diabetes**. *Sci Transl Med* 2015, **7**(316):316ps323.
7. Russ HA, Parent AV, Ringler JJ, Hennings TG, Nair GG, Shveygert M, Guo T, Puri S, Haataja L, Cirulli V *et al*: **Controlled induction of human pancreatic progenitors produces functional beta-like cells in vitro**. *EMBO J* 2015, **34**(13):1759-1772.
8. Jia S, Ivanov A, Blasevic D, Muller T, Purfurst B, Sun W, Chen W, Poy MN, Rajewsky N, Birchmeier C: **Insm1 cooperates with Neurod1 and Foxa2 to maintain mature pancreatic beta-cell function**. *EMBO J* 2015, **34**(10):1417-1433.
9. Bramswig NC, Everett LJ, Schug J, Dorrell C, Liu C, Luo Y, Streeter PR, Naji A, Grompe M, Kaestner KH: **Epigenomic plasticity enables human pancreatic alpha to beta cell reprogramming**. *J Clin Invest* 2013, **123**(3):1275-1284.
10. Dor Y, Brown J, Martinez OI, Melton DA: **Adult pancreatic beta-cells are formed by self-duplication rather than stem-cell differentiation**. *Nature* 2004, **429**(6987):41-46.
11. Thorel F, Nepote V, Avril I, Kohno K, Desgraz R, Chera S, Herrera PL: **Conversion of adult pancreatic alpha-cells to beta-cells after extreme beta-cell loss**. *Nature* 2010, **464**(7292):1149-1154.
12. van der Meulen T, Mawla AM, DiGruccio MR, Adams MW, Nies V, Dolleman S, Liu S, Ackermann AM, Caceres E, Hunter AE *et al*: **Virgin Beta Cells Persist throughout Life at a Neogenic Niche within Pancreatic Islets**. *Cell Metab* 2017, **25**(4):911-926 e916.
13. Chera S, Baronnier D, Ghila L, Cigliola V, Jensen JN, Gu G, Furuyama K, Thorel F, Gribble FM, Reimann F *et al*: **Diabetes recovery by age-dependent conversion of pancreatic delta-cells into insulin producers**. *Nature* 2014, **514**(7523):503-507.
14. Greenwald WW, Li H, Benaglio P, Jakubosky D, Matsui H, Schmitt A, Selvaraj S, D'Antonio M, D'Antonio-Chronowska A, Smith EN *et al*: **Subtle changes in chromatin loop contact propensity are associated with differential gene regulation and expression**. *Nat Commun* 2019, **10**(1):1054.
15. Campbell SA, Hoffman BG: **Chromatin Regulators in Pancreas Development and Diabetes**. *Trends Endocrinol Metab* 2016, **27**(3):142-152.
16. Muller C, Leutz A: **Chromatin remodeling in development and differentiation**. *Curr Opin Genet Dev* 2001, **11**(2):167-174.
17. van Arensbergen J, Dussaud S, Pardanaud-Glavieux C, Garcia-Hurtado J, Sauty C, Guerci A, Ferrer J, Ravassard P: **A distal intergenic region controls pancreatic endocrine differentiation by acting as a transcriptional enhancer and as a polycomb response element**. *PloS one* 2017, **12**(2):e0171508.

18. Alvarez-Dominguez JR, Donaghey J, Rasouli N, Kenty JHR, Helman A, Charlton J, Straubhaar JR, Meissner A, Melton DA: **Circadian Entrainment Triggers Maturation of Human In Vitro Islets**. *Cell Stem Cell* 2020, **26**(1):108-122 e110.
19. Mellor J: **The dynamics of chromatin remodeling at promoters**. *Mol Cell* 2005, **19**(2):147-157.
20. Lawlor N, Marquez EJ, Orchard P, Narisu N, Shamim MS, Thibodeau A, Varshney A, Kursawe R, Erdos MR, Kanke M *et al*: **Multiomic Profiling Identifies cis-Regulatory Networks Underlying Human Pancreatic beta Cell Identity and Function**. *Cell Rep* 2019, **26**(3):788-801 e786.
21. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ: **Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position**. *Nat Methods* 2013, **10**(12):1213-1218.
22. Andrey G, Mundlos S: **The three-dimensional genome: regulating gene expression during pluripotency and development**. *Development* 2017, **144**(20):3646-3658.
23. Miguel-Escalada I, Bonas-Guarch S, Cebola I, Ponsa-Cobas J, Mendieta-Esteban J, Atla G, Javierre BM, Rolando DMY, Farabella I, Morgan CC *et al*: **Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes**. *Nat Genet* 2019, **51**(7):1137-1148.
24. Birnbaum RY, Clowney EJ, Agamy O, Kim MJ, Zhao J, Yamanaka T, Pappalardo Z, Clarke SL, Wenger AM, Nguyen L *et al*: **Coding exons function as tissue-specific enhancers of nearby genes**. *Genome Res* 2012, **22**(6):1059-1068.
25. Kulaeva OI, Nizovtseva EV, Polikanov YS, Ulianov SV, Studitsky VM: **Distant activation of transcription: mechanisms of enhancer action**. *Mol Cell Biol* 2012, **32**(24):4892-4897.
26. Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J *et al*: **Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation**. *Cell* 2012, **148**(1-2):84-98.
27. Avrahami D, Li C, Zhang J, Schug J, Avrahami R, Rao S, Stadler MB, Burger L, Schubeler D, Glaser B *et al*: **Aging-Dependent Demethylation of Regulatory Elements Correlates with Chromatin State and Improved beta Cell Function**. *Cell Metab* 2015, **22**(4):619-632.
28. Dayeh T, Volkov P, Salo S, Hall E, Nilsson E, Olsson AH, Kirkpatrick CL, Wollheim CB, Eliasson L, Ronn T *et al*: **Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion**. *PLoS Genet* 2014, **10**(3):e1004160.
29. Dhawan S, Tschen SI, Zeng C, Guo T, Hebrok M, Matveyenko A, Bhushan A: **DNA methylation directs functional maturation of pancreatic beta cells**. *J Clin Invest* 2015, **125**(7):2851-2860.
30. Golson ML, Kaestner KH: **Epigenetics in formation, function, and failure of the endocrine pancreas**. *Mol Metab* 2017, **6**(9):1066-1076.
31. Pullen TJ, Rutter GA: **When less is more: the forbidden fruits of gene repression in the adult beta-cell**. *Diabetes Obes Metab* 2013, **15**(6):503-512.
32. Bhandare R, Schug J, Le Lay J, Fox A, Smirnova O, Liu C, Naji A, Kaestner KH: **Genome-wide analysis of histone modifications in human pancreatic islets**. *Genome Res* 2010, **20**(4):428-433.
33. Belgardt BF, Ahmed K, Spranger M, Latreille M, Denzler R, Kondratiuk N, von Meyenn F, Villena FN, Herrmanns K, Bosco D *et al*: **The microRNA-200 family regulates pancreatic beta cell survival in type 2 diabetes**. *Nat Med* 2015, **21**(6):619-627.
34. Calderari S, Diawara MR, Garaud A, Gauguier D: **Biological roles of microRNAs in the control of insulin secretion and action**. *Physiol Genomics* 2017, **49**(1):1-10.
35. Dumortier O, Fabris G, Van Obberghen E: **Shaping and preserving beta-cell identity with microRNAs**. *Diabetes Obes Metab* 2016, **18 Suppl 1**:51-57.
36. Motterle A, Sanchez-Parra C, Regazzi R: **Role of long non-coding RNAs in the determination of beta-cell identity**. *Diabetes Obes Metab* 2016, **18 Suppl 1**:41-50.

37. Pasquali L, Gaulton KJ, Rodriguez-Segui SA, Mularoni L, Miguel-Escalada I, Akerman I, Tena JJ, Moran I, Gomez-Marin C, van de Bunt M *et al*: **Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants**. *Nat Genet* 2014, **46**(2):136-143.
38. Cebola I: **Pancreatic Islet Transcriptional Enhancers and Diabetes**. *Curr Diab Rep* 2019, **19**(12):145.
39. Greenwald WW, Chiou J, Yan J, Qiu Y, Dai N, Wang A, Nariai N, Aylward A, Han JY, Kadakia N *et al*: **Pancreatic islet chromatin accessibility and conformation reveals distal enhancer networks of type 2 diabetes risk**. *Nat Commun* 2019, **10**(1):2078.
40. DiGruccio MR, Mawla AM, Donaldson CJ, Noguchi GM, Vaughan J, Cowing-Zitron C, van der Meulen T, Huising MO: **Comprehensive alpha, beta and delta cell transcriptomes reveal that ghrelin selectively activates delta cells and promotes somatostatin release from pancreatic islets**. *Mol Metab* 2016, **5**(7):449-458.
41. Benner C, van der Meulen T, Caceres E, Tigyi K, Donaldson CJ, Huising MO: **The transcriptional landscape of mouse beta cells compared to human beta cells reveals notable species differences in long non-coding RNA and protein-coding gene expression**. *BMC genomics* 2014, **15**:620.
42. van der Meulen T, Lee S, Noordeloos E, Donaldson CJ, Adams MW, Noguchi GM, Mawla AM, Huising MO: **Artemether Does Not Turn alpha Cells into beta Cells**. *Cell Metab* 2018, **27**(1):218-225 e214.
43. Mawla AM, Huising MO: **Navigating the Depths and Avoiding the Shallows of Pancreatic Islet Cell Transcriptomes**. *Diabetes* 2019, **68**(7):1380-1393.
44. Mawla AM, Huising MO: **epiRomics: a multi-omics R package to identify and visualize enhancers**. *bioRxiv* 2021:2021.2008.2019.456732.

II. Chapter 2: Comprehensive alpha, beta and delta cell transcriptomes reveal that ghrelin selectively activates delta cells and promotes somatostatin release from pancreatic islets

Michael R. DiGruccio, **Alex M. Mawla**, Cynthia J. Donaldson, Glyn M. Noguchi, Joan Vaughan, Christopher Cowing-Zitron, Talitha van der Meulen, Mark O. Huising

Molecular Metabolism 2016. 5(7):449-458. doi: 10.1016/j.molmet.2016.04.007

Contributions to Jointly Authored Works: As second author on this manuscript, I was responsible for the entirety of formal analysis and interpretation of all transcriptomic data. I analyzed all RNA-Seq data for the characterization of alpha, beta, and delta cell transcriptomes. I was responsible for all panels translating the analysis into meaningful biology. I was also responsible for creating and maintain the genome browser for these data through our lab website so other colleagues in our field can visualize the findings highlighted in the paper, and use these data as a resource to reference in their own studies.

Significance of Research: This manuscript provided a novel and comprehensive portrait of the transcriptomic landscape of delta cells and identified the ghrelin receptor as a pertinent aspect of regulating cross-talk between alpha, beta, and delta cells. This paper was also featured in an editorial by Molecular Metabolism.

Citations: 191

Original article

MOLECULAR METABOLISM

Comprehensive alpha, beta and delta cell transcriptomes reveal that ghrelin selectively activates delta cells and promotes somatostatin release from pancreatic islets

Michael R. DiGruccio [1], Alex M. Mawla [1], Cynthia J. Donaldson [2], Glyn M. Noguchi [1], Joan Vaughan [2], Christopher Cowing-Zitron [1], Talitha van der Meulen [1], Mark O. Huising [1,*]

ABSTRACT

Objective: Complex local crosstalk amongst endocrine cells within the islet ensures tight coordination of their endocrine output. This is illustrated by the recent demonstration that the negative feedback control by delta cells within pancreatic islets determines the homeostatic set-point for plasma glucose during mouse postnatal development. However, the close association of islet endocrine cells that facilitates paracrine crosstalk also complicates the distinction between effects mediated directly on beta cells from indirect effects mediated via local intermediates, such as somatostatin from delta cells.
Methods: To resolve this problem, we generated reporter mice that allow collection of pure pancreatic delta cells along with alpha and beta cells from the same islets and generated comprehensive transcriptomes for each islet endocrine cell type. These transcriptomes afford an unparalleled view of the receptors expressed by delta, alpha and beta cells, and allow the prediction of which signal targets which endocrine cell type with great accuracy.
Results: From these transcriptomes, we discovered that the ghrelin receptor is expressed exclusively by delta cells within the islet, which was confirmed by fluorescent in situ hybridization and qPCR. Indeed, ghrelin increases intracellular calcium in delta cells in intact mouse islets, measured by GCaMP6 and robustly potentiates glucose-stimulated somatostatin secretion on mouse and human islets in both static and perfusion assays. In contrast, des-acyl-ghrelin at the same dose had no effect on somatostatin secretion and did not block the actions of ghrelin.
Conclusions: These results offer a straightforward explanation for the well-known insulinostatic actions of ghrelin. Rather than engaging beta cells directly, ghrelin engages delta cells to promote local inhibitory feedback that attenuates insulin release. These findings illustrate the power of our approach to resolve some of the long-standing conundrums with regard to the rich feedback that occurs within the islet that is integral to islet physiology and therefore highly relevant to diabetes.

Keywords Ghrelin; Delta cell; Somatostatin release; Transcriptome; Beta cell; Alpha cell

1. INTRODUCTION

Pancreatic alpha and beta cells co-localize in an arrangement that facilitates accurate coordination of glucagon and insulin release from islets. While glucagon and insulin, along with Islet amyloid polypeptide (Iapp), are among the few islet signals that meet the classic definition of a hormone as a 'factor that is released into the general circulation to signal at a distant site in the body', a rich constellation of paracrine and neural interactions takes place within the islet to ensure tight control over their release [1–3]. Pancreatic delta cells are the third-most common endocrine cell type in the islets, and the somatostatin they release is an important inhibitor of both insulin and glucagon [4–6]. We recently described a novel negative feedback loop where the paracrine peptide Urocortin 3 (Ucn3) is co-released with insulin from beta cells and promotes glucose-stimulated somatostatin release via the type 2 corticotropin-releasing hormone (Crhr2) receptor expressed

The transcriptome data reported in this paper can be accessed at http://huisinglab.com/islet_txomes_2016/ and have been deposited in the Gene Expression Omnibus (GEO) repository under accession number GSE80673.

[1]Department of Neurobiology, Physiology & Behavior, College of Biological Sciences, University of California, One Shields Avenue, 196 Briggs Hall, Davis, CA, USA [2]Clayton Foundation Laboratories for Peptide Biology, The Salk Institute for Biological Studies, 10010 North Torrey Pines road, La Jolla, CA, USA

*Corresponding author. Tel.: +1 530 752 4670; fax: +1 530 752 5582.

E-mails: mrdigruccio@ucdavis.edu (M.R. DiGruccio), ammawla@ucdavis.edu (A.M. Mawla), donaldson@salk.edu (C.J. Donaldson), gmnoguchi@ucdavis.edu (G.M. Noguchi), vaughan@salk.edu (J. Vaughan), tvandermeulen@ucdavis.edu (T. van der Meulen), mhuising@ucdavis.edu (M.O. Huising).

Abbreviations: Crhr2, Corticotropin-releasing hormone receptor type 2; FISH, Fluorescent in situ hybridization; Ghsr, Growth hormone secretagogue receptor; GSSS, Glucose-stimulated somatostatin secretion; Iapp, Islet amyloid polypeptide; RPKM, Reads per kilobase gene model per million reads sequenced; Trpm2, Transient receptor potential melastatin 2; Ucn3, Urocortin 3; YFP, Yellow fluorescent protein

Received March 10, 2016 • Revision received April 14, 2016 • Accepted April 21, 2016 • Available online 3 May 2016

http://dx.doi.org/10.1016/j.molmet.2016.04.007

by delta cells [7]. Somatostatin then attenuates further insulin secretion and helps to maintain stable and tight control over plasma glucose. The appearance of Ucn3 across the beta cell mass correlates with the well-known uptick in plasma glucose levels [8–10]. The onset of Ucn3 is directly responsible for this phenomenon by initiating delta cell-dependent feedback on insulin release, thus determining the homeostatic set-point for plasma glucose [7]. Delta cell-mediated feedback breaks down early in diabetes, which leads to marked increases in plasma glucose fluctuations that resemble the glycemic volatility that impacts patients across the diabetes spectrum [7].

While these observations highlight the physiological importance of delta cell-dependent feedback, we know relatively little about the cues that control delta cells. The intimate co-localization of islet endocrine cells has also made it difficult to distinguish direct effects on beta cells from indirect actions that are mediated by locally produced intermediaries. We previously generated a transgenic mIns1-H2b-mCherry reporter mouse in which beta cells are labeled by the nuclear expression of mCherry driven by the mIns1 promoter and crossed these to S100b-eGFP reporter mice that we serendipitously discovered to label alpha cells [11]. These reporter mice enabled the comparison of mouse and human beta cell transcriptomes but did not fully resolve the transcriptomes of pancreatic alpha and delta cells. To improve substantially on these prior observations and to determine the factors that directly engage delta cells, we now generated a set of triple transgenic mouse models in which mIns1-H2b-mCherry [11] beta cells are crossed to mice with alpha or delta cells are marked by YFP in a Cre-dependent fashion. We used these reporter mice to generate unbiased and comprehensive transcriptomes of pancreatic delta and alpha cells, along with beta cells from the same islets. Our transcriptomes are validated by the strong enrichment for a large panel of known alpha, beta, and delta cell markers in the appropriate cell type. We then use this information to determine the signals that directly engage the delta cell by virtue of selectively expressed receptors and analyze ghrelin stimulation of delta cells as an example to highlight the utility of our approach. We discovered that ghrelin acts directly on mouse and human delta cells to promote somatostatin release. Our observations offer a straightforward explanation for the well-known insulinostatic actions of ghrelin (e.g. [12–16]) and illustrate the physiological importance of delta cell-mediated feedback within pancreatic islets. An accurate understanding of ghrelin's mechanism of action within pancreatic islets is highly relevant given ghrelin's central role in the regulation of energy and glucose metabolism [17].

2. RESEARCH DESIGN AND METHODS

2.1. Biological materials and ethics statements

All mouse procedures were approved by the UC Davis or the Salk Institute for Biological Studies Institutional Animals Care and Use Committee and were performed in compliance with the Animal Welfare Act and the Institute for Laboratory Animal Research (ILAR) Guide to the Care and Use of Laboratory Animals. Animals were maintained on a 12-h light/12-h dark cycle with free access to water and standard rodent chow. Static and dynamic hormone secretion experiments were carried out on C57BL/6NHsd mice, between 8 and 16 weeks of age, from Harlan (Indianapolis, IN). We obtained human islets via the Integrated Islet Distribution Program; the receipt was declared exempt from IRB review under 45 CRR 46.101 (b) category (4).

2.2. Immunofluorescence and FISH

Immunofluorescence was conducted as previously described [7,8]. Insulin was detected using guinea pig anti-insulin (Dako #A0564; 1:500), somatostatin using sheep anti-somatostatin (American Research Products Inc. #13-2366, 1:1000), Ucn3 using rabbit anti-Ucn3 (#6570; in house, 1:2000), Glucagon using rabbit anti-glucagon (Abcam #ab11022-1; 1:200), and YFP using goat anti-GFP (Rockland 600-101-215; 1:1000). All secondary antibodies were obtained from Jackson Laboratories Inc. RNA FISH was conducted using RNAscope probes for *Ghsr* and *Sst* (Advanced Cell Diagnostics) according to the manufacturers instructions.

2.3. Islet isolation and FACS sorting

Islet isolation was conducted as previously described [7,11,18]. Islets from mIns1-H2b-mCherry [11] (deposited with the Jackson laboratories as strain #28589) × Rosa-LSL-YFP [19] × Sst-Cre [20] or Gcg-Cre [21] triple transgenic animals were pooled by sex in 2 (Sst-Cre) or 3 (Gcg-Cre) replicate groups of a dozen animals. FACS sorting was conducted as described previously [7,11] with each sample collected directly in Trizol to ensure immediate cell lysis and preservation of RNA integrity.

2.4. Next generation sequencing and bioinformatics

RNA was isolated from Trizol-preserved samples by chloroform extraction and cleaned up over an RNeasy microcolumn essentially as previously described [11]. RNA quality was verified by Tapestation (Agilent, Santa Clara, CA). Indexed sequencing libraries were constructed using the TruSeq RNA sample Prep Kit v2 (Illumina Inc. San Diego, CA), sequenced at 50 cycles, and single read on an Illumina HiSeq 2000 platform. Results were validated by qPCR using Sybr chemistry and the primers listed in Table 1. Sequencing reads were mapped to the mouse genome version GenCode M5 (GRCm38.p3) using STAR v2.4 [22]. On average over 33 million reads were sequenced for each library with 89.9% of sequenced reads aligning (>63% unique alignment overall). FeatureCounts [23] was used to create count tables of the sorted bam files using reads aligning to RefSeq-defined exons. EdgeR version 3.12.0 [24] was used to conduct pairwise statistical comparisons. Wordles of transcript abundance were generated on wordle.net. Single cell RNAseq data from [25] were used to generate the violin plots in Figure 2C. Cells that had an RPKM value > 10 k of either *Sst*, or *Ins2*, or *Gcg* were defined as delta, beta, or alpha cells, respectively.

2.5. Functional imaging by GCaMP6

Sst-Cre mice and LSL-GCaMP6 mice (Jackson laboratories strain #24106) were crossed for functional imaging. Intact islets from bitransgenic offspring were plated on poly-D lysine-coated number 1.5 35 mm glass cover slip tissue culture dishes (Mattek) and maintained at 37 °C 5% CO_2 in RPMI 10% FBS, 5.5 mM glucose with pen/strep.

Table 1 — qPCR primer information.

Ref Seq ID	Gene	Primer	Sequence 5′→3′	Amplicon size (bp)
NM_008100	Gcg	qrodGcg.fwu1	TCACAGGGCACATTCACCAG	121
		qrodGcg.rvu1	CATCATGACGTTTGGCAATGTT	
NM_001185084	Ins2	qrodIns2.fwu1	GCTCTCTACCTGGTGTGTGGG	128
		qrodIns2.rvu1	CAAGGTCTGAAGGTCACCTGC	
NM_009215	Sst	qrodSst.fwu1	GACCCCAGACTCCGTCAGTTT	112
		qrodSst.rvu1	TCTCTGTCTGGTTGGGCTCG	
NM_021488	Ghsr	qmGhsr.fwu1	GACCAGAACCACAAACAGACAG	113
		qmGhsr.rvu1	GGCTCGAAAGACTTGGAAAA	
NM_013556	Hprt	qmHPRT.fwu	TCCTCCTCAGACCGCTTTT	90
		qmHPRT.rvu	CCTGGTTCATCATCGCTAATC	

Islets were allowed to adhere for 24 h prior to imaging using a Nikon A1R confocal microscope. During live cell fluorescence acquisition, islets were continuously superfused with 37 °C KRB bubbled with 95% O_2 and 5% CO_2. To measure baseline, high glucose, glucose plus ghrelin, and control responses, buffers were switched at the indicated time points. The position of delta cells within the islet in X, Y, and Z were first identified during a brief pulse with 16.8 mM glucose. Afterwards, islets were washed with 2.8 mM glucose for 30 min to allow fluorescence to return to baseline. For fluorescence intensity analysis, GCaMP6 glucose responsive regions of interest corresponding to single delta cells were identified and defined during the maximal depolarization by 30 mM KCl. Data analysis was performed using Nikon Elements software.

2.6. Hormone secretion

Secretion experiments were conducted in Krebs—Ringer Buffer (KRB). We hand-picked the required number of islets from a pool of all islets and assigned these as the next replicate to each subsequent treatment. Static somatostatin secretion was carried out using 50 or 100 islets per well for mouse islets under high or low glucose, respectively. 100 islets/well were used for human islet release. Dynamic somatostatin release was determined using a custom-built perfusion setup using 200 islets in 6 parallel chambers. We measured somatostatin by RIA [26] using anti-Sst14 antiserum (S201; diluted 1:50,000) [27] as described [7]. EC50 and minimal detectable dose are 12 pg/tube and 1 pg/tube, respectively. Insulin was measured by commercial RIA (Millipore) as previously described [18]. Ghrelin, des-acyl-ghrelin, Astressin2b, and the Sstr3 antagonist Sst3−ODN−8 (Carbamoyl−des−$AA^{1,2,4,5,12,13}$[D−Cys3,Tyr^{7},D−Agl^{8} (Me,2−napthoyl)]−SS; #315−260−15) [28] were generously synthesized and provided by Drs. Michael Beyermann and Jean Rivier.

2.7. Statistical analyses

Statistical analyses were carried out in Prism 6.0e for Mac (GraphPad Software, Inc.). We reported all values as mean values across bio-logical replicates and assumed normality, unless otherwise noted. We evaluated experiments with more than two treatments by ANOVA followed by Student's *t*-test to determine which means differed statistically, and we applied Welch's correction for unequal variance when necessary. We evaluated perfusion data by two-way ANOVA for treatment and the interaction of treatment and time for each block.

3. RESULTS

3.1. Generation and validation of alpha, beta and delta cell transcriptomes

To obtain transcriptomes of mouse alpha, beta and delta cells, we crossed our mIns1-H2b-mCherry mice [11] to either Sst-Cre [20] or Gcg-Cre [21] mice and a Rosa-LSL-YFP reporter [19] (Figure 1A—D). The resulting triple transgenic offspring enabled FACS purification of highly pure beta cells and delta or alpha cells that we used to generate comprehensive transcriptomes of mouse delta, beta, and alpha cells. Beta cell transcriptomes were markedly enriched for mCherry expression (Figure 1E), while alpha and delta cell transcriptomes were each marked by strong enrichment for YFP (Figure 1F), underpinning our robust FACS-based purification. Alpha, beta, and delta cells are all endocrine cells most noted for the release of glucagon, insulin, and somatostatin, respectively. Indeed, somatostatin transcripts are by far the most abundant transcripts in delta cells and account for 3.4 ± 0.25% of all aligned reads in delta cells (Figure 1G). Insulin (encoded by Ins2 and Ins1 genes in mouse) makes up 18.9 ± 2.4% of all reads in the beta cell transcriptome (Figure 1H), while glucagon transcripts make up 27.8 ± 0.9% of all reads in alpha cells (Figure 1I). Iapp is the second-most abundant transcript in our delta cell transcriptomes and while Iapp is best-known as a beta cell hormone, delta cells also express Iapp [29].

3.2. Transcriptome validation

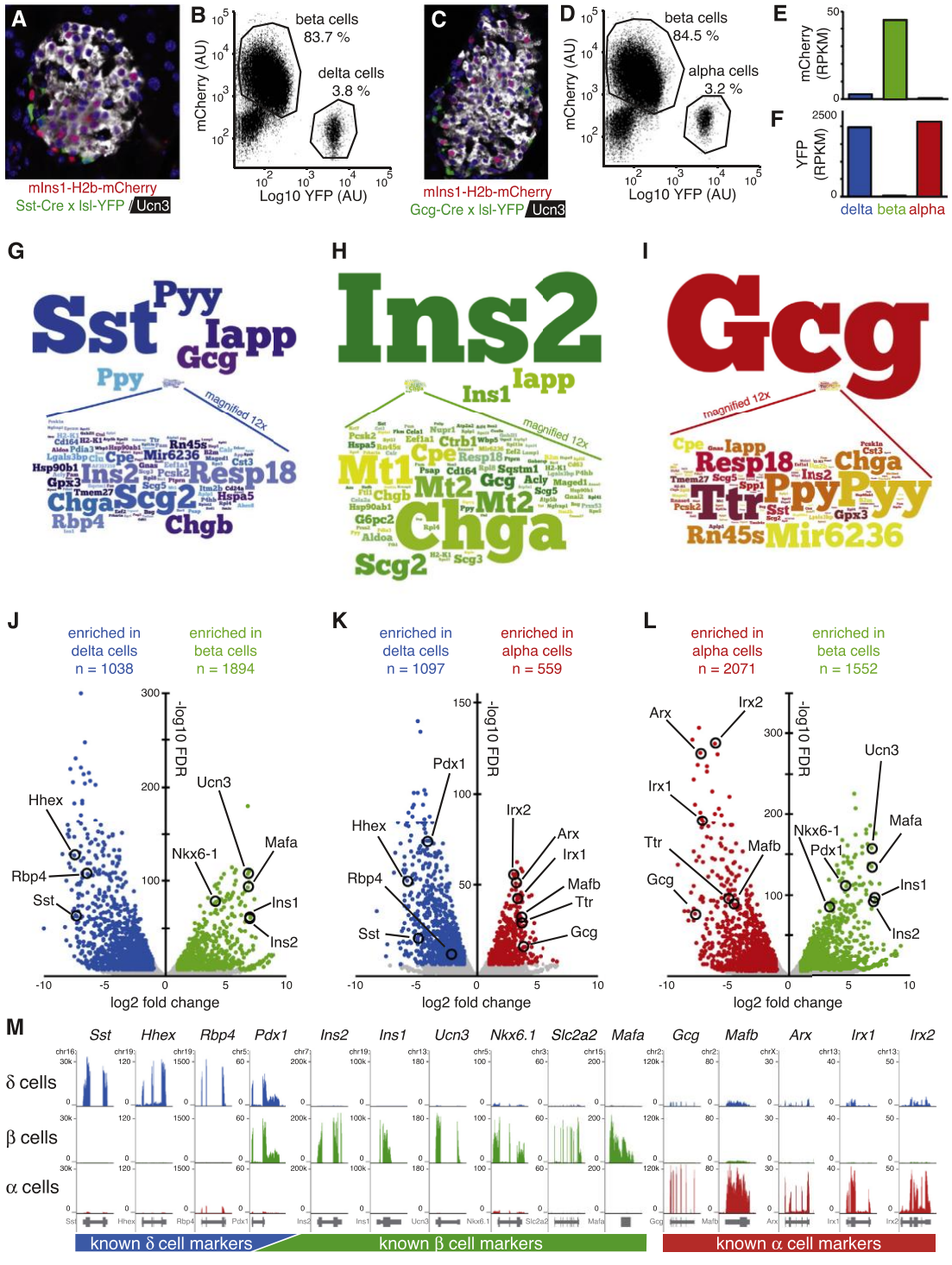
We observed on average 1369 genes that were differentially enriched in each pairwise comparison, as defined by a Log2-fold expression > 1 or < -1 and a FDR < 0.00001. Known delta cell markers such as *Sst*, *Rbp4,* and *Hhex* [30,31] were highly enriched in pairwise comparisons to beta and alpha cells (Figure 1J,K). Known beta cell markers such as *Ucn3*, *Mafa*, *Ins1*, *Ins2,* and *Nkx6-1* [11] were markedly enriched in beta compared to delta and alpha cells (Figure 1J,L). Alpha cell markers such as *Gcg*, *Ttr*, *Irx1*, *Irx2*, *Arx,* and *Mafb* [30,32,33] were enriched in alpha cells compared to delta and beta cells (Figure 1K,L). As *Pdx1* is expressed in beta and delta cells (Figure 1M), it was enriched in neither cell type upon pairwise comparison (Figure 1J).

3.3. Islet cell transcriptomes reveal GPCR expression profiles

Given the importance of paracrine interactions to control islet insulin and glucagon output [7], we assessed the expression of GPCRs in our islet transcriptomes in more detail. As noted, *Crhr2* is expressed exclusively by delta cells, but several GPCRs that are well-known for their beta cell expression are also expressed to a lesser extent by delta cells but absent from alpha cells (Figure 2A). This group includes glucagon-like peptide 1 receptor (*Glp1r*), the alpha 2 adrenergic receptor (*Adra2a*), and the glucagon receptor (*Gcgr*). In contrast, the gastric inhibitory polypeptide receptor (*Gipr*) is ubiquitously expressed in the islet and is detected in approximately equal measures by alpha and delta cells as well. In contrast, the galanin receptor 1 (*Galr1*) is expressed selectively in beta cells. In searching for GPCRs that, like *Crhr2*, are expressed selectively by delta cells, we noted the delta cell-specific expression of *Sstr1*. Alpha cells selectively express *Sstr2*, while *Sstr3* is expressed by alpha, beta and delta cells. There is no detectable expression of *Sstr4* or *Sstr5* in any of the endocrine cells of the mouse islets. Another example of a GPCR that caught our attention as expressed selectively by delta cells is the growth hormone secretagogue receptor (*Ghsr*), which encodes the GHS-1R that responds to ghrelin (Figure 2A). We validated the selective expression of *Ghsr* in delta cells by qPCR (Figure 2B). A recently published data set of high quality, single cell RNAseq of dissociated wild type mouse islets that contains populations of *Sst*-expression delta cells, *Ins2*-expression beta cells and *Gcg*-expression alpha cells [25] supported these observations, as *Ghsr* in these data was also detectable only in delta cells (Figure 2C). We further confirmed *Ghsr* mRNA expression in delta cells by FISH. We found that *Ghsr* mRNA (green) colocalizes with *Sst* mRNA (red) in cells at the islet periphery that co-express SST peptide (white) (Figure 2D) but not GCG peptide (white) (Figure 2E).

3.4. Ghrelin directly activates calcium in delta cells

GHS-1R primarily couples through Gαq [34], which results in phospholipase C-mediated intracellular calcium release. Thus, we evaluated the calcium response of delta cells to ghrelin in intact islets. We used islets from bitransgenic Sst-Cre [20] × LSL-GCaMP6 mice that express the Ca^{2+} sensor GCaMP6 [35] specifically and selectively in delta cells. After establishing baseline fluorescence under KRB with 2.8 mM glucose, increasing glucose to 16.8 mM glucose led to a



MOLECULAR METABOLISM

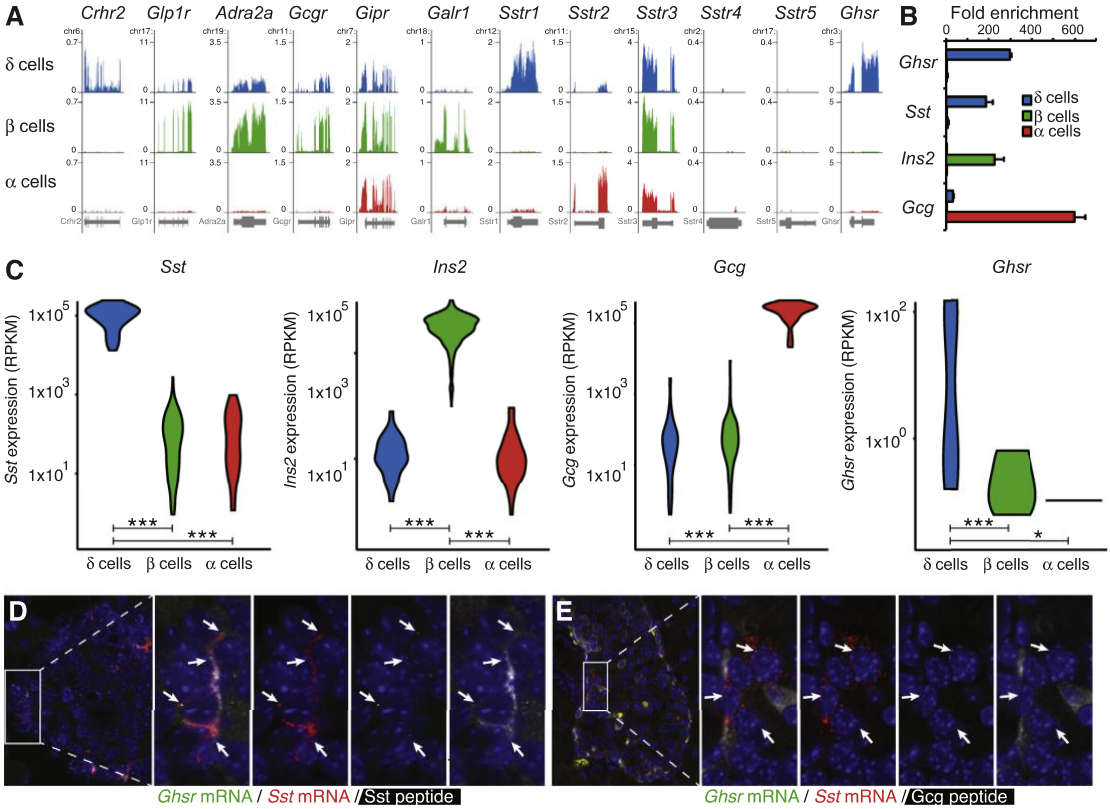


Figure 2: Delta cells selectively express *Ghsr*. A: Normalized browser plots illustrating the expression of a series of GPCRs in delta, beta and alpha cells of the mouse islet. B: Confirmation by qPCR that *Ghsr* message is selectively expressed by pancreatic delta cells. C: Violin plots of single cell RNA-seq of wild type mouse pancreatic islet cells [25] confirms that *Ghsr* expression is detectable only in delta cells. D: FISH confirmation of the expression of *Ghsr* gene (green dots) in pancreatic delta cells of wild type mice, colocalized with *Sst* message (red dots) and Sst peptide (white). E: mRNA for *Ghsr* (green dots) and *Sst* (red dots) co-localizes in a peripheral islet population that does not express Gcg peptide (white). *P $<$ 0.05; **P $<$ 0.01; ***P $<$ 0.001.

marked increase in delta cell GCaMP6 fluorescence intensity, which gradually tapered down (Figure 3; Supplementary Video 1). Upon subsequent addition of 100 nM ghrelin delta cells responded with a large initial Ca^{2+} peak that was followed by a distinct oscillatory response (Figure 3). The Ca^{2+} signal returned to baseline upon cessation of stimulation. Ghrelin stimulation under 2.8 mM glucose produced no increases GCaMP6 signal (not shown). Depolarization by 30 mM KCl at the conclusion of the experiment resulted in a robust GCaMP6 calcium peak, confirming delta cell viability and responsiveness throughout the experiment.

3.5. Ghrelin promotes somatostatin secretion

Given the increase in intracellular calcium upon activation of GHS-1R on delta cells and the importance of calcium in hormone secretion, we proceeded to determine if ghrelin promotes somatostatin secretion. Indeed, stimulation with ghrelin strongly potentiated glucose-stimulated somatostatin secretion (GSSS) in mouse islets (Figure 4A). Des-acyl-ghrelin, which lacks the octanyl post-translation modification at the Ser3 position and does not activate GHS-1R, did not potentiate GSSS. To determine if the actions ghrelin depended on the feedback mediated by beta cell-derived Ucn3 [7], we blocked the

Figure 1: Generation and validation of the delta, beta, and alpha cell transcriptomes. A: Islet of a mIns1-H2b-mCherry × Sst-Cre × LSL-YFP triple transgenic mouse labels all beta cells with nuclear mCherry (counterstained in white with Ucn3) and labels all delta with YFP (stained in green). B: FACS plots of dissociated beta and delta cells from this same cross. C: Islet of a mIns1-H2b-mCherry × Gcg-Cre × LSL-YFP triple transgenic mouse labels all beta cells with nuclear mCherry (counterstained in white with Ucn3) and labels alpha cells with YFP (stained in green). D: FACS plots of dissociated alpha and delta cells from this same cross. E, F: Beta cell transcriptomes are highly enriched for mCherry reads while delta and alpha cell transcriptomes are highly enriched for YFP reads, confirming our transgenic purification strategy. RPKM = reads per kilobase gene model per million reads sequenced. G–I: Graphical representation of the relative expression of the 100 most-abundantly detected transcripts in each cell type. Note that most of these genes are expressed at levels that are so much lower compared to the most abundant transcript in each endocrine cell type that they were magnified 12x to maintain legibility. J–L: Volcano plots listing the number of significantly enriched genes for each pairwise comparison. Selected markers of alpha, beta, and delta cell identity are highlighted for visual reference. M: Genome browser plots comparing the expression of a large panel of markers with well-established expression patterns validate the high degree of purity achieved in these delta, beta, and alpha cells transcriptomes.

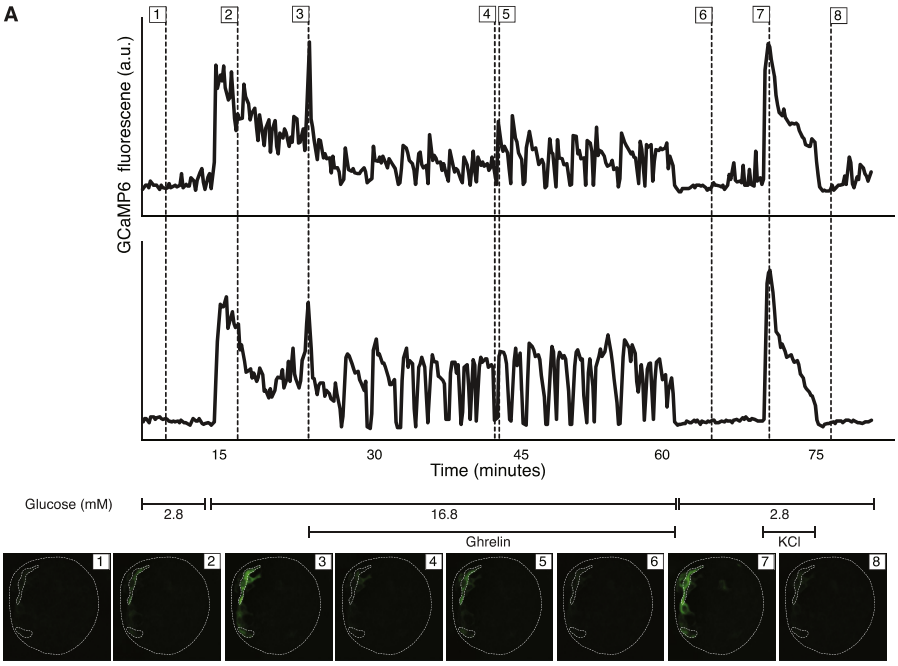


Figure 3: Ghrelin activates an intracellular calcium response as measured by GCaMP6 in intact islets. A: Calcium responses were measured over time in intact islets using the genetically encoded calcium sensor GCaMP6 expressed selectively in delta cells (Sst-Cre × LSL-GCaMP6). Calcium traces of two representative cells in response to changes in glucose concentration and stimulation with 100 nM ghrelin or 30 mM KCl, as indicated. Thumbnails of the islet at the indicated time points in each graph are given, with the top and bottom region of interest corresponding to the top and bottom trace, respectively. Movie of the full calcium trace for this figure is included as Supplemental Video 1.

actions of endogenous Ucn3 with the CRHR2-selective antagonist Astressin2b. Astressin2b attenuated GSSS as previously reported [7], but did not prevent the ghrelin-induced potentiation of GSSS (Figure 4B). This indicates that ghrelin engages delta cells directly, independent of Ucn3-mediated potentiation of GSSS [7]. Ghrelin, but not des-acyl-ghrelin, robustly potentiated GSSS from human islets *in vitro*, confirming that its actions are similar across rodents and primates (Figure 4C). We next tested the actions of ghrelin on somatostatin secretion in an islet perfusion setup and observed that islets demonstrated robust potentiation of GSSS in response to 100 nM ghrelin. Islets that did not receive ghrelin continued to maintain glucose-stimulated somatostatin levels (Figure 4D). Des-acyl-ghrelin did not elicit any response following a 15-minute pre-incubation and did not block the potentiation of GSSS by ghrelin (Figure 4E). Islets in all perfusion experiments responded to brief depolarization with 30 mM KCl with a strong somatostatin peak, confirming islet viability.

3.6. Ghrelin attenuates insulin release in a somatostatin-dependent manner

We next assessed whether the insulinostatic effect of ghrelin depended on intact somatostatin-mediated feedback within the islet. Stimulation of wild type mouse islets with ghrelin attenuated glucose-stimulated insulin secretion (Figure 5), in line with most published reports. Co-administration with an antagonist to the Sstr3 receptor [28], which is the only Sst receptor that is detectably expressed by primary mouse beta cells (Figure 2A), blocked the insulinostatic actions of ghrelin completely (Figure 5). These observations demonstrate that ghrelin's insulinostatic actions depend on intact Sstr3-dependent feedback within the islet. The fact that Sstr3 blockade promoted glucose-stimulated insulin secretion suggest that this Sstr3-mediated feedback exerts a tonic inhibitory effect on glucose-stimulated somatostatin secretion.

4. DISCUSSION

The importance of the neural, endocrine, and paracrine inputs to pancreatic islets that collectively determine the output of insulin and glucagon has long been recognized. However, the tools and techniques at our disposal to map this complex local crosstalk have lacked the resolution and specificity to accurately determine the origin of many signals and the identity of the islet cell receptive to its message. The recent onset of next-generation sequencing approaches has now made it attainable to conduct an unbiased evaluation of the expression of all genes in any given tissue that is quantitative over many orders of magnitude. The purity and quality of the input sample has become the factor limiting the usefulness of the data achieved in this way. For pancreatic islets, which are often equated with beta cells, but which contain many additional endocrine and non-endocrine cells, we resolved this issue by generating triple transgenic reporter mice that collectively enable the FACS-purification of alpha, beta, and delta cells from the same islets. From these highly pure input samples, we generated comprehensive gene expression compendia of the three principal endocrine cells within the islets. This now enables us to unravel with great accuracy some of the crosstalk pathways that had to date been unknown. Our current transcriptome data are a distinct improvement over the work we published previously [11] and offer a

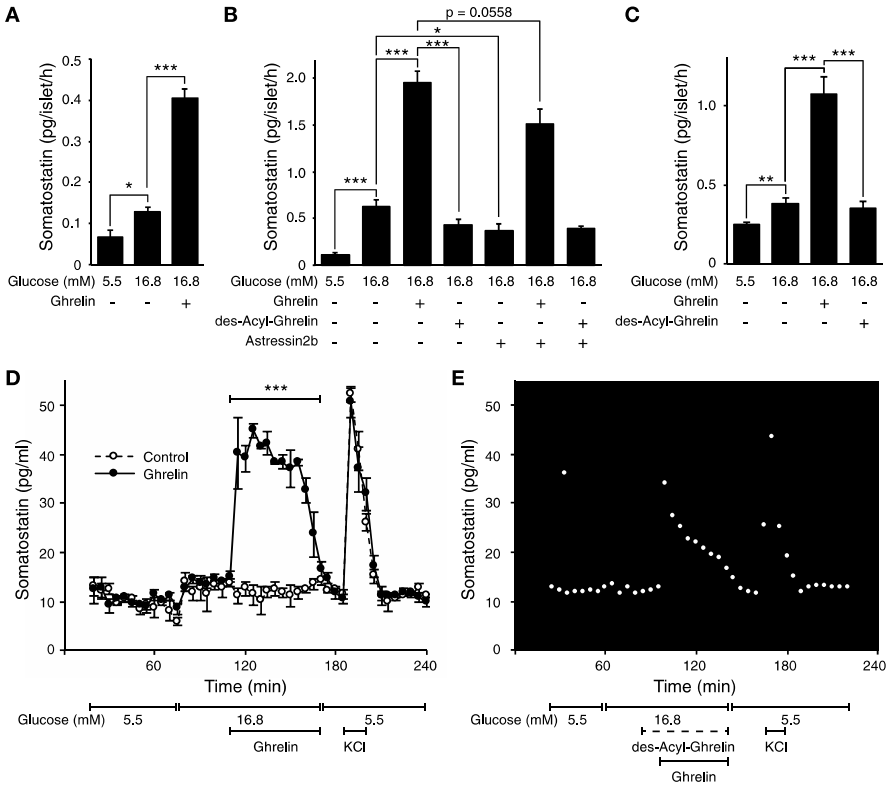


Figure 4: Ghrelin potentiates glucose-stimulated somatostatin secretion. A: Ghrelin significantly increases somatostatin release from mouse islets *in vitro*. B: Ghrelin, but not des-acyl-ghrelin, promotes somatostatin release. Astressin2b attenuates glucose-stimulated somatostatin release by preventing beta cell-derived Ucn3 from stimulating delta cells. Astressin2b does not prevent ghrelin from stimulating somatostatin release, indicating that ghrelin's actions are independent from the feedback mediated by Ucn3. C: Ghrelin, but not des-acyl-ghrelin, promotes glucose stimulated somatostatin secretion from human islets *in vitro*. D: Perfusion of mouse islets under high glucose with ghrelin (black circles) acutely potentiates glucose-stimulated somatostatin secretion compared to control (open circles). E: Application of des-acyl-ghrelin before and during stimulation does not affect the ability of ghrelin to potentiate glucose-stimulated somatostatin secretion. All mice were wild type, and all peptides were applied at 100 nM final concentration. Values represent the mean ± SEM for 8 (A), 7 (B) or 6 (C) wells per group (A–C) or 3 parallel perfusion chambers (D, E). *P < 0.05; **P < 0.01; ***P < 0.001.

significantly better resolution between pancreatic alpha and delta cells. Our data sets complement a recent, high quality single cell RNAseq data set of dissociated mouse islets [25]. The single cell approach offers valuable insight into the heterogeneity of gene expression amongst the different endocrine populations in the islet and complement our transcriptomes from FACS-purified populations of alpha, beta, and delta cells that demonstrate a far greater depth and dynamic range and therefore reliably detect the expression of a far greater number of genes, as expected.

While the most prevalent transcript in alpha, beta, and delta cell transcriptomes was *Gcg*, *Ins2,* and *Sst*, respectively, the relative abundance of *Gcg* in alpha cells (27.8%) and *Ins2* in beta cells (18.9%) was much greater than the relative abundance of *Sst* in delta cells, which made up merely 3.5% of all aligned reads. As our transgenic reporter strategy is very similar for delta and alpha cells (Figure 1), there is no technical reason for delta cells to exhibit such a relatively modest gene expression of their principal endocrine signal Sst, in comparison to the expression of *Gcg* and *Ins2* in alpha and beta cells, respectively. Instead, we think that this is a direct reflection of the fact that insulin and glucagon are hormones released into the systemic circulation to act at sites throughout the body, while somatostatin is responsible for local feedback within the islet.

Our transcriptome analyses reinforce the view that delta cells are similar to beta cells. This extends from the mechanistically similar control of glucose-stimulated exocytosis in both beta and delta cells that proceeds via K_{ATP} channels, voltage-gated sodium, and calcium channels [7,36] to include GPCR control that can potentiate glucose-stimulated exocytosis in both cell types. Some of these GPCRs, such as *Crhr2* and *Ghsr* on delta cells and *Galr1* on beta cells, are selectively expressed by only one cell type. In contrast, a series of other GPCRs that are routinely associated with beta cells, including *Glp1r*, *Adra2a,* and *Gcgr*, are expressed by delta cells at likely physiologically meaningful levels. We do not think that these reflect contamination by beta cells, as many known beta cell-specific genes are indeed highly selectively detected in our beta cell transcriptomes only and do not demonstrate similar levels of expression by delta cells (Figure 1). These observations betray significant overlap in delta and beta cell transcriptional programs, possibly afforded in part by the shared expression of Pdx1. These similarities likely reflect the close ontogenetic origins of beta and delta cells and are in line with observations

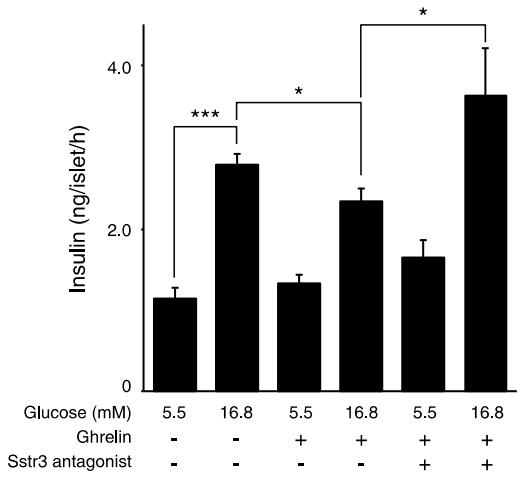


Figure 5: Attenuation of glucose-stimulated somatostatin secretion by ghrelin depends on local feedback mediated by Sst and Sst3r. Ghrelin significantly attenuates glucose-stimulated insulin secretion from wild type mouse islets *in vitro*. Co-stimulation with an Sstr3 antagonist fully prevents the insulinostatic actions of ghrelin. All peptides were applied at 100 nM final concentration. Values represent the mean ± SEM for 4 wells per group. *P < 0.05; **P < 0.01; ***P < 0.001.

that perturbation of a single transcription factor, Nkx6-1, suffices to turn beta cells into delta-like cells [37]. Their overlapping transcriptomes may contribute to the participation of beta and delta cells in a common functional unit that responds in unison to increases over resting glucose with increased exocytosis. This enables pulsatile, delta cell-mediated feedback to oscillate with insulin release, antiparallel to glucagon [38].

The transcriptomes we report here constitute an invaluable repository that informs on the expression, or lack thereof, of all genes in each of the main endocrine cell types of the islet, including delta cells. In addition to confirming that *Crhr2* is selectively expressed by delta cells [7], we discovered that the *Ghsr* gene is abundantly and selectively expressed in delta cells (Figure 2). Stimulation of mouse and human islets with ghrelin indeed leads to a robust increase in somatostatin secretion (Figure 4) and stimulates calcium responses in delta cells in intact islets (Figure 3). Furthermore, at equimolar concentrations, des-acyl-ghrelin in intact islets does not affect somatostatin secretion and does not block ghrelin-induced somatostatin release, in agreement with prior observations that des-acyl-ghrelin is a full agonist of the GHS-R1 only at very high concentrations [39] and does not block ghrelin [40].

Most studies to date have assumed that the *Ghsr* within islets is expressed by beta cells and thereby inhibits insulin release [41]. This implies that GHS-R1 in islets is not coupled to its canonical $G\alpha_{q/11}$ signaling cascade [42], as this would enhance rather than attenuate insulin secretion. Moreover, the insulinostatic effects of ghrelin are pertussis toxin-sensitive [43], indicative of the involvement of $G\alpha_i$. Our findings resolve this long-standing paradox [17] by demonstrating that ghrelin promotes somatostatin release, likely by engaging the canonical $G\alpha_{q/11}$ and Ca^{2+} cascade in delta cells, which would then inhibit insulin via beta cell somatostatin receptors coupled to $G\alpha_i$. Such a model is in good agreement with the preponderance of published data on the mechanism by which ghrelin inhibits insulin secretion, but offers a more parsimonious explanation compared to some of the prior models that have been proposed to resolve the discrepancy between ghrelin's well-accepted insulinostatic actions and the fact that its canonical signaling cascade would promote, not inhibit, exocytosis.

It has been suggested that *Ghsr* is expressed by alpha cells and promotes glucagon release [44]. This was based on *Ghsr* detection by in situ hybridization at the islet periphery, where alpha cells are intimately associated with delta cells [44], experiments on the alpha-TC-1 alpha cell line in which expression of *Ghsr* need not imply similar *Ghsr* expression by primary alpha cells, and studies on isolated islets under 12 mM glucose, a concentration not typically associated with highest alpha cell activity. More importantly, increases in glucagon stimulate insulin [45,46] and, therefore, cannot explain the insulinostatic actions of ghrelin that are generally accepted [12–16]. Hetero-dimerization between GHS-R1 and SSTR5 on the beta cell surface has been offered to explain how ghrelin inhibits insulin release in a $G\alpha_i$-dependent manner [47]. The fact that *Ghsr* is expressed by delta cells and that *Sstr5* is not detected in any mouse islet endocrine cell (Figure 2) makes this model untenable. More recently, coupling of the GHS-1R with transient receptor potential melastatin 2 (TRPM2) was suggested as an alternate explanation to reconcile ghrelin's insulin inhibition with $G\alpha_{q/11}$ and phospholipase C-mediated increases in Ca^{2+} [48]. Attenuation by ghrelin of non-selective glucose-induced cation currents via TRPM2 that promote insulin release led to a model in which TRPM2 was placed downstream of Ghrelin/GHSR-1R on beta cells [48]. The fact that we demonstrate that *Ghsr* is expressed exclusively by delta not by beta cells would not alter the hierarchy of Trmp2 downstream of ghrelin/GHS-1R but would merely insert somatostatin as an intermediate factor between ghrelin and Trpm2.

Our observation that beta cells do not express *Ghsr* is in agreement with a prior report that detected no *Ghsr* on HIT-T15 beta cells [49]. This study observed that ghrelin promotes beta cell proliferation and prevents beta cell apoptosis. Since these effects were also observed in response to des-acyl-ghrelin and depended on adenyl cyclase and protein kinase A, it was suggested that they are mediated via an as yet unidentified receptor distinct from GHS-R1 [49]. This is certainly an intriguing possibility, and we do not rule out the presence of an alternate receptor with high affinity for ghrelin and des-acyl-ghrelin in the islet. If an alternate ghrelin receptor indeed exists within the islet, this suggests the possibility of secondary effects of ghrelin stimulation that are mediated indirectly via local intermediates. Nevertheless, we report here that the actions of ghrelin (and not des-acyl-ghrelin) are in full agreement with the selective activation of delta cell GHS-R1 by ghrelin. There has been a prior report that suggested ghrelin had no effect on basal somatostatin release but modestly inhibited arginine-induced somatostatin secretion in perfused rat pancreas [50]. We are unable to offer an explanation that reconciles this observation with our current data.

We think that many of the discrepancies with regard to the mechanistic understanding of ghrelin's insulinostatic actions can be traced to observations that place Ghsr expression on alpha or beta cell lines. As the canonical signaling cascade in response to ghrelin would promote exocytosis, stimulation of insulin (from beta cell lines) or glucagon (from alpha cell lines) in response to ghrelin is the default response. However, caution should be exercised when extrapolating these observations on cell lines to the primary endocrine cells they resemble most closely. All islet endocrine cells derive from a common lineage. Key characteristics of primary islet endocrine cells are often lost in cell lines, with traits normally restricted to one particular cell type reappearing in cell lines that are thought to resemble another islet endocrine cell. The transcriptomes we report here constitute a valuable

repository to validate the cell type that expresses any gene that is detectably expressed within pancreatic islets.

5. CONCLUSIONS

We contribute significantly to the understanding of the physiological role of ghrelin by offering an alternate mechanistic explanation how ghrelin achieves its well-known insulinostatic actions, which is relevant given ghrelin's central role in metabolism [17]. Moreover, our observations underscore that delta cells play an underappreciated but important physiological role in controlling insulin output by local feedback within pancreatic islets [7]. The high quality transcriptomes from delta, beta, and alpha cells that were instrumental in this discovery will be invaluable to unravel the complex crosstalk that tightly balances insulin and glucagon release from healthy islets and breaks down in diabetes.

FUNDING

This research was supported by by a Career Development Award from the Juvenile Diabetes Research Foundation (2-2013-54) and an Individual Biomedical Research Award from the Hartwell Foundation (201500731), both to MOH.

AUTHOR CONTRIBUTIONS

MRD conducted functional imaging on intact islets and contributed to writing and discussions. AM and CC-Z conducted all bioinformatics analyses and contributed to discussions. CJD and JV measured somatostatin secretion, generated tracers and and maintained the RIA for somatostatin. GN conducted immunofluorescence and FISH experiments. TvdM conducted experiments and contributed to the writing, editing and discussions. MOH designed the study and directed the project, generated delta, beta and alpha cell transcriptomes, conducted static and dynamic somatostatin secretion experiments and wrote the paper.

ACKNOWLEDGMENTS

The authors thank the members of the Huising lab for constructive input and discussion. Human islets were obtained through the NIDDK-supported Integrated Islet Distribution Program (IIDP). Dr. Michael Beyermann and Jean Rivier (Salk Institute for Biological Studies) are gratefully acknowledged for providing ghrelin and des-acyl-ghrelin.

CONFLICT OF INTEREST

The authors have no conflicts to declare.

APPENDIX A. SUPPLEMENTARY DATA

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.molmet.2016.04.007.

REFERENCES

[1] Caicedo, A., 2013. Paracrine and autocrine interactions in the human islet: more than meets the eye. Seminars in Cell & Developmental Biology 24(1): 11–21.

[2] Yang, Y.H., Szabat, M., Bragagnini, C., Kott, K., Helgason, C.D., Hoffman, B.G., et al., 2011. Paracrine signalling loops in adult human and mouse pancreatic islets: netrins modulate beta cell apoptosis signalling via dependence receptors. Diabetologia 54(4):828–842.

[3] Amisten, S., Salehi, A., Rorsman, P., Jones, P.M., Persaud, S.J., 2013. An atlas and functional analysis of G-protein coupled receptors in human islets of Langerhans. Pharmacology & Therapeutics 139(3):359–391.

[4] Hauge-Evans, A.C., King, A.J., Carmignac, D., Richardson, C.C., Robinson, I.C., Low, M.J., et al., 2009. Somatostatin secreted by islet delta-cells fulfills multiple roles as a paracrine regulator of islet function. Diabetes 58(2):403–411.

[5] Taborsky Jr., G.J., Smith, P.H., Porte Jr., D., 1978. Interaction of somatostatin with the A and B cells of the endocrine pancreas. Metabolism 27(9 Suppl. 1): 1299–1302.

[6] Brown, M., Rivier, J., Vale, W., 1976. Biological activity of somatostatin and somatostatin analogs on inhibtion of arginine-induced insulin and glucagon release in the rat. Endocrinology 98(2):336–343.

[7] van der Meulen, T., Donaldson, C.J., Caceres, E., Hunter, A.E., Cowing-Zitron, C., Pound, L.D., et al., 2015. Urocortin3 mediates somatostatin-dependent negative feedback control of insulin secretion. Nature Medicine 21(7):769–776.

[8] van der Meulen, T., Xie, R., Kelly, O.G., Vale, W.W., Sander, M., Huising, M.O., 2012. Urocortin 3 marks mature human primary and embryonic stem cell-derived pancreatic alpha and beta cells. PLoS One 7(12):e52181.

[9] Blum, B., Hrvatin, S.S., Schuetz, C., Bonal, C., Rezania, A., Melton, D.A., 2012. Functional beta-cell maturation is marked by an increased glucose threshold and by expression of urocortin 3. Nature Biotechnology 30(3):261–264.

[10] Rozzo, A., Meneghel-Rozzo, T., Delakorda, S.L., Yang, S.B., Rupnik, M., 2009. Exocytosis of insulin: in vivo maturation of mouse endocrine pancreas. Annals of the New York Academy of Sciences 1152:53–62.

[11] Benner, C., van der Meulen, T., Caceres, E., Tigyi, K., Donaldson, C.J., Huising, M.O., 2014. The transcriptional landscape of mouse beta cells compared to human beta cells reveals notable species differences in long non-coding RNA and protein-coding gene expression. BMC Genomics 15(1):620.

[12] Dezaki, K., Hosoda, H., Kakei, M., Hashiguchi, S., Watanabe, M., Kangawa, K., et al., 2004. Endogenous ghrelin in pancreatic islets restricts insulin release by attenuating Ca2+ signaling in beta-cells: implication in the glycemic control in rodents. Diabetes 53(12):3142–3151.

[13] Dezaki, K., Damdindorj, B., Sone, H., Dyachok, O., Tengholm, A., Gylfe, E., et al., 2011. Ghrelin attenuates cAMP-PKA signaling to evoke insulinostatic cascade in islet beta-cells. Diabetes 60(9):2315–2324.

[14] Reimer, M.K., Pacini, G., Ahren, B., 2003. Dose-dependent inhibition by ghrelin of insulin secretion in the mouse. Endocrinology 144(3):916–921.

[15] Tong, J., Prigeon, R.L., Davis, H.W., Bidlingmaier, M., Kahn, S.E., Cummings, D.E., et al., 2010. Ghrelin suppresses glucose-stimulated insulin secretion and deteriorates glucose tolerance in healthy humans. Diabetes 59(9):2145–2151.

[16] Yada, T., Damdindorj, B., Rita, R.S., Kurashina, T., Ando, A., Taguchi, M., et al., 2014. Ghrelin signalling in beta-cells regulates insulin secretion and blood glucose. Diabetes, Obesity and Metabolism 16(Suppl. 1):111–117.

[17] Muller, T.D., Nogueiras, R., Andermann, M.L., Andrews, Z.B., Anker, S.D., Argente, J., et al., 2015. Ghrelin. Molecular Metabolism 4(6):437–460.

[18] Huising, M.O., van der Meulen, T., Vaughan, J.M., Matsumoto, M., Donaldson, C.J., Park, H., et al., 2010. CRFR1 is expressed on pancreatic beta cells, promotes beta cell proliferation, and potentiates insulin secretion in a glucose-dependent manner. Proceedings of the National Academy of Sciences U S A 107(2):912–917.

[19] Srinivas, S., Watanabe, T., Lin, C.S., William, C.M., Tanabe, Y., Jessell, T.M., et al., 2001. Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. BMC Developmental Biology 1:4.

[20] Taniguchi, H., He, M., Wu, P., Kim, S., Paik, R., Sugino, K., et al., 2011. A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. Neuron 71(6):995—1013.
[21] Herrera, P.L., 2000. Adult insulin- and glucagon-producing cells differentiate from two independent cell lineages. Development 127(11):2317—2322.
[22] Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29(1):15—21.
[23] Liao, Y., Smyth, G.K., Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30(7):923—930.
[24] Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26(1):139—140.
[25] Xin, Y., Kim, J., Ni, M., Wei, Y., Okamoto, H., Lee, J., et al., 2016. Use of the Fluidigm C1 platform for RNA sequencing of single mouse pancreatic islet cells. Proceedings of the National Academy of Sciences U S A 113(12):3293—3298.
[26] Brazeau, P., Vale, W., Burgus, R., Ling, N., Butcher, M., Rivier, J., et al., 1973. Hypothalamic polypeptide that inhibits the secretion of immunoreactive pituitary growth hormone. Science 179(4068):77—79.
[27] Vale, W., Rivier, J., Ling, N., Brown, M., 1978. Biologic and immunologic activities and applications of somatostatin analogs. Metabolism 27(9 Suppl. 1): 1391—1401.
[28] Reubi, J.C., Schaer, J.C., Wenger, S., Hoeger, C., Erchegyi, J., Waser, B., et al., 2000. SST3-selective potent peptidic somatostatin receptor antagonists. Proceedings of the National Academy of Sciences U S A 97(25):13973—13978.
[29] Mulder, H., Lindh, A.C., Sundler, F., 1993. Islet amyloid polypeptide gene expression in the endocrine pancreas of the rat: a combined in situ hybridization and immunocytochemical study. Cell and Tissue Research 274(3): 467—474.
[30] Artner, I., Hang, Y., Mazur, M., Yamamoto, T., Guo, M., Lindner, J., et al., 2010. MafA and MafB regulate genes critical to beta-cells in a unique temporal manner. Diabetes 59(10):2530—2539.
[31] Zhang, J., McKenna, L.B., Bogue, C.W., Kaestner, K.H., 2014. The diabetes gene Hhex maintains delta-cell differentiation and islet function. Genes & Development 28(8):829—834.
[32] Petri, A., Ahnfelt-Ronne, J., Frederiksen, K.S., Edwards, D.G., Madsen, D., Serup, P., et al., 2006. The effect of neurogenin3 deficiency on pancreatic gene expression in embryonic mice. Journal of Molecular Endocrinology 37(2): 301—316.
[33] Su, Y., Jono, H., Misumi, Y., Senokuchi, T., Guo, J., Ueda, M., et al., 2012. Novel function of transthyretin in pancreatic alpha cells. FEBS Letters 586(23): 4215—4222.
[34] Alexander, S.P., Mathie, A., Peters, J.A., 2011. Guide to receptors and channels (GRAC), 5th edition. British Journal of Pharmacology 164(Suppl. 1): S1—S324.
[35] Chen, T.W., Wardill, T.J., Sun, Y., Pulver, S.R., Renninger, S.L., Baohan, A., et al., 2013. Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature 499(7458):295—300.
[36] Braun, M., Ramracheya, R., Amisten, S., Bengtsson, M., Moritoh, Y., Zhang, Q., et al., 2009. Somatostatin release, electrical activity, membrane currents and exocytosis in human pancreatic delta cells. Diabetologia 52(8): 1566—1578.
[37] Schaffer, A.E., Taylor, B.L., Benthuysen, J.R., Liu, J., Thorel, F., Yuan, W., et al., 2013. Nkx6.1 controls a gene regulatory network required for establishing and maintaining pancreatic Beta cell identity. PLoS Genetics 9(1): e1003274.
[38] Salehi, A., Qader, S.S., Grapengiesser, E., Hellman, B., 2007. Pulses of somatostatin release are slightly delayed compared with insulin and anti-synchronous to glucagon. Regulatory Peptides 144(1—3):43—49.
[39] Gauna, C., van de Zande, B., van Kerkwijk, A., Themmen, A.P., van der Lely, A.J., Delhanty, P.J., 2007. Unacylated ghrelin is not a functional antagonist but a full agonist of the type 1a growth hormone secretagogue receptor (GHS-R). Molecular and Cellular Endocrinology 274(1—2):30—34.
[40] Tong, J., Davis, H.W., Summer, S., Benoit, S.C., Haque, A., Bidlingmaier, M., et al., 2014. Acute administration of unacylated ghrelin has no effect on basal or stimulated insulin secretion in healthy humans. Diabetes 63(7):2309—2319.
[41] Kageyama, H., Funahashi, H., Hirayama, M., Takenoya, F., Kita, T., Kato, S., et al., 2005. Morphological analysis of ghrelin and its receptor distribution in the rat pancreas. Regulatory Peptides 126(1—2):67—71.
[42] Howard, A.D., Feighner, S.D., Cully, D.F., Arena, J.P., Liberator, P.A., Rosenblum, C.I., et al., 1996. A receptor in pituitary and hypothalamus that functions in growth hormone release. Science 273(5277):974—977.
[43] Dezaki, K., Kakei, M., Yada, T., 2007. Ghrelin uses Galphai2 and activates voltage-dependent K+ channels to attenuate glucose-induced Ca2+ signaling and insulin release in islet beta-cells: novel signal transduction of ghrelin. Diabetes 56(9):2319—2327.
[44] Chuang, J.C., Sakata, I., Kohno, D., Perello, M., Osborne-Lawrence, S., Repa, J.J., et al., 2011. Ghrelin directly stimulates glucagon secretion from pancreatic alpha-cells. Molecular Endocrinology 25(9):1600—1611.
[45] Samols, E., Marri, G., Marks, V., 1965. Promotion of insulin secretion by glucagon. Lancet 2(7409):415—416.
[46] Samols, E., Marri, G., Marks, V., 1966. Interrelationship of glucagon, insulin and glucose. The insulinogenic effect of glucagon. Diabetes 15(12):855—866.
[47] Park, S., Jiang, H., Zhang, H., Smith, R.G., 2012. Modification of ghrelin receptor signaling by somatostatin receptor-5 regulates insulin release. Proceedings of the National Academy of Sciences U S A 109(46):19003—19008.
[48] Kurashina, T., Dezaki, K., Yoshida, M., Sukma Rita, R., Ito, K., Taguchi, M., et al., 2015. The beta-cell GHSR and downstream cAMP/TRPM2 signaling account for insulinostatic and glycemic effects of ghrelin. Scientific Reports 5: 14041.
[49] Granata, R., Settanni, F., Biancone, L., Trovato, L., Nano, R., Bertuzzi, F., et al., 2007. Acylated and unacylated ghrelin promote proliferation and inhibit apoptosis of pancreatic beta-cells and human islets: involvement of 3′,5′-cyclic adenosine monophosphate/protein kinase A, extracellular signal-regulated kinase 1/2, and phosphatidyl inositol 3-Kinase/Akt signaling. Endocrinology 148(2):512—529.
[50] Egido, E.M., Rodriguez-Gallardo, J., Silvestre, R.A., Marco, J., 2002. Inhibitory effect of ghrelin on insulin and pancreatic somatostatin secretion. European Journal of Endocrinology 146(2):241—244.

III. Chapter 3: RoiViz: an R package for visualizing ROI intensity from GCAMP6 calcium imaging

Alex M. Mawla & Mark O. Huising

Pre-print on BiorXiv. Advanced draft in process of submission.

Contributions to Jointly Authored Works: As first author of this manuscript, I created and programmed this approach, developed the R package, and wrote the manuscript. This was intended to act as a simple and quick method to appropriately translate and then visualize calcium-imaging data produced in the lab of Dr. Mark O. Huising.

Significance of Research: Currently, this package has been used to generate and present visuals in at least five well-received papers, as well as in conference talks and poster presentations. This is intended to act as a short manuscript to share the application of this package so that it may be used by other researchers who utilize calcium imaging data.

RoiViz: an R package for visualizing ROI intensity from GCAMP6 calcium imaging

Alex M. Mawla[1], Mark O. Huising[1,2*]

[1]Department of Neurobiology, Physiology & Behavior, College of Biological Sciences, University of California, Davis, CA 95616, USA

[2]Department of Physiology and Membrane Biology, School of Medicine, University of California, Davis, CA 95616, USA

* Corresponding author:

Dr. Mark O. Huising, Ph.D.

Professor

University of California, Davis

Department of Neurobiology, Physiology & Behavior, College of Biological Sciences

Department of Physiology and Membrane Biology, School of Medicine

One Shields Avenue,

Davis, CA, 95616

Ph: (530) 752-4670

Email: mhuising@ucdavis.edu

Abstract

Summary: RoiViz is an R package designed to seamlessly visualize region of interest (ROI) intensity data derived from utilizing reporter line signaling, such as GCAMP6 calcium imaging acquired using the genetically encoded calcium reporter GCaMP6 [1]. Its sole purpose is to contextualize tabular data of relative intensity values as a function of time, and generates high-resolution, publication-quality eps files that can be saved in a portable document format or edited further as needed using off the shelf graphics software.

Availability and Implementation: RoiViz is released under Artistic-2.0 License. The source code and documents are freely available through Github (https://github.com/Huising-Lab/RoiViz).

Contact: mhuising@ucdavis.edu or ammawla@ucdavis.edu

Supplementary information: Supplementary data are available at *BiorXiv* online or *Github.*

Introduction

Microscopy coupled with the use of genetically-encoded or dye-based fluorescent reporter lines for imaging is a widely-used approach to ascertain the functional profile of different cell types over time and in response to different stimuli. In the context of pancreatic endocrine islet biology, where RoiViz was originally developed, such approaches can be used to reflect patterns of beta, alpha, or delta cells response to glucose, various hormones, or exogeneous drugs. The functional profiling of these cells is intricate and is characterized by beta cells some respond in gap junction-coordinated synchronous responses, while others respond in short, asynchronous bursts. Please see http://huisinglab.com/wp-content/movies/Alpha_to_beta_calcium.mp4 or supplemental data for a visual illustration.

Calcium signaling as a secondary messenger acts as a useful proxy to identify changes in cell function; for example, in pancreatic islet beta cells, calcium is required for insulin release [2-4]. GCAMP6 [1] is a widely used genetically encoded calcium indicator through the combination of calmodulin, green fluorescence protein (GFP), and M13 – a myosin light-chain kinase peptide sequence. Its use in transgenic lines allows for high-resolution tracking of a cell's intracellular calcium response [5, 6]. This can be finely tuned by inserting GCAMP6 under promoter control, to define cell-specificity [7]. In the context of pancreatic islet biology, for instance, floxed alleles of GCaMP6 are now routinely used in conjunction with cell type-specific Cre driver strains to direct the expression of GCaMP6 specifically in the islet cell type of interest [8-11]

Data such as this has been generated to elucidate the behavior of neuronal cells and complex neuronal circuitry [5, 12-15], cardiac muscle [16], or G protein-coupled receptor (GPCR) activation via olfactory responses [17]. Our ability to track and quantify the calcium behavior across populations of cells is rapidly increasing, and it is now entirely feasible to visualize calcium responses of hundreds or even thousands of cells over time. However, while tabular data can be utilized in diverse ways to track behavior of cells, a gap exists in visualizing these data in a succinct manner without a loss of the single cell resolution

at which the data were originally acquired. We developed RoiViz to visualize calcium responses quickly and seamlessly across populations of cells with minimal programming skills required. To date, RoiViz has been used and optimized in the generation of high-resolution, publication quality visuals in several papers [8, 11, 18-20]. Here, we provide this tool as an open-source package available on GitHub, and for later availability on the R BioConductor repository for ease of use and access. The tool was developed to visualize calcium responses detected using GCaMP6. This tool should also be usable with data acquired using calcium dyes or dyes or genetically encoded biosensors that report on changes in the concentration of other metabolites or second messengers, provided that the data demonstrate a good signal to noise ratio.

Functions

RoiViz takes in a user-provided comma-separated values (csv) file containing intensity imaging data, originally designed for use with GCAMP6 calcium responses. The tool is designed to interpret data providing each cell type represented as a column, as a function of time across rows. The tool is quick and easy to implement, using a function of the same name - *RoiViz*, which will quickly generate three, high-resolution publication quality encapsulated postscript (eps) files in three main color schemes [green, grey scale, and black inverted]. The tool will also cluster cells of similar behavior, accepting different methods of association, such as Euclidean distance. The user can then track individual cells, labeled by column name, and add annotated data of substance exposure across time using commercially available third-party graphics software packages.

This tool was designed to allow visualization of biological relevance of the same, or combined, cell types' responses to exogenous substance exposure, without the need for programming skills, and to generate high-quality images for use later.

Results

RoiViz is developed as an R package to be made available through BioConductor [21], and is available under Artistic-2.0 License. RoiViz is designed to take in tabular data of intensity values derived

from microscopy tracking reporter line fluorescence, such as GCAMP6 calcium imaging, and generate visualizations of cell behavior in response to exogenous substance exposure. RoiViz has routinely been used in previously cited research and is now available in open source. Finally, RoiViz is easy to use, with a full walkthrough with sample data accessible through its companion vignette.

Funding

This work was supported by grants from the National Institutes of Health (NIDDK110276) the Juvenile Diabetes Research Foundation (2-SRA-2021-1054-M-N) and the American Diabetes Association (1-19-IBS-078) to M.O.H. A.M.M. was supported by the Stephen F. and Bettina A. Sims Immunology Fellowship and the Amazon AWS Machine Learning Research Fellowship.

Conflict of interest

The authors have no conflicts of interests to declare.

References

1. Nakai J, Ohkura M, Imoto K: **A high signal-to-noise Ca(2+) probe composed of a single green fluorescent protein**. *Nat Biotechnol* 2001, **19**(2):137-141.
2. Klec C, Ziomek G, Pichler M, Malli R, Graier WF: **Calcium Signaling in Beta-cell Physiology and Pathology: A Revisit**. *Int J Mol Sci* 2019, **20**(24).
3. Baylor SM, Hollingworth S: **Calcium indicators and calcium signalling in skeletal muscle fibres during excitation-contraction coupling**. *Prog Biophys Mol Biol* 2011, **105**(3):162-179.
4. Clapham DE: **Calcium signaling**. *Cell* 2007, **131**(6):1047-1058.
5. Chen T-W, Wardill TJ, Sun Y, Pulver SR, Renninger SL, Baohan A, Schreiter ER, Kerr RA, Orger MB, Jayaraman V *et al*: **Ultrasensitive fluorescent proteins for imaging neuronal activity**. *Nature* 2013, **499**(7458):295-300.
6. Barnett LM, Hughes TE, Drobizhev M: **Deciphering the molecular mechanism responsible for GCaMP6m's Ca2+-dependent change in fluorescence**. *PloS one* 2017, **12**(2):e0170934.
7. Ji G, Feldman ME, Deng KY, Greene KS, Wilson J, Lee JC, Johnston RC, Rishniw M, Tallini Y, Zhang J *et al*: **Ca2+-sensing transgenic mice: postsynaptic signaling in smooth muscle**. *J Biol Chem* 2004, **279**(20):21461-21468.
8. Croze ML, Flisher MF, Guillaume A, Tremblay C, Noguchi GM, Granziera S, Vivot K, Castillo VC, Campbell SA, Ghislain J *et al*: **Free fatty acid receptor 4 inhibitory signaling in delta cells regulates islet hormone secretion in mice**. *Mol Metab* 2021, **45**:101166.
9. Mallo M: **Controlled gene activation and inactivation in the mouse**. *Front Biosci* 2006, **11**:313-327.
10. Srinivas S, Watanabe T, Lin CS, William CM, Tanabe Y, Jessell TM, Costantini F: **Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus**. *BMC Dev Biol* 2001, **1**:4.
11. van der Meulen T, Mawla AM, DiGruccio MR, Adams MW, Nies V, Dolleman S, Liu S, Ackermann AM, Caceres E, Hunter AE *et al*: **Virgin Beta Cells Persist throughout Life at a Neogenic Niche within Pancreatic Islets**. *Cell Metab* 2017, **25**(4):911-926 e916.
12. Akerboom J, Chen TW, Wardill TJ, Tian L, Marvin JS, Mutlu S, Calderon NC, Esposti F, Borghuis BG, Sun XR *et al*: **Optimization of a GCaMP calcium indicator for neural activity imaging**. *J Neurosci* 2012, **32**(40):13819-13840.
13. Dana H, Mohar B, Sun Y, Narayan S, Gordus A, Hasseman JP, Tsegaye G, Holt GT, Hu A, Walpita D *et al*: **Sensitive red protein calcium indicators for imaging neural activity**. *eLife* 2016, **5**.
14. Madisen L, Garner AR, Shimaoka D, Chuong AS, Klapoetke NC, Li L, van der Bourg A, Niino Y, Egolf L, Monetti C *et al*: **Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance**. *Neuron* 2015, **85**(5):942-958.
15. Tian L, Hires SA, Mao T, Huber D, Chiappe ME, Chalasani SH, Petreanu L, Akerboom J, McKinney SA, Schreiter ER *et al*: **Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators**. *Nat Methods* 2009, **6**(12):875-881.
16. Tallini YN, Ohkura M, Choi BR, Ji G, Imoto K, Doran R, Lee J, Plan P, Wilson J, Xin HB *et al*: **Imaging cellular signals in the heart in vivo: Cardiac expression of the high-signal Ca2+ indicator GCaMP2**. *Proc Natl Acad Sci U S A* 2006, **103**(12):4753-4758.
17. Greer PL, Bear DM, Lassance JM, Bloom ML, Tsukahara T, Pashkovski SL, Masuda FK, Nowlan AC, Kirchner R, Hoekstra HE *et al*: **A Family of non-GPCR Chemosensors Defines an Alternative Logic for Mammalian Olfaction**. *Cell* 2016, **165**(7):1734-1748.
18. DiGruccio MR, Mawla AM, Donaldson CJ, Noguchi GM, Vaughan J, Cowing-Zitron C, van der Meulen T, Huising MO: **Comprehensive alpha, beta and delta cell transcriptomes reveal that**

ghrelin selectively activates delta cells and promotes somatostatin release from pancreatic islets. *Mol Metab* 2016, **5**(7):449-458.

19. Huang JL, Lee S, Hoek P, van der Meulen T, Van R, Huising MO: **Genetic deletion of Urocortin 3 does not prevent functional maturation of beta cells**. *J Endocrinol* 2020, **246**(1):69-78.
20. Huising MO, van der Meulen T, Huang JL, Pourhosseinzadeh MS, Noguchi GM: **The Difference delta-Cells Make in Glucose Control**. *Physiology (Bethesda)* 2018, **33**(6):403-411.
21. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J *et al*: **Bioconductor: open software development for computational biology and bioinformatics**. *Genome Biol* 2004, **5**(10):R80.

IV. Chapter 4: Virgin Beta Cells Persist throughout Life at a Neogenic Niche within Pancreatic Islets

Talitha van der Meulen, **Alex M. Mawla**, Michael R. DiGruccio, Michael W. Adams, Vera Nies, Sophie Dólleman, Siming Liu, Amanda M. Ackermann, Elena Cáceres, Anna E. Hunter, Klaus H. Kaestner, Cynthia J. Donaldson, Mark O. Huising

Cell Metabolism, 2017. 25(4):911-926. doi: 10.1016/j.cmet.2017.03.017

Contributions to Jointly Authored Works: As second author of this manuscript, I was responsible for the entirety of formal analysis and interpretation of all genomic data. I analyzed all RNA-Seq data for alpha, beta, delta, and transdifferentiated cell transcriptomes, alongside virgin and mature beta cell transcriptomes, and a separate developmental series experiment. I was responsible for the exploration of these results, their interpretation, and identified meaningful biological pathways, which played an integral role in explaining the inability of virgin beta cells to appropriately sense and respond to glucose. I was responsible for panels translating the analysis into meaningful biology in Figures 2, 3, 6, S3, S4, and S6. I was also responsible for creating and maintaining browser plot resources through our lab website for other colleagues in the field to utilize to visualize our findings, and to test theirs quickly.

Significance of Research: This paper described and characterized the novel discovery of a subpopulation of pancreatic beta cells that are characterized by a lack of maturity markers. These cells are functionally immature and suggest an avenue to understanding what could be a source of beta cell regeneration. This manuscript received significant coverage by outside news outlets, with an Altmetric score of 183. This manuscript was also selected as 'best of 2017' by the editors of Cell Metabolism. Finally, this manuscript was responsible for the basis of rewarding my professor, Dr. Mark O. Huising, with the 2017-2018 University of California, Davis College of Biological Science Faculty Research Award.

Citations: 130

Published in final edited form as:

Cell Metab. 2017 April 04; 25(4): 911–926.e6. doi:10.1016/j.cmet.2017.03.017.

Virgin Beta Cells Persist throughout Life at a Neogenic Niche within Pancreatic Islets

Talitha van der Meulen[1], **Alex Mawla**[1], **Michael R. DiGruccio**[1], **Michael W. Adams**[2], **Vera Nies**[3], **Sophie Dólleman**[3], **Siming Liu**[1], **Amanda M. Ackermann**[4], **Elena Cáceres**[3], **Anna E. Hunter**[1], **Klaus H. Kaestner**[5], **Cynthia J. Donaldson**[3], **Mark O. Huising**[1,6,*]

[1]Department of Neurobiology, Physiology & Behavior, College of Biological Sciences, University of California, Davis, CA 95616, USA

[2]Waitt Advanced Biophotonics Center, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA

[3]Clayton Foundation Laboratories for Peptide Biology, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA

[4]Division of Endocrinology and Diabetes, The Children's Hospital of Philadelphia, 3400 Civic Center Boulevard, Philadelphia, PA, 19104, USA

[5]Department of Genetics and Institute for Diabetes, Obesity, and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

[6]Department of Physiology and Membrane Biology, School of Medicine, University of California, Davis, CA 95616, USA

Summary

Postnatal maintenance or regeneration of pancreatic beta cells is considered to occur exclusively via the replication of existing beta cells, but clinically meaningful restoration of human beta cell mass by proliferation has never been achieved. We discovered a population of immature beta cells that is present throughout life and forms from non-beta precursors at a specialized micro-environment or 'neogenic niche' at the islet periphery. These cells express insulin, but lack other key beta cell markers, are transcriptionally immature, incapable of sensing glucose and unable to support calcium influx. They constitute an intermediate stage in the transdifferentiation of alpha cells to cells that are functionally indistinguishable from conventional beta cells. We thus identified a lifelong source of new beta cells at a specialized site within healthy islets. By comparing co-existing immature and mature beta cells within healthy islets we stand to learn how to mature insulin-expressing cells into functional beta cells.

*Lead Contact: mhuising@ucdavis.edu.

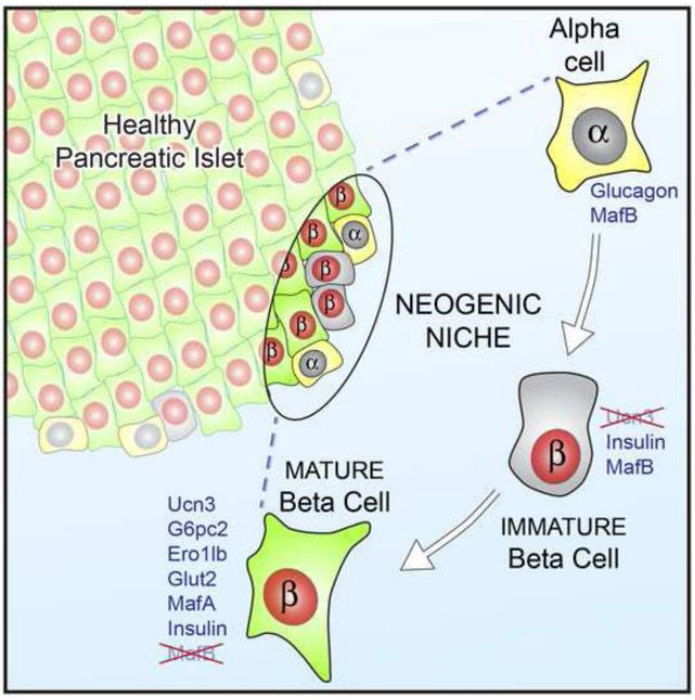
Author Contributions

Conceptualization, TvdM, MOH; Methodology, MWA, MOH; Software, MWA; Formal Analysis and Data Curation, AM, TvdM, MOH; Investigation, TvdM, MRD, VN, SD, SL, EC, AEH, AMA, CJD, MOH; Resources, AMA, KHK; Writing, TvdM, MOH; Review and Editing, TvdM, MRD, KHK, CJD, MOH; Visualization, MOH; Supervision, Project Administration and Funding Acquisition, MOH.

Accession numbers

The sequencing data associated with this publication has been deposited in GEO under GSE88778, GSE88779 and GSE90766.

Graphical Abstract



Keywords

beta cell maturation; transdifferentiation; beta cell neogenesis; Urocortin3; islet architecture; stem cell; diabetes; GCaMP6; neogenic niche; alpha cell

Introduction

The insulin deficiency that characterizes Type 1 Diabetes (T1D) and Type 2 Diabetes (T2D) has led to strong interest in processes that control beta cell mass. The current view is that beta cell mass is determined by the net effect of islet neogenesis, beta cell proliferation and hyperplasia, balanced by dedifferentiation and beta cell death through apoptosis (Bonner-Weir et al., 2010). Mice increase beta cell mass by self-replication to compensate for increased metabolic demand in the context of obesity or pregnancy (Cox et al., 2016; Parsons et al., 1992), although beta cell proliferation rates decline sharply with age (Brennand et al., 2007; Teta et al., 2005). In humans, increases in beta cell mass in response to similar conditions are modest at best. The mechanisms responsible for this increase are not fully understood. They have long been considered less reliant on beta cell replication (Butler et al., 2010), although standard replication markers may underestimate human beta cell proliferation rates during post-mortem conditions (Sullivan et al., 2015). Regardless, the

existence of progenitors within the pancreas that support the regeneration of functional beta cells is a highly relevant - but controversial - topic in diabetes.

Several recent reports demonstrate that forced expression of lineage-inappropriate transcription factors causes islet cells to switch identity (Collombat et al., 2007; Collombat et al., 2009; Gao et al., 2014; Papizan et al., 2011; van der Meulen and Huising, 2015), which has been attributed to the similar chromatin states of alpha and beta cells (Bramswig et al., 2013). This establishes that beta cells can arise from non-beta endocrine cells in the islet via direct transdifferentiation. Indeed, near-complete ablation of pre-existing beta cells is eventually followed by the restoration of beta cell mass via transdifferentiation of non-beta endocrine cells in mice (Chera et al., 2014; Thorel et al., 2010). These studies provided important proof of principle that insulin independence can be regained by transdifferentiation, but have also led to the notion that it is triggered by severe beta cell ablation and the associated pancreas remodeling (Habener and Stanojevic, 2012).

Instead, we demonstrate here that the islet periphery contains a 'neogenic niche'; a privileged microenvironment that supports lifelong conversion between alpha and beta cells. This is evident from the presence of a distinct population of immature beta cells at the islet periphery that does not yet express the late maturation marker Urocortin3 (Ucn3) (Blum et al., 2012; van der Meulen and Huising, 2014; van der Meulen et al., 2012). Similar Ucn3-negative beta cells are readily identifiable in human pancreas of different ages and in donors with T1D. Ucn3 negative beta cells are transcriptionally immature, lack cell surface Glut2, cannot sense glucose, and do not support calcium influx in response to depolarization. By lineage tracing we demonstrate that these Ucn3-negative beta cells are 'virgin' beta cells that represent an intermediate stage in the transdifferentiation of alpha cells into mature beta cells. On the basis of their transcriptional signature and functional responses within intact islets, we demonstrate that beta cells that arise from alpha cells at the islet edge are functionally indistinguishable from conventional beta cells. Conversely, we also observe mature beta cells that adopt an alpha cell fate and function at the islet edge. We propose that this ongoing plasticity within the neogenic niche can be targeted to regenerate beta cells.

Results

Ucn3-negative immature beta cells persist throughout life

Ucn3 is a late maturation marker for primary and stem cell-derived beta cells (Blum et al., 2012; van der Meulen et al., 2012) that continued to increase progressively weeks after beta cell expression of Nkx6–1 and Mafa reached steady state, and coincided with the gradual loss of Mafb from beta cells (Figure 1A) (Artner et al., 2010). However, Ucn3-negative beta cells persisted in islets of adult mice (Figure 1B). The fraction of beta cells that did not yet co-express Ucn3 comprised 15% of all beta cells at postnatal day 2 before stabilizing at approximately 1 – 2 % of all beta cells from 3 weeks to 14 months of age (Figure 1C). Ucn3-negative beta cells did not reflect proliferating beta cells that had transiently down regulated Ucn3, as Ki67-postive beta cells continued to stain for Ucn3 (Figure 1D).

Immature beta cells are located at the islet edge

A striking aspect unique to Ucn3-negative immature beta cells was their frequent proximity to the islet edge. We therefore quantified the relative position of approximately 17,000 individual islet cells across hundreds of islets from 23 individual animals across a range of ages, normalized for islet size (Figure 1E). Identity of each cell was assigned manually. Most cells were either alpha or mature beta cells (Figure 1F, G). We reserved the designation of immaturity for only those insulin+ cells where no Ucn3 was detected (Figure 1H) and separately quantified the distribution of insulin+ beta cells with weak, but detectable, Ucn3 (Ucn3-low beta cells). We reasoned that the latter cells may reflect maturing beta cells, or may have resulted from the inherent limitations in quantifying gradual increases in signal intensities by antibody-based methods and concluded that the prudent approach was to account for these cells separately (Figure 1I). We then compared the normalized cumulative distribution of each population (Figure 1J). Mature beta and alpha cells distributed preferentially at the islet core and periphery, respectively, which recapitulated the characteristic mouse islet architecture and validated our approach. Both Ucn3 low and Ucn3 negative beta cells localized peripherally, just internally from alpha cells in the mouse islet (Figure 1J). When broken down by age, immature and mature beta cell populations in the first postnatal week distributed in proximity, but segregated as adult architecture is established (Shih et al., 2013) (Figure S1).

To confirm these findings independent of staining we crossed mIns1-H2b-mCherry (Benner et al., 2014) and Ucn3-eGFP reporter mice (Figure S2A, B). Islets from bitransgenic offspring revealed a small population of mCherry single positive beta cells at the islet periphery, while beta cells in the rest of the islet were consistently co-positive for eGFP and mCherry (Figure 2A, B; Movie S1).

Ucn3-negative beta cells at the islet edge are ‘virgin’ beta cells

While Ucn3 is an excellent beta cell maturation marker, dedifferentiating beta cells also lose Ucn3 (Blum et al., 2014; van der Meulen et al., 2015). Therefore, Ucn3 negative beta cells at the islet edge could either be new beta cells or have lost Ucn3 expression secondary to dedifferentiation. To distinguish between both scenarios, we lineage-traced mature beta cells by crossing Ucn3-Cre (Figure S2C–F) to the mT/mG reporter mouse that switches from membrane-tdTomato (mT) to membrane-eGFP (mG) upon Cre expression (Muzumdar et al., 2007). Ucn3-negative beta cells at the islet periphery had not acquired a green mature beta lineage-label and instead retained mT (Figure 1K). This ruled out that these were once mature beta cells that dedifferentiated, but is entirely consistent with the scenario that peripheral Ucn3-negative cells are ‘virgin’ beta cells that arose from a peripheral non-beta precursor within the islet.

Expression profile of the immature beta cells at the neogenic niche

We then purified mCherry positive immature beta cells and mCherry/eGFP co-positive mature beta cells (Figure 2C) from the same islets by FACS (Figure 2D) for transcriptome analysis. Overall, immature and mature beta cells were more similar to each other than to non-beta cells from the same islets (Figure 2E) and revealed no differential expression for 75% of detectable genes (Figure 2F). The remaining genes contained several clusters

that were selectively enriched in either immature or mature beta cells (Figure 2G). Ucn3 and several transcription factors associated with beta cell identity, such as Pdx1, Nkx6–1, Insm1, Mnx1, Neurod1, Pax6, and Mafa were enriched in mature beta cells. (Figure 2H). In contrast, Mafb and Neurog3 were enriched in immature beta cells (Figure 2H). This was confirmed by immunofluorescence, where a majority of Ucn3-negative immature beta cells did not yet stain for Mafa and retained Mafb (Figure 2I, J). A minority of Ucn3-negative immature beta cells lacked nuclear Pdx1, but most immature cells already showed nuclear staining for Nkx6–1 (Figure 2K, L). On the basis of their transcriptome, immature beta cells resembled peri- and early post-natal time points, while mature beta cells resembled the later post-natal stages (Figure S3).

Ucn3 negative beta cells are transcriptionally immature

We next addressed if Ucn3 negative beta cells were functional. We observed significant reductions in immature cells of essential beta cell genes, including many genes required for glucose-stimulated insulin secretion, the TCA cycle and oxidative phosphorylation (Figure 3A–D). In contrast, key metabolic genes that are either considered ‘disallowed’ or contribute to their regulation (Dhawan et al., 2015; Martinez-Sanchez et al., 2016; Piccand et al., 2014; Pullen and Rutter, 2013) were enriched in immature beta cells (Figure 3E). Islets were also co-stained for insulin, Ucn3, and G6pc2, which hydrolyzes glucose-6-phosphate (Pound et al., 2013) or the oxidoreductase Ero1lb, which is required for disulfide bond formation in pro-insulin (Zito et al., 2010). Both markers co-labeled Ucn3-positive mature beta cells, and the majority of Ucn3 negative beta cells did not express G6pc2 or Ero1lb (Figure 3F–I). We did not observe G6pc2-negative beta cells with Ucn3, but detected a minority of Ucn3-positive beta cells without Ero1lb. This indicates that G6pc2 expression generally preceded Ero1lb and Ucn3.

Ucn3 negative beta cells are functionally immature

The reduced expression of genes encoding critical insulin secretion components, including Slc2a2, established that Ucn3 negative beta cells are transcriptionally immature, but did not directly demonstrate their functional status. We determined that Ucn3-negative peripheral beta cells lacked cell-surface expression of the essential Glut2 glucose transporter that is encoded by Slc2a2 (Figure 4A, B, S4), which suggested they were incapable of sensing glucose. Indeed, glucose uptake experiments in intact mIns1-H2b-mCherry islets using the fluorescent non-hydrolysable glucose analog 6-NBDG revealed rapid glucose uptake across the entire extracellular and intracellular space of the islet after 50 minutes, with the exception of a small number of mIns-H2b-mCherry positive beta cells at the islet periphery (Figure 4C, D; Movie S2).

We next administered the beta cell toxin Streptozotocin (STZ; 120 mg/kg on 2 consecutive days), which is taken up specifically by beta cells in a Glut2-dependent manner, to acutely ablate beta cells. While STZ treatment kills a large fraction of beta cells, beta cells that survive STZ treatment dedifferentiate, downregulate Ucn3, and are therefore difficult to distinguish from immature beta cells. We therefore used Ucn3-Cre x mT/mG bitransgenic mice where immature beta cells would remain Ucn3 lineage-negative (red), while dedifferentiated beta cells remain Ucn3 lineage-positive (green) despite the down

regulation of Ucn3 expression. Indeed, STZ treatment killed most mature beta cells, leading to a collapse in islet architecture with remaining Ucn3 lineage-positive beta cells down regulating insulin at 48 hours after the second STZ injection. Ucn3 lineage-negative beta cells on the other hand persisted and continued to show normal insulin expression (Figure 4E). Citrate controls once again revealed clusters of Ucn3 lineage-negative immature beta cells at the periphery of otherwise uniformly Ucn3 lineage-positive islets (Figure 4F). As the purpose of this experiment was to test the ability of immature beta cells to resist acute STZ-mediated cell death, we did not track their fate long-term.

As STZ administration *in vivo* provides only a snapshot of the islet at the conclusion of the experiment, we applied STZ (5 mM) *ex vivo* to Ucn3-eGFP x mIns1-H2b-mCherry bitransgenic islets, while imaging continuously in the presence of the nuclear dead cell marker Sytox Blue. STZ proceeded to kill all Ucn3-eGFP positive beta cells over the course of the 10-hour treatment (Figure 4G). In contrast, mCherry single positive immature beta cells are among the last beta cells to take up Sytox Blue (Movie S3). Collectively, these three experiments demonstrate that immature beta cells have impaired glucose uptake and are protected from STZ-mediated beta cell death, in line with their lack of cell-surface Glut2.

As ATP generated by glycolytic flux triggers beta cell depolarization downstream of Glut2, this experiment did not address if immature beta cells were capable of supporting voltage-triggered inward calcium currents. We therefore directly recorded these cells in voltage-clamp mode and observed that immature beta cells were unable to respond to stepwise depolarization with the opening of L-type voltage-gated calcium channels (Figure 4H, I).

Immature beta cells are present in human islets

In human neonatal islets, Ucn3-negative beta cells preferentially localize to the interface between clusters of alpha and beta cells or to the periphery. (Figure 5A–B). This suggests that a spatially distinct neogenic niche exists at an early age in human islets before alpha and beta cells progressively intermingle into the adult human islet architecture. Moreover, we observed residual insulin-positive, Ucn3-negative beta cells in human donors with diagnosed Type 1 Diabetes (Figure 5D, E), although it is impossible to know without lineage tracing whether these cells reflect regeneration of new beta cells or dedifferentiation of existing ones.

Immature beta cells are an intermediate stage in alpha to beta transdifferentiation

The location of immature beta cells at the neogenic niche suggested the presence of a local precursor distinct from existing mature beta cells. This may be a progenitor *sensu strictu* that is destined to become a beta cell, or it could be another differentiated cell type that, under the right circumstances, becomes a beta cell. This led us to hypothesize that the persistent population of immature beta cells we discovered reflects an intermediate stage in the transdifferentiation of alpha to mature beta cells. We therefore lineage-labeled alpha cells using bitransgenic offspring of a cross between Gcg-Cre (Herrera, 2000) and mT/mG reporter mice and readily observed cells with an alpha cell lineage-label (green) that expressed Ucn3 instead of glucagon (Figure 6A). The acquisition of an alpha cell lineage-label by beta cells could reflect an artifact of transgenesis. Under such a stochastic

scenario, one would expect lineage-labeled beta cells to occur randomly across the cross-sectional area of the islet (Figure 6B). Alternatively, beta cells may acquire an alpha cell lineage-label if they go through an insulin/glucagon bi-hormonal progenitor stage during embryonic development. As beta cell mass increases rapidly in early postnatal life by clonal expansion, this would lead to clusters of clonally related beta cells with an alpha cell lineage-label. Indeed, such a distribution was observed in adult islets of a Gcg-Cre line that labeled bi-hormonal cells during pancreas development (Shiota et al., 2013), but is never observed in the line we used (Herrera, 2000). However, if beta cells that carry an alpha cell lineage-label reflect transdifferentiation of alpha to beta cells at the periphery of adult islets, one would expect that beta cells with an alpha cell lineage-label preferentially co-distribute with alpha cells (Figure 6B). Indeed, we observed these cells to be peripherally distributed just internally from the alpha cell population (Figure 6C). Moreover, we occasionally observed immature beta cells with an alpha cell lineage-label that did not yet co-express Ucn3 at the islet edge (Figure 6D). Nevertheless, to more firmly establish that alpha to beta transdifferentiation takes place in adult islets at the neogenic niche, we used Gcg-CreER x lsl-eYFP bitransgenic reporter mice (Ackermann et al., 2017) in a pulse-chase approach to lineage-label all alpha cells at 2 months of age. Two days after the lineage-labeling, the fraction of beta cells with an alpha cell lineage-label was < 0.2%, but this fraction had increased significantly 4 months after tamoxifen (Figure 6E, F). This is consistent with a scenario where alpha cells transdifferentiate via a transient, immature stage that expresses insulin, but not Ucn3. Interestingly, the total fraction of glucagon-positive cells that carries an alpha cell lineage-label also dropped from nearly 100% to slightly less than 97% over the same period (Figure 6G), suggesting that new alpha cells arose from a non-alpha source during the same window.

Based on the absence of mT/mG co-positive ‘yellow’ cells in the pancreas of bitransgenic Ins-Cre x LSL-mT/mG reporter mice it had previously been concluded that there was no postnatal contribution of non-beta cells to the beta cell mass (Xiao et al., 2013); a conclusion that appears to contradict our current observations. However, beta cells are only (mT/mG co-positive) for a short, 4-day window after activation of Ins1-CreER with a single tamoxifen pulse (Figure S5). Moreover, this study attempted to detect this transient ‘yellow’ intermediate stage in dissociated whole pancreas (which consists of >98% non-islet cells) and was by design underpowered to detect a relatively small but ongoing contribution of non-beta cells to the beta cell mass. By using isolated islets and a constitutive Cre to accumulate transdifferentiation events, we achieved a markedly better sensitivity.

Beta cells of alpha cell descent are transcriptionally similar to conventional beta cells

We compared the transcriptomes of beta cells of alpha cell descent with conventional beta cells from the same islets, using mIns1-H2b-mCherry mice crossed to Gcg-Cre and lsl-YFP mice (DiGruccio et al., 2016). In this strategy, alpha cells that transdifferentiated into beta cells acquired mCherry, but maintained the YFP lineage-label and can be FACS-sorted accordingly (Figure 6H, I, Movie S4). We applied a stringent doublet exclusion strategy to ensure that the 0.3% mCherry/YFP co-positive events did not reflect doublets. We validated this by imaging cytometry and confirmed that transdifferentiated cells expressed mRNA for both mCherry and YFP (Figure S6). Approximately 81% of all detectable genes were shared

by all three cell populations. The remaining 19% of detectable genes were split between alpha and beta cells. Less than 1% of all genes were enriched in trans-differentiated cells (Figure 6J). This is relevant, as these genes may contain markers to identify beta cells from alpha cell descent in human islets, where lineage-tracing is not possible. Transdifferentiated cells more closely resembled beta cells, but clustered intermediate from the alpha cells that they once were and the beta cells that they became (Figure 6K, L).

Beta cells that arise via transdifferentiation from alpha cells are functionally mature

To determine whether transdifferentiation from alpha cells at the neogenic niche generates functionally mature beta cells, we assessed their ability to respond to glucose stimulation using a floxed allele of the genetically encoded calcium indicator GCaMP6. This strategy served the dual purpose of a traditional Cre-dependent lineage-labeling approach with the ability to dynamically record the behavior of the lineage-labeled cells within intact islets in real time. We conducted these experiments on the mIns1-H2b-mCherry background to distinguish beta from non-beta cells. We first established the prototypical responses of alpha and beta cells to known alpha and beta cell-selective cues. To selectively stimulate alpha cells, we applied the peptide hormone AVP, which activates V1b receptors that are selectively expressed by alpha cells (DiGruccio et al., 2016) (Figure 7A) and dose-dependently stimulated glucagon release (Dunning et al., 1984) (Figure 7B). Alpha cells within intact islets responded simultaneously with a calcium response to the initial application of each escalating dose of AVP, which was followed by a period of uncoordinated, alpha cell-autonomous calcium activity (Figure 7C; Movie S5). To selectively stimulate beta cells, we continuously perfused with 16.8 mM glucose, which activated a pulsatile, synchronized calcium response uniquely characteristic of healthy beta cells (Figure 7D; Movie S5). To establish the functional responses of transdifferentiated cells at the neogenic niche, we used Gcg-Cre to lineage-label alpha cells with GCaMP6. As before, AVP stimulated an alpha cell-autonomous response. However, upon switching to 16.8 mM glucose, a new population of cells at the islet surface that had not responded to AVP now started firing in the synchronized, pulsatile fashion that is characteristic of functionally mature beta cells (Figure 7E; Movie S6). While their GCaMP6 expression indicated alpha cell lineage history, their functional behavior and Ins1-dependent expression of nuclear mCherry each confirmed current beta cell identity; AVP-responsive alpha cells maintained a dark nuclear shade (Figure 7E, F). A brief depolarization (30 mM KCl) activated all GCaMP6-expressing cells, regardless of current identity to indicate cell viability throughout the experiment.

Mature beta cells transdifferentiate into alpha cells at the islet edge

Our findings thus far raised the question whether the privileged micro-environment of the neogenic niche specifically supports the conversion of alpha cells into beta cells, or supports lineage plasticity in general. The observation that new lineage-negative alpha cells appeared during the 4 months following the tamoxifen-induced labeling of the entire alpha cells (Figure 6G) suggests that new alpha cells are continuously generated from a non-alpha progenitor. Indeed, when we used Ucn3-Cre to lineage-label mature beta cells, we observed Ucn3 lineage-labeled cells at the islet edge that expressed glucagon instead of Ucn3 (Figure 7G). To determine if such alpha cells of mature beta descent fully completed

their transdifferentiation, islets from mIns1-H2b-mCherry mice that expressed GCaMP6 in the alpha (Gcg-Cre) or mature beta lineage (Ucn3-Cre) were simultaneously subjected to the same set of alpha and beta cell-specific stimuli as before (Movie S7). We readily identified Ucn3 lineage-labeled GCaMP6-positive cells that nevertheless responded in typical alpha cell fashion to AVP under basal glucose conditions and were indistinguishable from conventional Gcg-Cre-expressing alpha cells in the adjacent islet (Figure 7H). These cells no longer participated in the synchronized pulsatile calcium response following subsequent stimulation with 16.8 mM glucose, which implies that they were no longer coupled to the beta cell mass via gap junctions. They also no longer had an mCherry+ nucleus, but responded robustly to depolarization by KCl (Figure 7H, I). This established that alpha cells can transdifferentiate from mature beta cells at the edge of healthy islets.

Discussion

Here we describe a novel population of Ucn3-negative beta cells that occurs throughout life at a spatially distinct ‘neogenic niche’ at the islet periphery. These cells express insulin, but are transcriptionally immature and lack key beta cell markers such as G6pc2, Ero1lb and cell-surface Glut2. Indeed, these beta cells do not sense glucose and cannot support calcium influx following depolarization, indicating that they are functionally immature. They represent virgin beta cells that constitute a transitional stage in the transdifferentiation from alpha into mature, functional beta cells. We also observe the converse transition from mature beta into functional alpha cells. Based on the totality of these observations, we propose that the islet periphery contains a specialized ‘neogenic niche’; a specific micro-environment where the necessary cues and conditions to facilitate islet cell plasticity are met. The differentiation of pancreatic beta cells during embryonic development is driven by spatio-temporal gradients that determine the formation of endocrine progenitors at the trunk of the budding ducts followed by their delamination from the ductal epithelia to eventually cluster into islets (Shih et al., 2013). Our discovery suggests the continued importance of the 3-dimensional environment in determining the plasticity and maintenance of islet cell fate.

The continuous formation of new beta cells from a progenitor pool in adult islets is a distinct departure from the long-standing paradigm that maintenance of beta cell mass occurs exclusively through beta cell self-replication (Brennand et al., 2007; Dor et al., 2004). However, since the latter study was conducted using a low-efficiency (30%) lineage-labeling strategy dependent on the rat insulin promoter, it was underpowered to rule out a relatively small contribution from non-beta cells. Moreover, this strategy could not have differentiated between the immature and mature beta cells we report here, as both express insulin. Our observations are easily reconciled with the common view that beta cell self-renewal is the major mechanism to regenerate or expand beta cell mass in mice, provided sufficient beta cells remain (Cox et al., 2016). Our discovery of the neogenic niche extends this model to establish that beta cells can and do arise via alternative paths within islets of healthy, non-diabetic individuals. This is particularly relevant to established T1D, where proliferation of any residual beta cells by itself is unlikely to support meaningful regeneration of functional beta cell mass, given that human beta cells rarely proliferate (Meier et al., 2008). This emphasizes the need to pursue alternative beta cell sources to restore functional beta cell mass.

Transdifferentiation from alpha to beta cells without genetic perturbation of transcription factor expression was first described following near-complete beta cell ablation in mice (Thorel et al., 2010). The process is commonly presumed to be induced by insulin deficiency or pancreatic injury and remodeling (Chera et al., 2014; Habener and Stanojevic, 2012). Instead, our observations demonstrate that transdifferentiation between alpha and beta cells takes place within a specialized niche at the periphery of healthy islets. It follows that the near-complete removal of existing beta cells as the predominant source of beta cell regeneration may have merely unmasked transdifferentiation as an alternate contributor to beta cell regeneration. This represents an important conceptual shift: it suggests that the signal(s) that control plasticity at the neogenic niche are associated with the constellation of cell types and autonomic nerves that converge at the periphery of healthy islets instead of being induced in response to major pancreas remodeling. Of course, this does not rule out a scenario where transdifferentiation is further stimulated under certain (patho)physiological circumstances, as was recently demonstrated when blockade of glucagon-dependent amino acid clearance by the liver promoted alpha cell proliferation, accompanied by increased alpha to beta transdifferentiation (Solloway et al., 2015).

Islet architecture differs markedly between adult human and mouse islets (Brissova et al., 2005; Cabrera et al., 2006; Dolensek et al., 2015; Kim et al., 2009). However, in younger human donor islets, alpha and beta cells segregate into distinct areas or adopt a mantle/core topology similar to mouse islets (Figure 5). We readily observe Ucn3-negative beta cells at the alpha/beta interface in islets of neonatal and infant pancreas donors and can also identify such cells in adult human pancreas. This suggests that a neogenic niche similar to the one we discovered in mice may exist in human islets, at least in younger ages. Moreover, we continue to observe Ucn3 negative beta cells within islets of patients with decades-long history of Type 1 Diabetes, possibly reflecting subclinical beta cell regeneration supported by the islet plasticity we describe here. However, human islet architecture becomes less defined with age and Ucn3 expression is lost from beta cells in diabetes (Blum et al., 2014; van der Meulen et al., 2015). The fact that a lack of Ucn3 marks both immature and de-differentiated beta cells complicates its utility as a marker to establish whether immature beta cells continue to congregate in a spatially distinct niche. Nevertheless, our observations could indicate human alpha cell-autonomous potential to adopt a beta cell identity that may be re-engaged by recreating the appropriate conditions that converge at the neogenic niche.

Our discovery that a population of immature beta cells persists in the islets of healthy adult mice and, possibly, humans, is also a strong testament to the significant heterogeneity that has long been known amongst beta cells (Giordano et al., 1991; Kiekens et al., 1992). Insulin-expressing Pancreatic Multipotent Progenitor cells (PMPs) were reported that are capable of extensive self-replication, self-renewal and contribute to multiple pancreatic and neural cell types (Smukler et al., 2011). These and similar cells found in small clusters outside of normal islets (Beamish et al., 2016) share a lack of cell-surface Glut2 with the immature beta cells we discovered at the niche. However, the genes enriched in the PMP transcriptome (Razavi et al., 2015) demonstrate virtually no overlap with those enriched in immature beta cells (Figure S4). Moreover, while we consider Ucn3-negative immature beta cells to be a reservoir of beta cell progenitors, we do not have evidence to suggest that they possess the multipotent progenitor properties described for PMPs. Recently, a distinct

population of approximately 15% of all beta cells was described to resist immune attack in non-obese diabetic (NOD) mice. These beta cells are characterized by reduced expression of beta cell identity genes including Glut2 and likely reflect the de-differentiation of existing beta cells (Rui et al., 2017). While these cells are distinct from the virgin beta cells that we describe, these observations nevertheless suggest that immature beta cells by virtue of their reduced expression of beta cell identity genes – including the major auto-antigens insulin and G6pc2 – may escape auto-immune diabetes like they resist STZ-mediated beta cell death.

Other recent papers describe a distinct population of beta 'hub' cells that control the response of islets to stimulation (Johnston et al., 2016) or report heterogeneity based on a transgenic reporter strain for Flattop (Bader et al., 2016). However, based on several lines of evidence, it is clear that the Ucn3-negative immature beta cells we discovered are completely distinct from both. Flattop negative beta cells constitute approximately 20% of the total beta cell mass and hub cells make up 5 – 10% of the beta cells, with both beta cell types found throughout the islet (Figure S4). This contrasts with the 1.5% of beta cell number and the peripheral location of Ucn3-negative immature beta cells. Moreover, Flattop positive and negative beta cells both robustly express Ucn3 protein (Bader et al., 2016); *vice versa* Ucn3-negative immature and Ucn3-expressing mature beta cells express similar levels of Flattop mRNA (Figure S4), indicating that Flattop and Ucn3 do not mark the same beta cell population. In human islets four sub-populations of beta cells were recently described based on the expression of ST8SIA1 and CD9 (Dorrell et al., 2016). ST8SIA1 positive cells share traits with the immature beta cells we describe here (Figure S4), although others (including ST8SIA1 itself) are distinct between these beta cell sub-populations.

In summary, the plasticity of beta cell identity is a double-edged sword that provides an opportunity to regenerate beta cells from endogenous progenitors, while contributing to their dedifferentiation in diabetes (Jonas et al., 1999; Talchai et al., 2012). The lessons we learn from beta cell maturation at the neogenic niche will be widely applicable under different scenarios where beta cell maturity is either desired or compromised. This includes the generation of functionally mature human beta cells from stem cells to alleviate the pressing need for an unlimited source of human beta cells to cure T1D - a feat that despite impressive recent progress in the field (Pagliuca et al., 2014; Rezania et al., 2014; Russ et al., 2015) - has not been achieved. We envision that the signaling responses and epigenetic mechanisms that underlie beta cell maturation at the niche overlap with those that are impaired when beta cells dedifferentiate in diabetes. It follows that the discovery of the specific spatio-temporal cues that govern islet cell fate at the neogenic niche can be wielded to take advantage of the significant plasticity in beta cell regeneration. These insights could apply to generate functionally mature beta cells from endogenous sources or stem cell-derived progenitors and to block the adverse consequences of the beta cell identity crisis in diabetes.

STAR Methods

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Mark Huising (mhuising@ucdavis.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals—Commercial male C57BL6/NHsd mice were obtained from Harlan (Indianapolis, IN; now Envigo) at 3, 6 and 9 weeks and 3, 8 and 14 months of age and used immediately for pancreas collection. A number of transgenic mouse lines were employed. The dual color lsl-mT/mG reporter mouse (*Gt(ROSA)26Sor*$^{tm4(ACTB-tdTomato,-eGFP)Luo}$/J, JAX strain 007576) (Muzumdar et al., 2007). The lsl-eYFP reporter mouse (B6.129X1-*Gt(ROSA)26Sor*$^{tm1(EYFP)Cos}$/J) (Srinivas et al., 2001). Calcium levels in the cell were visualized using the lsl-GCaMP6 mouse line (B6;129S6-*Gt(ROSA)26Sor*$^{tm96(CAG-GCaMP6s)Hze}$/J) (Madisen et al., 2015). Two Ucn3 BAC transgenic reporter mice were used that are based on BAC clone RP23–332L13, which contains the Ucn3 gene flanked by more than 197 kb of genomic context: the Ucn3-Cre line (B6.FVB(Cg)-Tg(Ucn3-cre)KF43Gsat/Mmucd and the Ucn3-eGFP line (Tg(Ucn3-eGFP)HY36Gsat/Mmucd). For alpha cell labeling we employed a Gcg-Cre mouse line (Herrera, 2000) or a Gcg-CreER line (Ackermann et al., 2017). To label beta cells we used the B6.Cg-Tg(Ins1-HIST1H2BB/mCherry)5091Mhsg/J mouse line (Benner et al., 2014). To lineage label beta cells in Figure S6, we used the Ins1-CreER mouse line (B6.Cg-Tg(Ins1-cre/ERT)1Lphi/J) (Wicksteed et al., 2010). All transgenic lines are maintained by back crossing to commercially obtained C57BL6/NHsd (Envigo). Unless indicated otherwise, adult mice of both sexes were used between 3 and 9 months of age. Animals were maintained in group-housing on a 12-h light/12-h dark cycle with free access to water and standard rodent chow. All animal procedures were approved by the Salk Institute for Biological Studies and UC Davis Institutional Animal Care and Use Committees and performed in compliance with the Animal Welfare Act and the ILAR Guide to the Care and Use of Laboratory Animals.

Validation of novel transgenic mouse models—We observed close overlap between the expression of Ucn3 peptide and the expression of GFP in the Ucn3-eGFP mouse line Tg(Ucn3-eGFP)HY36Gsat/Mmucd (Figure S2). We evaluated two distinct Ucn3-Cre lines and validated the Ucn3-Cre(KF43) line to selectively and specifically label the Ucn3 expression domain in within the islets (Figure S2). In this line, over 98% of all insulin-positive beta cells in bitransgenic offspring from the Ucn3-Cre x mT/mG cross carries the mG lineage-label, validating this reporter line. We also tested the B6.FVB(Cg)-Tg(Ucn3-cre)KF31Gsat/Mmucd) line, but found the efficiency and specificity of this line to be undesirable (Figure S2).

Primary cell cultures—Primary islets were cultured in RPMI (5.5 mM glucose, 10% FBS, pen/strep) under 5% CO_2 at 37° C in 10 cm petri dishes (*i.e.* not tissue culture treated). Islets for microscopy or dissociated islet cells for electrophysiology experiments were cultured overnight on uncoated #1.5 glass-bottom 35 mM culture dishes (MatTek Corporation, Ashland, MA). No cell lines were used in this study.

Human Subjects—Human donor pancreata were obtained via the Network of Pancreatic Islet Donors with Diabetes (nPOD). Gender and age are indicated in the figure legend. The UC Davis Institutional Review Board declared the human pancreas histological specimens

used in this study exempt from IRB review under 45 CFR 46.101 (b) Category (4) on December 11, 2014.

METHOD DETAILS

Islet isolation—Islets were isolated by injecting collagenaseP (0.8 mg/mL in HBSS; Roche Diagnostics) (Invitrogen) via the common bile duct while the ampulla of Vater was clamped (Huising et al., 2010). The entire pancreas was collected following the injection of 2mL collagenase solution and, after addition of 2 more ml of collagenase solution, was incubated at 37°C for 13 min. Pancreata were dissociated by gentle manual shaking followed by three washes with cold HBSS containing 5% NCS. The digested suspension was passed through a nylon mesh (pore size 425 μm; Small Parts Inc.), and islets were isolated by density gradient centrifugation on a Histopaque gradient (1.077 g/mL density; Sigma) for 20 min at 1400 × g without brake. Islets were collected from the interface, washed once with cold HBSS containing 5% NCS, and hand-picked several times under a dissecting microscope prior to culture in RPMI (5.5 mM glucose, 10% FBS, pen/strep).

Flow and imaging cytometry—Islets were dissociated by incubation in 0.25% trypsin/EDTA for 2 minutes complemented by gentle trituration with a p200 pipette, washed once in RPMI (5.5 mM glucose, 10% FBS, pen/strep) and immediately processed on the cytometer. Conventional FACS separation was conducted based on red and green fluorescence, using 1 μg/ml Dapi for live/dead exclusion, as previously described (DiGruccio et al., 2016; van der Meulen et al., 2015) and samples were collected directly in Trizol reagent for library preparation. Samples for imaging cytometry were run on an Amnis Imagestream MarkII imaging cytometer (EMD Millipore, Billerica, MA) at the Gladstone Institutes in San Francisco.

Next generation sequencing—FACS-sorted samples were collected directly into Trizol reagent. RNA isolation and library construction was done by Illumina's TruSeq RNA sample Prep Kit v2 (Illumina Inc. San Diego, CA), sequenced at 50 cycles, and single read on an Illumina HiSeq 2000 platform as previously described (DiGruccio et al., 2016).

Immunohistochemistry—Immunohistochemistry was conducted as follows: slides were washed 3 times for 5 minutes each in KPBS, antibodies were diluted in KPBS supplemented with 2% donkey serum and 0.4% Triton-X100 and applied overnight at 4 degrees Celsius. After 3 more washes in KPBS, slides were incubated with secondary antibodies (obtained from Jackson ImmunoResearch (Westgrove, PA) and used at 1:600 final dilution), also in donkey block, now for 45 minutes at room temperature. Three more washes completed the procedure. Where applicable, nuclei were counterstained by Dapi at 1 μg/ml final concentration and slides were embedded in Prolong Gold Antifade (Thermo Fisher Scientific, Waltham, MA) and imaged on either a Zeiss LSM780 confocal microscope or a Nikon A1R+ confocal microscope. The antiserum for Ero1lb (Zito et al., 2010) was generously provided by Dr. David Ron, the antiserum for G6pc2 (Hutton and Eisenbarth, 2003) was originally generated by the late Dr. John Hutton and generously provided by Drs. Jay Walters and Howard Davidson.

Glucose uptake—To measure glucose uptake, we incubated intact islets from freshly isolated mIns1-H2b-mCherry reporter mice overnight on uncoated #1.5 glass-bottom 35 mM culture dishes (MatTek Corporation, Ashland, MA) in RPMI (10% FBS, 5.5 mM glucose, pen/strep). The next day, Z-stacks of islets were continuously acquired as the non-hydrolysable glucose analog 6-NBDG (6-(*N*-(7-Nitrobenz-2-oxa-1,3-diazol-4-yl)amino)-6-Deoxyglucose; Thermo Fisher Scientific, Waltham, MA) was added at a final concentration of 0.3 mM, using a Nikon A1R+ confocal microscope in resonant scanning mode. The relative rate of glucose uptake was determined by drawing ROIs of individual mIns1-H2b-mCherry+ beta cells that either had or had not taken up glucose.

Streptozotocin treatment—Ucn3-Cre x mTmG mice were treated with two doses of 120 mg/kg of streptozotocin (EMD Millipore, Billerica MA) dissolved fresh in 100 mM sodium citrate (pH 4.5) on consecutive days. Citrate controls were included and animals were monitored closely around the clock to prevent hypoglycemia from STZ-induced insulin release. Mice were euthanized 48 hours after the second STZ injection when hyperglycemia indicated the destruction of most beta cells and processed for immunohistochemistry. To study streptozotocin-induced beta cell death *ex vivo*, we cultured mIns1-H2b-mCherry x Ucn3-eGFP bitransgenic islets as described above and followed them for 15 hours in the presence of 5 mM STZ and the nuclear dead cell marker Sytox Blue (500 nM, Thermo Fisher Scientific, Waltham, MA) while acquiring a Z-stack every 30 minutes on a Nikon A1R+ confocal microscope in resonant scanning mode.

Pulse-chase of Gcg-CreER mice—Gcg-CreER x lsl-eYFP bitransgenic mice were treated at 2 months of age with tamoxifen (100 μg/g BW) 3 times over the course of 5 days and collected 2 days (pulse) or 4 months (chase) after the last injection. These samples were embedded in paraffin and a GFP antiserum was used to detect eYFP protein following antigen retrieval by microwave in 10 mM sodium citrate pH 6.0.

Electrophysiology—Dispersed immature and mature beta cells from mIns1-H2b-mCherry x Ucn3-eGFP bitransgenic islets were patched in the whole cell configuration and held at −70 mV. All cells were subjected to the voltage clamp protocol depicted in Figure 4H. Recordings were performed using the following internal solution: (mM) Cs-aspartate 87, CsCl 20, $MgCl_2$ 1, MgATP 5, HEPES 10 pH 7.2, EGTA 10 and external solution: (mM) NaCl 138, KCl 5.4, $MgCl_2$ 1, $CaCl_2$10, HEPES 10 pH 7.4, glucose 2.8. Glass capillary electrodes (Suttter Instruments) were pulled to have a resistance of 7–12 MΩ. All electrical signals were acquired using a HEKA EPC 10 USB dual head stage amplifier. Each head stage was mounted on a single Sutter micromanipulator driven by the MPC 225 controller. All cells were visualized using a Zeiss Axiovert inverted microscope with 10x, 40x and 63x objectives with filters for blue, GFP, YFP, and red fluorophores. The scope was mounted upon a Siskiyou x y translator bolted to a TMC anti-vibration table. All islet cells were continuously perfused with external solution maintained at 33°C via a stage mounted incubation system. Temperature control was achieved with a SF-20 in line heater (Harvard Apparatus). Data was acquired using HEKA Patchmaster software version 2×90.2. Signal data was analyzed using Clampfit 10 (Molecular devices).

Glucagon secretion—Static glucagon secretion experiments were carried out on 30 wild type mouse islets per well in Krebs Ringer Buffer (KRB). Islets were isolated the day prior to the secretion assay, cultured overnight in RPMI (5.5 mM glucose, 10% FBS, pen/strep) and transferred to KRB containing 5.5 mM glucose an hour before the start of the assay. Islets were picked to the wells plate for final secretion in 10% of the final assay volume. The remaining 90% of volume consisted of KRB without glucose, with the indicated concentration of AVP added, to effectively reduce the final glucose concentration to 0.5 mM at the start of the secretion assay. AVP peptide was synthesized in-house and generously provided by Dr. Jean Rivier (Salk Institute).

Calcium responses in intact islets—We used islets from a triple transgenic offspring of a cross between mIns1-H2b-mCherry, lsl-GCaMP6 and either Gcg-Cre or Ucn3-Cre to label the alpha or mature beta cell lineages, respectively. Live islets were cultured overnight after the islet prep, placed on 35mm dishes with glass bottom (#1.5; MatTek Corporation), allowed to attach overnight and imaged in x, y, z and t on a Nikon A1R+ confocal microscope using a 40x lens with a long working distance. Treatments were continuously perfused over the islets using a Masterflex peristaltic pump at 2.5 ml per minute. Each protocol concluded with a 30 mM potassium chloride pulse to demonstrate viability and responsiveness throughout the treatment. Individual cells in individual z-planes were defined as regions of interest (ROI) and the green fluorescence intensity within the ROIs was plotted over time as a measure of calcium activity.

QUANTIFICATION AND STATISTICAL ANALYSIS

Bio-informatics analysis—Read refinement for quality and adapter contamination was performed using FASTQC, Scythe, and Sickle. After quality control and filtering, the libraries comparing immature to mature beta cells had an average library size of 21.4 million reads. The libraries comparing alpha and beta cells to beta cells of alpha cell-descent had an average library size of 19.2 million reads. The libraries of the postnatal maturation time series had an average library size of 12.8 million reads. Libraries were then aligned to mouse genome version GenCode M8 (GRCm38.p4 M8; mm10) using STAR (Dobin et al., 2013) with the two-pass method with default parameters, with the exception of a tighter mismatch restriction of no greater than 3% per read. Average unique read alignment was 87.5%, 86.6% and 86.9% for the three sets of libraries, respectively. Bigwigs were generated using the wiggle output option in STAR and Genome Utilities in the UCSC Genome Browser (Kent et al., 2002). Gene-level quantification was performed on sorted BAM files using featureCounts (Liao et al., 2014) with default parameters, counted by Gencode defined exons, and aggregated to the gene level. Differential expression analyses were performed using the edgeR generalized linear model approach and maximum likelihood method testing (Robinson et al., 2010). Results were filtered for statistical significance using the thresholds: FDR < 0.001 and an absolute $\log_2$FC > 1 across all experiments. In comparing immature and mature beta cells from adult islets, we observed 3394 differentially expressed genes (2117 Immature enriched, 1277 Mature enriched) that met these thresholds. Using the same criteria, we observed 2480 genes to be differentially expressed between any of the pair-wise comparisons in comparing alpha and beta cells with trans-differentiated cells. Heat maps were generated based on the n most up- and n most down-regulated genes between each

pairwise comparison, with duplicate genes removed, where n = 750 for Figure 2G and n = 200 for Figure 6K. To enable the most direct comparison, the same analysis pipeline— from raw sequencing reads, to quantification and differential analyses— was also applied to compare our data against mouse PMPs (Razavi et al., 2015) and human beta cell subtypes (Dorrell et al., 2016). Distance matrices for the latter comparison were generated using the R package lattice, with RPKM values from each experiment used to correlate relatedness.

Bio-informatics visualization—All clustering and heat map visuals were generated using Gene Cluster 3.0 (de Hoon et al., 2004) on log transformed RPKM values normalized by the median and hierarchically clustered by the city block distance method. The R-package ComplexHeatmap (Gu et al., 2016) was used on the clustering output using default parameters, with the exception of the maturation time series experiment where the Pearson instead of the Euclidean distance method was used. All Venn diagrams were generated using the R library package VennDiagram (Chen and Boutros, 2011), using the edgeR output and our FDR and $\log_2$FC filtering criteria for gene selection. We included a lower RPKM threshold of 1 when comparing our immature beta cell transcriptomes with those of the PMPs (Razavi et al., 2015). KEGG gene set enrichment testing was performed on the immature beta data set using the R Bioconductor package GAGE (Luo et al., 2009) on the edgeR output that met the statistical threshold. Pathway visuals were created using the R-package Pathview (Luo and Brouwer, 2013).

Distribution algorithm—The distribution algorithm we developed to quantify the distribution of cell types across the cross-sectional surface of the islet is build up into three main parts: segmentation, classification, and distribution measurement. All classification was done manually for each cell by a person. The segmentation and the distribution measurements are handled using two separate Matlab functions that interface with Bitplane's Imaris software. We utilized Imaris in this process to provide an accurate visual representation of the segmentations for verification and classification. After loading the data set into Imaris and converting the data to 32-bit float, the segmentation function is called in to segment the islet volume and nuclear volumes separately using a simple intensity based threshold for binary segmentation. The user is prompted to input which channel(s) should be used for each of these segmentation steps. The segmentation function then creates two surface objects in Imaris, one for the islet sub volume and the second for the nuclei. Both the islet and nuclei surface objects are visually verified for accuracy, followed by the manual classification of each individual cell into distinct bins by cell type. Once the classification is complete, the distribution measurement function is called to find the distributions of each classified group created by the researcher. The distribution measurement function creates a new channel that is the distance transformation with respect to the interior edge of the islet. The distance transformation channel intensity is then recorded at the center of each classified nucleus. Recorded measurements are written to a csv file containing all measurements along with the sample information.

Statistical Analysis—Data were analyzed by t-test, corrected for multiple comparisons using the Holm-Sidak method where appropriate. The distribution of cells across the islet was tested using a non-parametric Kruskal-Wallis test using Dunn's multiple comparisons

test. A post-hoc Kolmogorov-Smirnov test was conducted to compute the D statistic on the basis of the cumulative frequency distributions computed from raw data. Data are represented as mean ± SEM across, with n defined in the corresponding figure legend. Differences were considered significant when $p<0.05$. Statistics were computed using Prism (GraphPad Software, La Jolla, CA).

DATA AND SOFTWARE AVAILABILITY

Sequencing data sets—Sequencing data associated with this publication have been deposited in GEO under GSE88778, GSE88779 and GSE90766.

Distribution algorithm—The Matlab routine for the distribution algorithm described in this paper can be found here: http://huisinglab.com/cell_metabolism_2017/index.html.

ADDITIONAL RESOURCES

A resource website featuring UCSC browser plots searchable by plain text query for the RNAseq data presented in this paper can be found at: http://huisinglab.com/cell_metabolism_2017/index.html.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We gratefully acknowledge Dr. David Ron for sharing the Ero1lb antiserum and Drs. Jay Walters and Howard Davidson for sharing the G6pc2 antiserum generated by the late Dr. John Hutton. We thank Ms. Bridget McLaughlin from the UC Davis Flow Cytometry Shared Resource Laboratory (funded by NCI P30 CA093373 and NIH NCRR C06-RR12088, S10 RR12964 and S10 RR 026825) for help with flow cytometry and Dr. Marielle Cavrois for help with the Amnis Imagestream imaging cytometer (DOD grant W81XWH-11-1-0562). We thank Ms. Giselle Blanco for technical assistance. The research described in this paper was funded by a Career Development Award from the Juvenile Diabetes Research Foundation (2-2013-54) and an Individual Biomedical Research Award from the Hartwell Foundation to MOH.

References

Ackermann AM, Zhang J, Heller A, Briker A, and Kaestner KH (2017). High-fidelity Glucagon-CreER mouse line generated by CRISPR-Cas9 assisted gene targeting. Mol Metab in press.

Artner I, Hang Y, Mazur M, Yamamoto T, Guo M, Lindner J, Magnuson MA, and Stein R. (2010). MafA and MafB regulate genes critical to beta-cells in a unique temporal manner. Diabetes 59, 2530–2539. [PubMed: 20627934]

Bader E, Migliorini A, Gegg M, Moruzzi N, Gerdes J, Roscioni SS, Bakhti M, Brandl E, Irmler M, Beckers J, et al. (2016). Identification of proliferative and mature beta-cells in the islets of Langerhans. Nature 535, 430–434. [PubMed: 27398620]

Beamish CA, Strutt BJ, Arany EJ, and Hill DJ (2016). Insulin-positive, Glut2-low cells present within mouse pancreas exhibit lineage plasticity and are enriched within extra-islet endocrine cell clusters. Islets 8, 65–82. [PubMed: 27010375]

Benner C, van der Meulen T, Caceres E, Tigyi K, Donaldson CJ, and Huising MO (2014). The transcriptional landscape of mouse beta cells compared to human beta cells reveals notable species differences in long non-coding RNA and protein-coding gene expression. BMC Genomics 15, 620. [PubMed: 25051960]

Blum B, Hrvatin SS, Schuetz C, Bonal C, Rezania A, and Melton DA (2012). Functional beta-cell maturation is marked by an increased glucose threshold and by expression of urocortin 3. Nat. Biotechnol 30, 261–264. [PubMed: 22371083]

Blum B, Roose AN, Barrandon O, Maehr R, Arvanites AC, Davidow LS, Davis JC, Peterson QP, Rubin LL, and Melton DA (2014). Reversal of beta cell de-differentiation by a small molecule inhibitor of the TGFbeta pathway. eLife 3, e02809. [PubMed: 25233132]

Bonner-Weir S, Li WC, Ouziel-Yahalom L, Guo L, Weir GC, and Sharma A. (2010). Beta-cell growth and regeneration: replication is only part of the story. Diabetes 59, 2340–2348. [PubMed: 20876724]

Bramswig NC, Everett LJ, Schug J, Dorrell C, Liu C, Luo Y, Streeter PR, Naji A, Grompe M, and Kaestner KH (2013). Epigenomic plasticity enables human pancreatic alpha to beta cell reprogramming. J Clin Invest 123, 1275–1284. [PubMed: 23434589]

Brennand K, Huangfu D, and Melton D. (2007). All beta Cells Contribute Equally to Islet Growth and Maintenance. PLoS Biol 5, e163. [PubMed: 17535113]

Brissova M, Fowler MJ, Nicholson WE, Chu A, Hirshberg B, Harlan DM, and Powers AC (2005). Assessment of human pancreatic islet architecture and composition by laser scanning confocal microscopy. J Histochem Cytochem 53, 1087–1097. [PubMed: 15923354]

Butler AE, Cao-Minh L, Galasso R, Rizza RA, Corradin A, Cobelli C, and Butler PC (2010). Adaptive changes in pancreatic beta cell fractional area and beta cell turnover in human pregnancy. Diabetologia 53, 2167–2176. [PubMed: 20523966]

Cabrera O, Berman DM, Kenyon NS, Ricordi C, Berggren PO, and Caicedo A. (2006). The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. Proc Natl Acad Sci U S A 103, 2334–2339. [PubMed: 16461897]

Chen H, and Boutros PC (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics 12, 35. [PubMed: 21269502]

Chera S, Baronnier D, Ghila L, Cigliola V, Jensen JN, Gu G, Furuyama K, Thorel F, Gribble FM, Reimann F, et al. (2014). Diabetes recovery by age-dependent conversion of pancreatic delta-cells into insulin producers. Nature.

Collombat P, Hecksher-Sorensen J, Krull J, Berger J, Riedel D, Herrera PL, Serup P, and Mansouri A. (2007). Embryonic endocrine pancreas and mature beta cells acquire alpha and PP cell phenotypes upon Arx misexpression. J Clin Invest 117, 961–970. [PubMed: 17404619]

Collombat P, Xu X, Ravassard P, Sosa-Pineda B, Dussaud S, Billestrup N, Madsen OD, Serup P, Heimberg H, and Mansouri A. (2009). The ectopic expression of Pax4 in the mouse pancreas converts progenitor cells into alpha and subsequently beta cells. Cell 138, 449–462. [PubMed: 19665969]

Cox AR, Lam CJ, Rankin MM, King KA, Chen P, Martinez R, Li C, and Kushner JA (2016). Extreme obesity induces massive beta cell expansion in mice through self-renewal and does not alter the beta cell lineage. Diabetologia 59, 1231–1241. [PubMed: 27003683]

de Hoon MJ, Imoto S, Nolan J, and Miyano S. (2004). Open source clustering software. Bioinformatics 20, 1453–1454. [PubMed: 14871861]

Dhawan S, Tschen SI, Zeng C, Guo T, Hebrok M, Matveyenko A, and Bhushan A. (2015). DNA methylation directs functional maturation of pancreatic beta cells. J Clin Invest 125, 2851–2860. [PubMed: 26098213]

DiGruccio MR, Mawla AM, Donaldson CJ, Noguchi GM, Vaughan J, Cowing-Zitron C, van der Meulen T, and Huising MO (2016). Comprehensive alpha, beta and delta cell transcriptomes reveal that ghrelin selectively activates delta cells and promotes somatostatin release from pancreatic islets. Mol Metab 5, 449–458. [PubMed: 27408771]

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. [PubMed: 23104886]

Dolensek J, Rupnik MS, and Stozer A. (2015). Structural similarities and differences between the human and the mouse pancreas. Islets 7, e1024405. [PubMed: 26030186]

Dor Y, Brown J, Martinez OI, and Melton DA (2004). Adult pancreatic beta-cells are formed by self-duplication rather than stem-cell differentiation. Nature 429, 41–46. [PubMed: 15129273]

Dorrell C, Schug J, Canaday PS, Russ HA, Tarlow BD, Grompe MT, Horton T, Hebrok M, Streeter PR, Kaestner KH, et al. (2016). Human islets contain four distinct subtypes of beta cells. Nat Commun 7, 11756. [PubMed: 27399229]

Dunning BE, Moltz JH, and Fawcett CP (1984). Actions of neurohypophysial peptides on pancreatic hormone release. Am J Physiol 246, E108–114. [PubMed: 6364829]

Gao T, McKenna B, Li C, Reichert M, Nguyen J, Singh T, Yang C, Pannikar A, Doliba N, Zhang T, et al. (2014). Pdx1 maintains beta cell identity and function by repressing an alpha cell program. Cell Metab 19, 259–271. [PubMed: 24506867]

Giordano E, Bosco D, Cirulli V, and Meda P. (1991). Repeated glucose stimulation reveals distinct and lasting secretion patterns of individual rat pancreatic B cells. J Clin Invest 87, 2178–2185. [PubMed: 2040700]

Gu Z, Eils R, and Schlesner M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 32, 2847–2849. [PubMed: 27207943]

Habener JF, and Stanojevic V. (2012). alpha-cell role in beta-cell generation and regeneration. Islets 4, 188–198. [PubMed: 22847495]

Herrera PL (2000). Adult insulin- and glucagon-producing cells differentiate from two independent cell lineages. Development 127, 2317–2322. [PubMed: 10804174]

Huising MO, van der Meulen T, Vaughan JM, Matsumoto M, Donaldson CJ, Park H, Billestrup N, and Vale WW (2010). CRFR1 is expressed on pancreatic beta cells, promotes beta cell proliferation, and potentiates insulin secretion in a glucose-dependent manner. Proc. Natl. Acad. Sci. USA 107, 912–917. [PubMed: 20080775]

Hutton JC, and Eisenbarth GS (2003). A pancreatic beta-cell-specific homolog of glucose-6-phosphatase emerges as a major target of cell-mediated autoimmunity in diabetes. Proc Natl Acad Sci U S A 100, 8626–8628. [PubMed: 12861077]

Johnston NR, Mitchell RK, Haythorne E, Pessoa MP, Semplici F, Ferrer J, Piemonti L, Marchetti P, Bugliani M, Bosco D, et al. (2016). Beta Cell Hubs Dictate Pancreatic Islet Responses to Glucose. Cell Metab 24, 389–401. [PubMed: 27452146]

Jonas JC, Sharma A, Hasenkamp W, Ilkova H, Patane G, Laybutt R, Bonner-Weir S, and Weir GC (1999). Chronic hyperglycemia triggers loss of pancreatic beta cell differentiation in an animal model of diabetes. J Biol Chem 274, 14112–14121. [PubMed: 10318828]

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, and Haussler D. (2002). The human genome browser at UCSC. Genome Res 12, 996–1006. [PubMed: 12045153]

Kiekens R, In ‘t Veld P, Mahler T, Schuit F, Van De Winkel M, and Pipeleers D. (1992). Differences in glucose recognition by individual rat pancreatic B cells are associated with intercellular differences in glucose-induced biosynthetic activity. J Clin Invest 89, 117–125. [PubMed: 1729264]

Kim A, Miller K, Jo J, Kilimnik G, Wojcik P, and Hara M. (2009). Islet architecture: A comparative study. Islets 1, 129–136. [PubMed: 20606719]

Liao Y, Smyth GK, and Shi W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930. [PubMed: 24227677]

Luo W, and Brouwer C. (2013). Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics 29, 1830–1831. [PubMed: 23740750]

Luo W, Friedman MS, Shedden K, Hankenson KD, and Woolf PJ (2009). GAGE: generally applicable gene set enrichment for pathway analysis. BMC Bioinformatics 10, 161. [PubMed: 19473525]

Madisen L, Garner AR, Shimaoka D, Chuong AS, Klapoetke NC, Li L, van der Bourg A, Niino Y, Egolf L, Monetti C, et al. (2015). Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. Neuron 85, 942–958. [PubMed: 25741722]

Martinez-Sanchez A, Pullen TJ, Chabosseau P, Zhang Q, Haythorne E, Cane MC, Nguyen-Tu MS, Sayers SR, and Rutter GA (2016). Disallowance of Acot7 in beta-Cells Is Required for Normal Glucose Tolerance and Insulin Secretion. Diabetes 65, 1268–1282. [PubMed: 26861785]

Meier JJ, Butler AE, Saisho Y, Monchamp T, Galasso R, Bhushan A, Rizza RA, and Butler PC (2008). Beta-cell replication is the primary mechanism subserving the postnatal expansion of beta-cell mass in humans. Diabetes 57, 1584–1594. [PubMed: 18334605]

Muzumdar MD, Tasic B, Miyamichi K, Li L, and Luo L. (2007). A global double-fluorescent Cre reporter mouse. Genesis 45, 593–605. [PubMed: 17868096]

Pagliuca FW, Millman JR, Gurtler M, Segel M, Van Dervort A, Ryu JH, Peterson QP, Greiner D, and Melton DA (2014). Generation of functional human pancreatic beta cells in vitro. Cell 159, 428–439. [PubMed: 25303535]

Papizan JB, Singer RA, Tschen SI, Dhawan S, Friel JM, Hipkens SB, Magnuson MA, Bhushan A, and Sussel L. (2011). Nkx2.2 repressor complex regulates islet beta-cell specification and prevents beta-to-alpha-cell reprogramming. Genes Dev 25, 2291–2305. [PubMed: 22056672]

Parsons JA, Brelje TC, and Sorenson RL (1992). Adaptation of islets of Langerhans to pregnancy: increased islet cell proliferation and insulin secretion correlates with the onset of placental lactogen secretion. Endocrinology 130, 1459–1466. [PubMed: 1537300]

Piccand J, Strasser P, Hodson DJ, Meunier A, Ye T, Keime C, Birling MC, Rutter GA, and Gradwohl G. (2014). Rfx6 maintains the functional identity of adult pancreatic beta cells. Cell reports 9, 2219–2232. [PubMed: 25497096]

Pound LD, Oeser JK, O'Brien TP, Wang Y, Faulman CJ, Dadi PK, Jacobson DA, Hutton JC, McGuinness OP, Shiota M, et al. (2013). G6PC2: a negative regulator of basal glucose-stimulated insulin secretion. Diabetes 62, 1547–1556. [PubMed: 23274894]

Pullen TJ, and Rutter GA (2013). When less is more: the forbidden fruits of gene repression in the adult beta-cell. Diabetes Obes Metab 15, 503–512. [PubMed: 23121289]

Razavi R, Najafabadi HS, Abdullah S, Smukler S, Arntfield M, and van der Kooy D. (2015). Diabetes enhances the proliferation of adult pancreatic multipotent progenitor cells and biases their differentiation to more beta-cell production. Diabetes 64, 1311–1323. [PubMed: 25392245]

Rezania A, Bruin JE, Arora P, Rubin A, Batushansky I, Asadi A, O'Dwyer S, Quiskamp N, Mojibian M, Albrecht T, et al. (2014). Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. Nat Biotechnol.

Robinson MD, McCarthy DJ, and Smyth GK (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140. [PubMed: 19910308]

Rui J, Deng S, Arazi A, Perdigoto AL, Liu Z, and Herold KC (2017). Beta Cells that Resist Immunological Attach Develop during Progression of Autoimmune Diabetes in NOD Mice. Cell Metab in press.

Russ HA, Parent AV, Ringler JJ, Hennings TG, Nair GG, Shveygert M, Guo T, Puri S, Haataja L, Cirulli V, et al. (2015). Controlled induction of human pancreatic progenitors produces functional beta-like cells in vitro. EMBO J 34, 1759–1772. [PubMed: 25908839]

Shih HP, Wang A, and Sander M. (2013). Pancreas organogenesis: from lineage determination to morphogenesis. Annu Rev Cell Dev Biol 29, 81–105. [PubMed: 23909279]

Shiota C, Prasadan K, Guo P, El-Gohary Y, Wiersch J, Xiao X, Esni F, and Gittes GK (2013). alpha-Cells are dispensable in postnatal morphogenesis and maturation of mouse pancreatic islets. Am J Physiol Endocrinol Metab 305, E1030–1040. [PubMed: 23982158]

Smukler SR, Arntfield ME, Razavi R, Bikopoulos G, Karpowicz P, Seaberg R, Dai F, Lee S, Ahrens R, Fraser PE, et al. (2011). The adult mouse and human pancreas contain rare multipotent stem cells that express insulin. Cell Stem Cell 8, 281–293. [PubMed: 21362568]

Solloway MJ, Madjidi A, Gu C, Eastham-Anderson J, Clarke HJ, Kljavin N, Zavala-Solorio J, Kates L, Friedman B, Brauer M, et al. (2015). Glucagon Couples Hepatic Amino Acid Catabolism to mTOR-Dependent Regulation of alpha-Cell Mass. Cell reports 12, 495–510. [PubMed: 26166562]

Srinivas S, Watanabe T, Lin CS, William CM, Tanabe Y, Jessell TM, and Costantini F. (2001). Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. BMC developmental biology 1, 4. [PubMed: 11299042]

Sullivan BA, Hollister-Lock J, Bonner-Weir S, and Weir GC (2015). Reduced Ki67 Staining in the Postmortem State Calls Into Question Past Conclusions About the Lack of Turnover of Adult Human beta-Cells. Diabetes 64, 1698–1702. [PubMed: 25488899]

Talchai C, Xuan S, Lin HV, Sussel L, and Accili D. (2012). Pancreatic beta cell dedifferentiation as a mechanism of diabetic beta cell failure. Cell 150, 1223–1234. [PubMed: 22980982]

Teta M, Long SY, Wartschow LM, Rankin MM, and Kushner JA (2005). Very slow turnover of beta-cells in aged adult mice. Diabetes 54, 2557–2567. [PubMed: 16123343]

Thorel F, Nepote V, Avril I, Kohno K, Desgraz R, Chera S, and Herrera PL (2010). Conversion of adult pancreatic alpha-cells to beta-cells after extreme beta-cell loss. Nature 464, 1149–1154. [PubMed: 20364121]

van der Meulen T, Donaldson CJ, Caceres E, Hunter AE, Cowing-Zitron C, Pound LD, Adams MW, Zembrzycki A, Grove KL, and Huising MO (2015). Urocortin3 mediates somatostatin-dependent negative feedback control of insulin secretion. Nat Med 21, 769–776. [PubMed: 26076035]

van der Meulen T, and Huising MO (2014). Maturation of Stem Cell-Derived Beta cells Guided by the Expression of Urocortin 3. Rev. Diabet. Stud 11, 115–132. [PubMed: 25148370]

van der Meulen T, and Huising MO (2015). Role of transcription factors in the transdifferentiation of pancreatic islet cells. J Mol Endocrinol 54, R103–117. [PubMed: 25791577]

van der Meulen T, Xie R, Kelly OG, Vale WW, Sander M, and Huising MO (2012). Urocortin 3 marks mature human primary and embryonic stem cell-derived pancreatic alpha and beta cells. PLoS One 7, e52181. [PubMed: 23251699]

Wicksteed B, Brissova M, Yan W, Opland DM, Plank JL, Reinert RB, Dickson LM, Tamarina NA, Philipson LH, Shostak A, et al. (2010). Conditional gene targeting in mouse pancreatic ss-Cells: analysis of ectopic Cre transgene expression in the brain. Diabetes 59, 3090–3098. [PubMed: 20802254]

Xiao X, Chen Z, Shiota C, Prasadan K, Guo P, El-Gohary Y, Paredes J, Welsh C, Wiersch J, and Gittes GK (2013). No evidence for beta cell neogenesis in murine adult pancreas. J Clin Invest 123, 2207–2217. [PubMed: 23619362]

Zito E, Chin KT, Blais J, Harding HP, and Ron D. (2010). ERO1-beta, a pancreas-specific disulfide oxidase, promotes insulin biogenesis and glucose homeostasis. J Cell Biol 188, 821–832. [PubMed: 20308425]

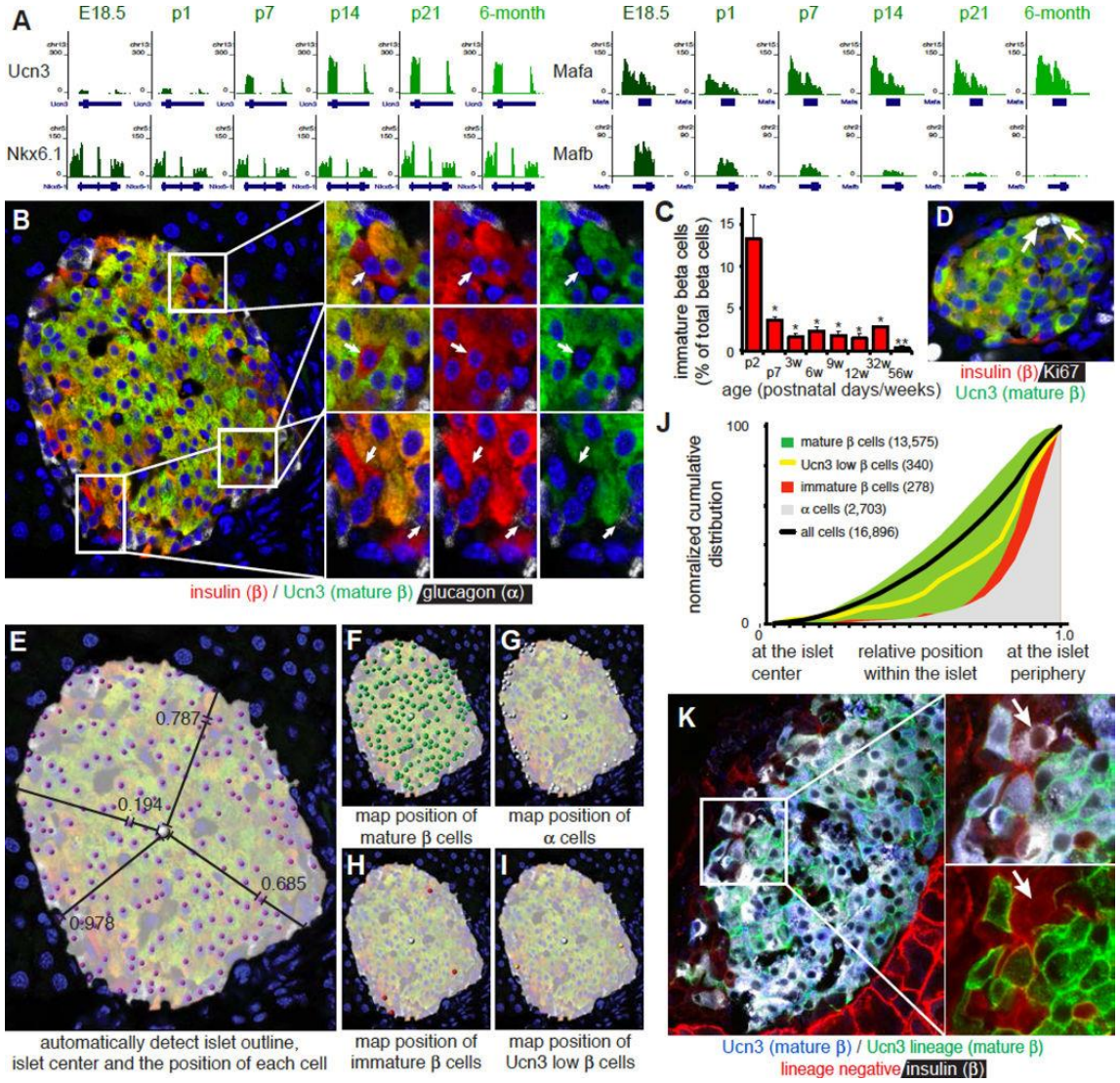


Figure 1. The absence of Ucn3 marks beta cells in the neogenic niche.

(A) Gene expression of Ucn3, Nkx6.1, Mafa and Mafb by RNAseq on FACS sorted beta cells during perinatal maturation. Gene structure and chromosome number are indicated for each panel.

(B) Detection of insulin, glucagon and Ucn3 in a 3-month old islet. Insets show Ucn3-negative beta cells.

(C) Fraction of Ucn3-negative beta cells at different ages (counted n=3 animals per time point, 10 islets each). Error bars reflect SEM, * P <0.05, ** P < 0.01.

(D) Ki67-positive beta cells maintain Ucn3 expression indicating that proliferating beta cells and Ucn3-negative beta cells are not the same.
(E-I) Image analysis to detect the islet outline and center of mass to compute the position of a cell relative to the center and nearest edge. Cells are manually classified as (F) mature beta cell (insulin and Ucn3 co-positive), (G) alpha cell (glucagon positive), (H) immature beta cell (insulin-positive, Ucn3-negative) or (I) Ucn3 low beta cell (insulin-positive, Ucn3-low).
(J) Normalized cumulative distribution of Ucn3-negative beta cells compared to mature beta cells of all ages combined (3, 6 and 9 weeks; 3, 8 and 14 months; 3 animals per age, 16,896 cells). See Figure S1 for distributions by age and Table S1A for the P values and D statistics for each pairwise comparison.
(K) Ucn3-negative, insulin positive beta cells (white, but not blue, indicated by the arrow) at the islet periphery of Ucn3-Cre x mT/mG mice are Ucn3-lineage negative (they express mTomato, instead of mGFP). These cells make up 0.75% ± 0.56% (n = 3) of all insulin-positive beta cells. See also Figure S2.

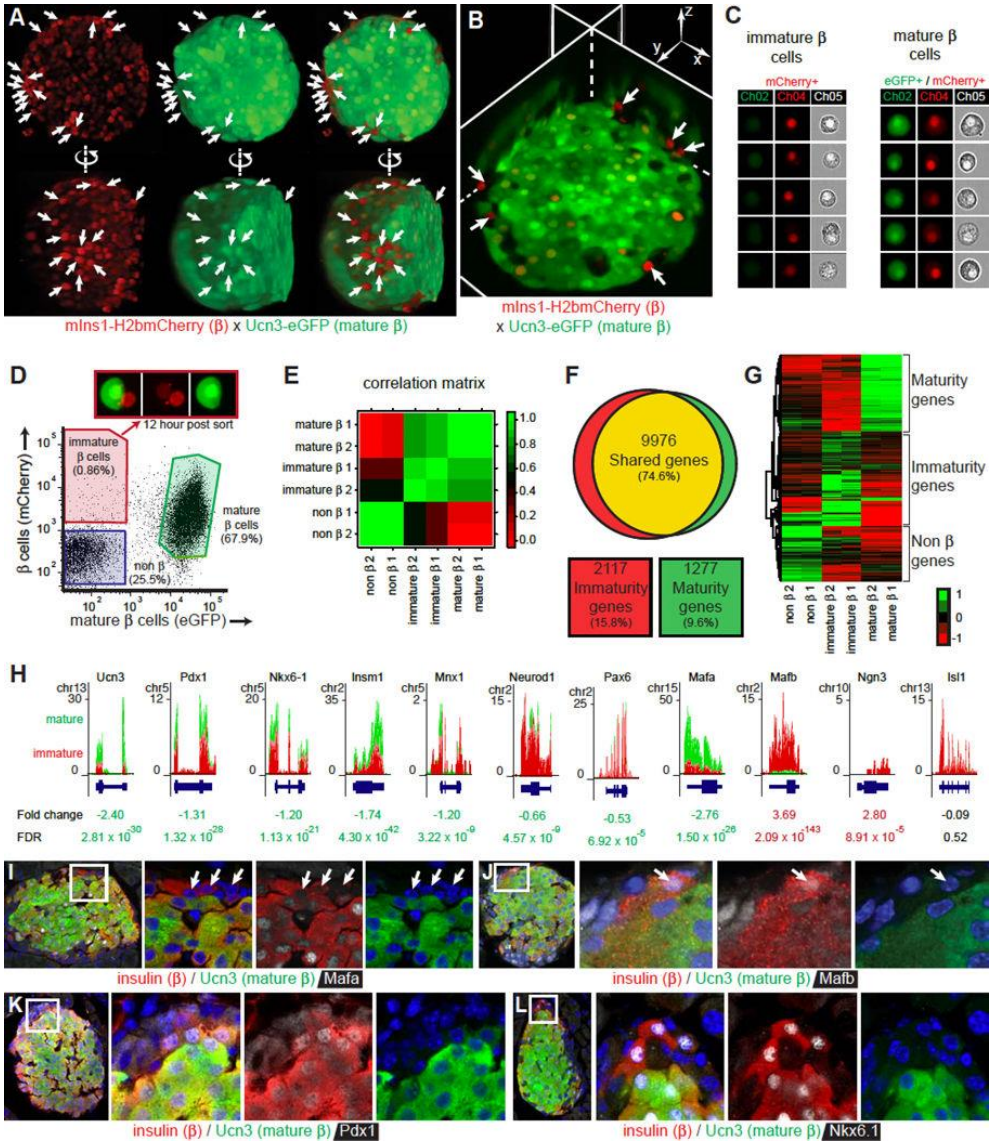


Figure 2. Comparison of mature and immature beta cells from the same islets.

(A) View from two angles of a 3D reconstructed islet expressing mIns1-H2b-mCherry (all beta cells) and Ucn3-EGFP (mature beta cells only). Arrows indicate immature beta cells. See also Figure S2.

(B) Immature beta cells (arrows) in the neogenic niche revealed by virtual slicing of a 3D reconstructed islet. See also Movie S1.

(C) Imaging cytometer analysis of individual immature and mature beta cells.

(D) FACS strategy to obtain Ucn3-eGFP mIns1-H2b-mCherry co-positive mature and mCherry single positive immature beta cells from the same islets. Immature beta cells start expressing Ucn3-eGFP in culture 12 hours after sorting (inset).
(E) Correlation matrix of the 200 top differentially expressed (100 enriched, 100 depleted) genes between mature and immature cells.
(F) Venn diagram comparing gene expression (RPKM>1 in either population) between mature and immature beta cells. Expression was considered different when the absolute $\log_2$FC > 1 and FDR < 0.001.
(G) Heat map of the most differentially expressed genes between mature and immature beta cells.
(H) Expression of key genes in mature and immature beta cells by RNAseq. See also Figure S3.
(I-L) Colocalization of insulin, Ucn3 and Mafa (I), Mafb (J), Pdx1 (K), or Nkx6–1 (L) in an adult mouse islet. Arrows indicate immature beta cells.

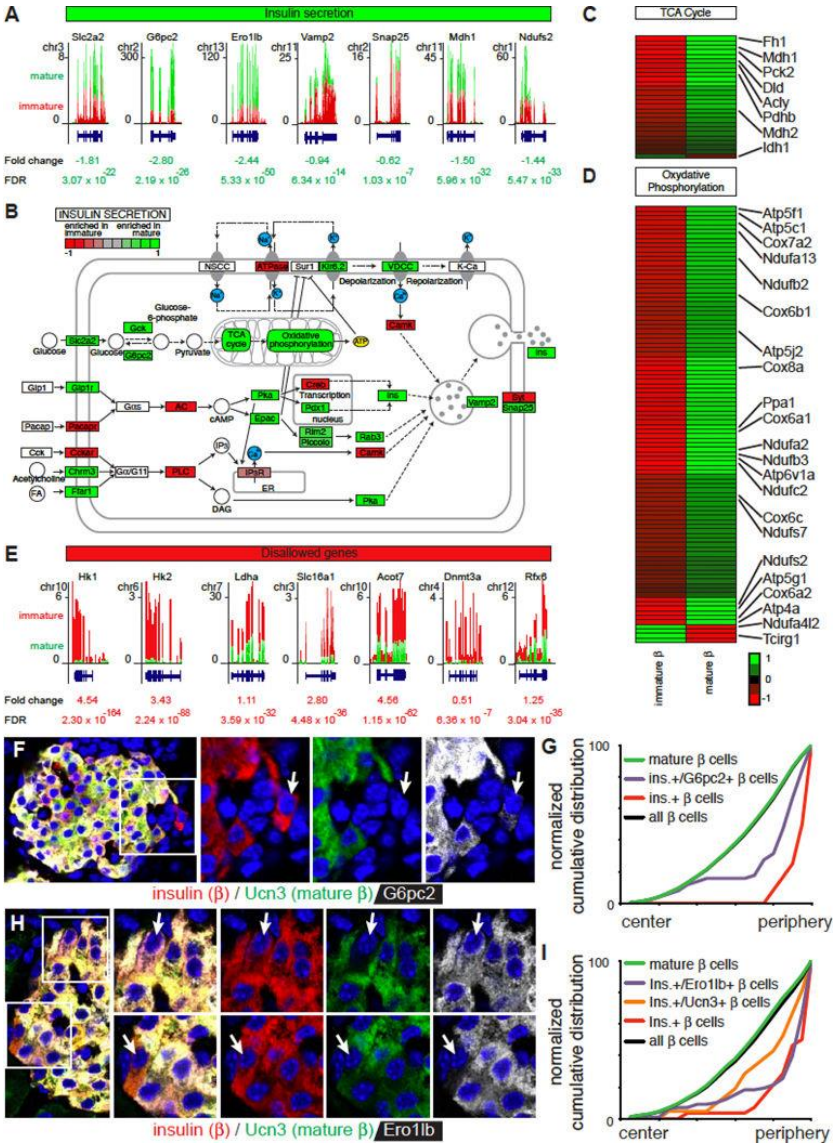


Figure 3. Ucn3 negative beta cells are transcriptionally immature.

(A) Gene expression of select genes involved in insulin secretion in mature and immature beta cells by RNAseq.

(B) Visualization of differential expression of the Kegg pathway analysis for insulin secretion (FDR < 0.001).

(C) Heat map of the differential expression of tricarboxylic acid (TCA) cycle genes between immature and mature beta cells.

(D) Heat map of the differential expression of genes involved in oxidative phosphorylation between immature and mature beta cells.
(E) Gene expression of 'disallowed' genes in mature and immature beta cells by RNAseq.
(F) Immunofluorescence detection of G6pc2 (white), insulin (red) and Ucn3 (green) in a mouse islet. Arrow indicates an immature beta cell.
(G) Distribution of 2329 G6pc2-positive and negative beta cells within mouse islets (2329 cells). See Table S1B for the P values and D statistics for each pairwise comparison.
(H) Immunofluorescence detection of Ero1lb (white), insulin (red) and Ucn3 (green) in a mouse islet. Arrows indicate examples of beta cells with Ero1lb but not Ucn3 (top) and with Ucn3 but not Ero1lb (bottom).
(I) Distribution of 2043 Ero1lb-positive and negative beta cells within mouse islets (2073 cells).

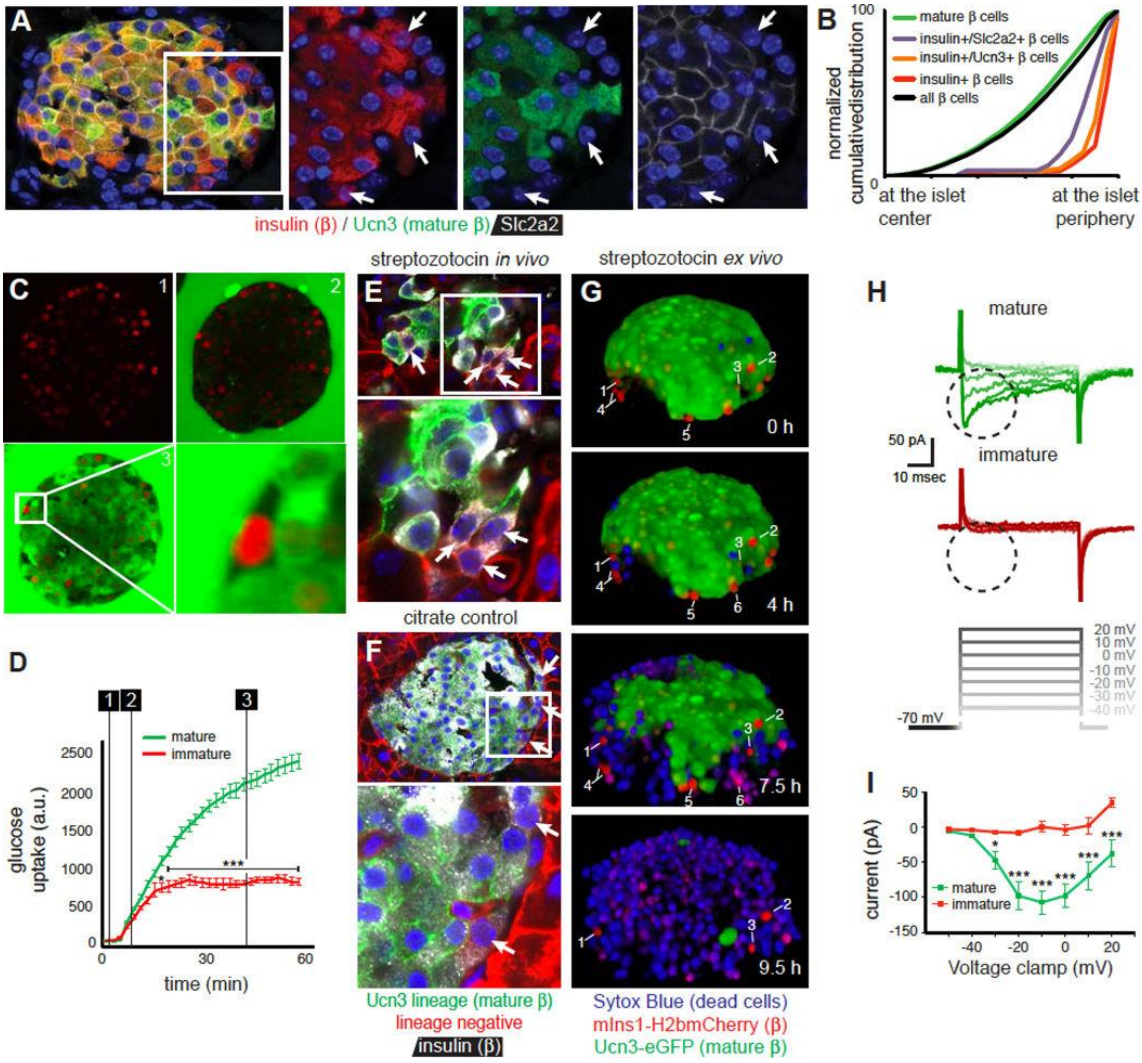


Figure 4. Ucn3 negative beta cells are functionally immature.

(A) Ucn3-negative beta cells at the periphery (arrows) do not express cell-surface Glut2.

(B) Distribution of Glut2-negative beta cells at the periphery of the islet (6783 cells total). See Table S1B for the P values and D statistics for each pairwise comparison and Figure S4 for a comparison of immature beta cells with other heterogeneous populations of beta cells.

(C) Uptake of the glucose analog 6-NBDG over time by all beta cells except for immature beta cells at the periphery of intact islets.

(D) Quantification of 6-NBDG uptake. Numbered lines corresponds to panel C. Data represent mean ± SEM for 3 immature and 7 mature beta cells. * $P < 0.05$, *** $P < 0.001$. See also Movie S2.

(E) Treatment with a high dose of the Glut2-dependent beta cell toxin streptozotocin ablates a majority of beta cells. Remaining Ucn3 lineage-positive beta cells are often insulin negative, while immature beta cells (arrows) survive owing to the lack of cell surface Glut2.
(F) Ucn3 lineage-negative beta cells at the islet edge (arrows) in citrate controls.
(G) STZ-induced death of Ucn3-eGFP x mIns1-H2b-mCherry mature beta cells in intact islets in real time. The death of beta cells is marked by the acute loss of eGFP protein and the nuclear uptake of the dead cell marker Sytox Blue. In contrast, mCherry single positive immature beta cells (numbered) do not take up Sytox Blue. Individual immature beta cells are labeled for clarity. Note that immature beta cells #4–6 disappear from the Z-stack as the islet volume expands due to the extensive cell death. See also Movie S3.
(H) Immature beta cells cannot support calcium influx following depolarization. Representative traces of depolarization-induced inward calcium currents (circled) from mIns1-mCherry+ immature beta cells and Ucn3-eGFP positive mature beta cells from the same preparations.
(I) Full Current-Voltage (I-V) plot contrasting the voltage-dependent inward calcium current in Ucn3-eGFP positive mature beta cells with the lack thereof in mIns1-mCherry+ immature beta cells from the same preparations. Data represent mean ± SEM for 4 immature and 8 mature beta cells from four individual animals. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

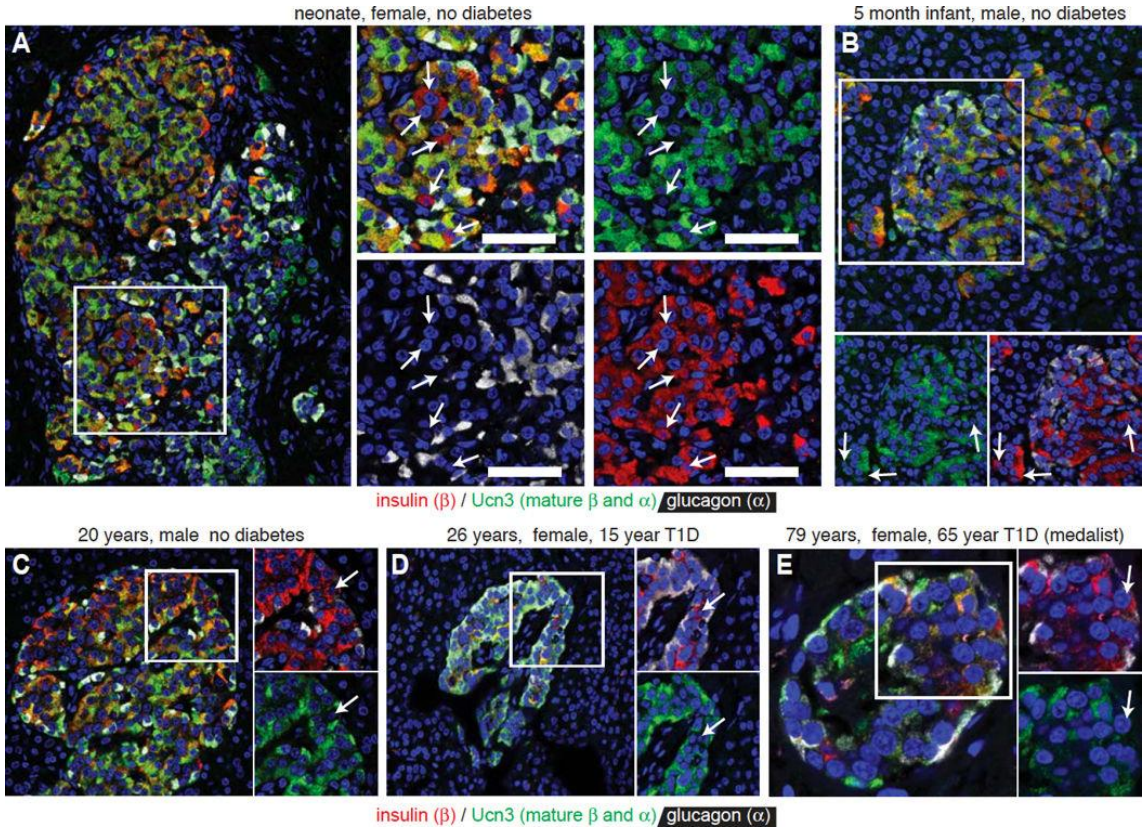


Figure 5. Ucn3-negative beta cells are present in human islets of young and adult donors and donors with T1D.

Expression of insulin, glucagon and Ucn3. Arrows indicate Ucn3-negative beta cells.

(A) Neonate, female, no diabetes, nPOD #6200.

(B) Infant, 5 months old, male, no diabetes, nPOD #6115.

(C) Adult, 20 years old, male, no diabetes, nPOD #6238.

(D) Adult, 26 years old, female, 15 years with diagnosed T1D, nPOD #6196.

(E) Adult, 79 years old, female, 56 years with diagnosed T1D (medalist), nPOD #6065.

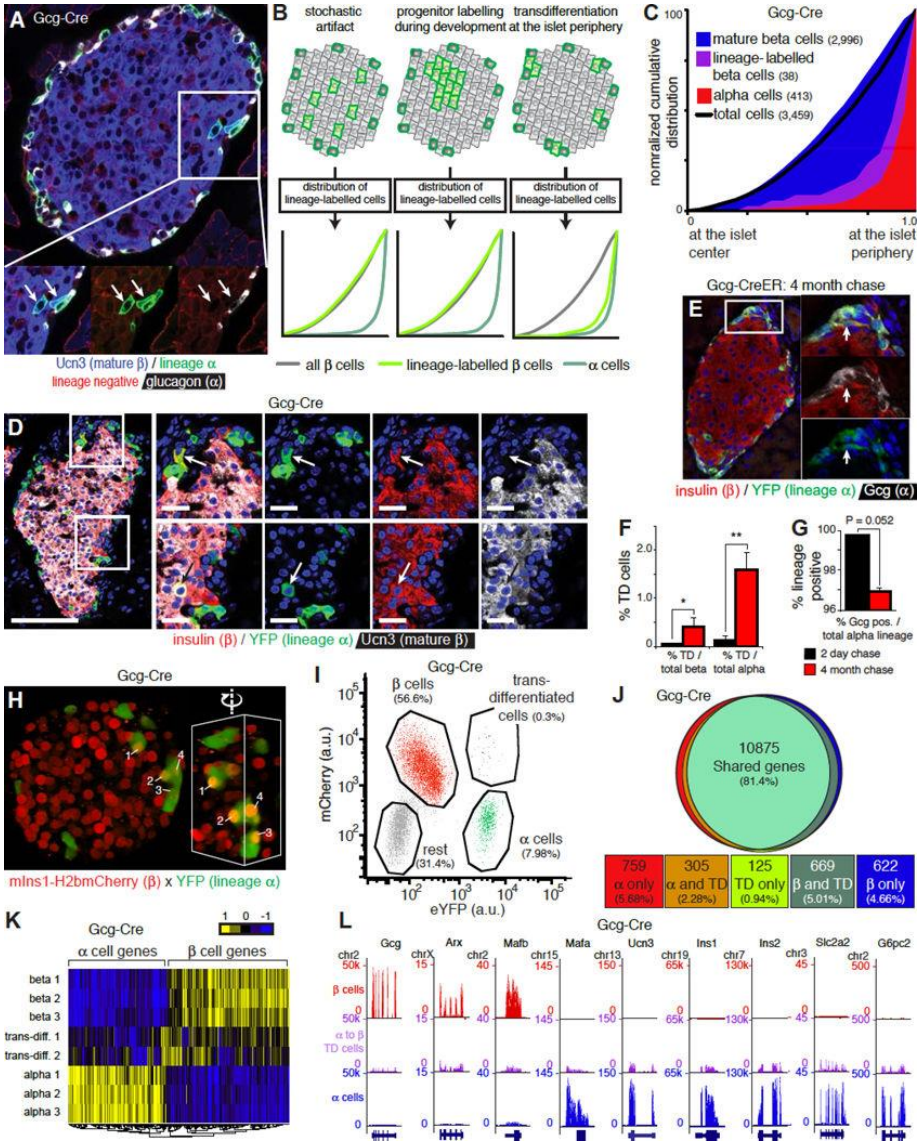


Figure 6. Immature beta cells reflect a transient stage in the transdifferentiation between alpha and beta cells at the islet edge.

(A) A pair of cells at the islet periphery with an alpha lineage mark (arrows) that now express Ucn3 instead of glucagon.

(B) Three potential scenarios that could account for the detection of beta cells with an alpha cell lineage-label and their predicted distribution across the cross-sectional islet area: 1) random labeling of beta cells by the 'leaky' expression of Cre recombinase in beta cells, 2) labeling of bi-hormonal progenitors during development is predicted to lead to randomly localized clusters of lineage-labeled beta cells that expanded from a single bi-hormonal

progenitor, 3) transdifferentiation of alpha cells into beta cells at the periphery of adult islets. See also Figure S5.

(C) Observed distribution of alpha lineage-labeled beta cells (3459 cells). See Table S1C for P values and D statistics for each pairwise comparison.

(D) Alpha lineage-labeled islet, visualized by lsl-YFP, that features two transdifferentiated cells. One of these co-expresses Ucn3 and is mature (bottom, arrow), the other cell is an immature beta cell of alpha cell-descent that expresses insulin but not yet Ucn3 (top, arrow).

(E) Lineage-labeling all alpha cells in 2-month old mice via Gcg-CreER followed by a 4-month chase demonstrates that alpha cells continue to transdifferentiate into beta cells.

(F) The fraction of beta cells of alpha cell descent increased significantly four months after lineage-labeling all alpha cells, as measured relative to either total beta cell or alpha cell number. * $P < 0.05$, ** $P < 0.01$.

(G) Conversely, the alpha cell fraction, which is nearly completely lineage-labeled immediately after tamoxifen administration, is notably diluted 4 months later by alpha cells without a lineage-label.

(H) Islets from triple transgenic offspring of a cross between mIns1-mCherry x Gcg-Cre x lsl-YFP reveal the presence of mCherry positive beta cells that carry the YFP alpha cell lineage-label at the edge of the intact islet. Inset shows detail of the same cluster from different angle to emphasize the nuclear mCherry in transdifferentiated cells. See also Movie S4.

(I) FACS strategy to isolate transdifferentiated cells along with alpha and beta cells from dissociated islets of mIns1-mCherry x Gcg-Cre x lsl-YFP triple transgenic islets. See also Figure S6.

(J) Venn diagram of genes that are detectably expressed (RPKM > 1) among alpha, beta and transdifferentiated cells. Expression was considered different when the absolute $\log_2$FC > 1 and FDR < 0.001.

(K) Heat map of alpha, beta and transdifferentiated cells based on the most differentially expressed genes from panel 6J.

(L) Expression of key genes in alpha, beta and transdifferentiated cells by RNAseq. Gene structure and chromosome number are indicated for each panel.

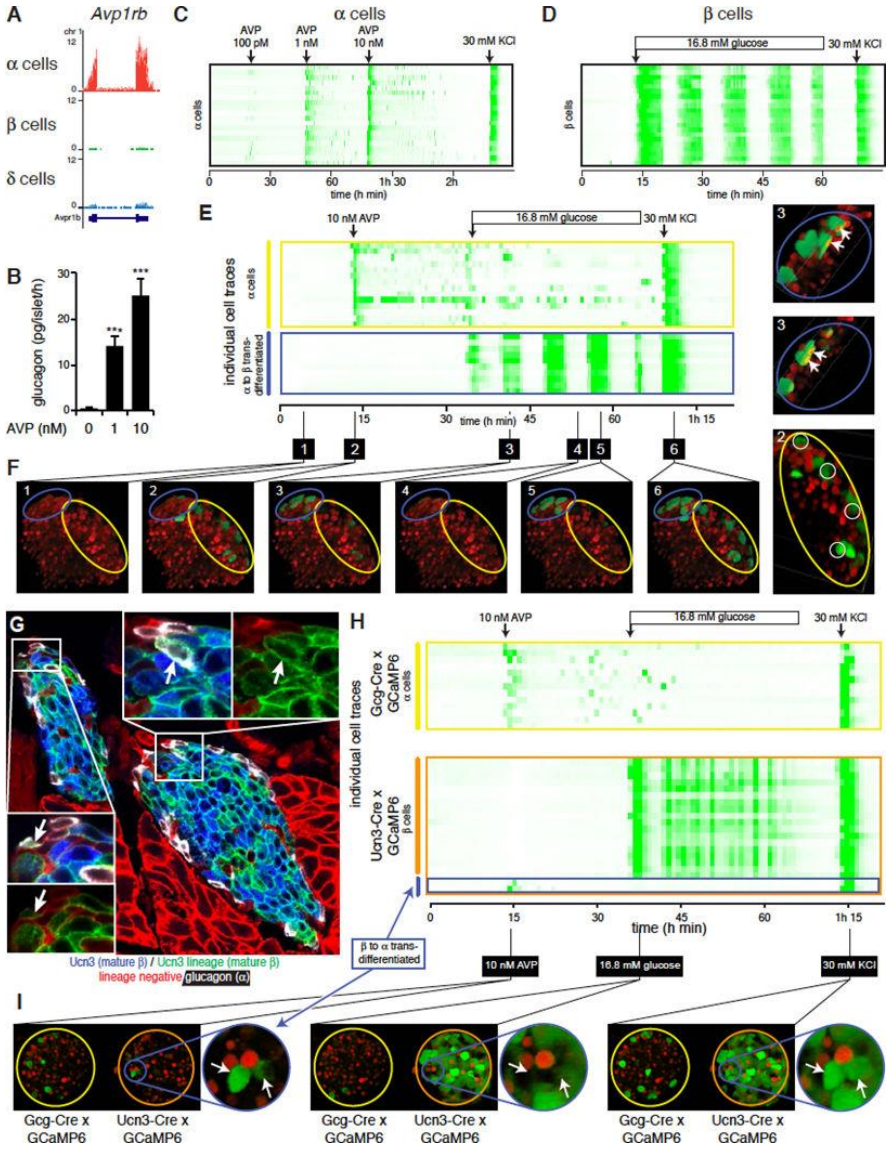


Figure 7. Transdifferentiated cells are functionally mature

(A) Expression of arginine vasopressin (AVP) receptor Avp1rb in alpha, beta and delta cells.

(B) Glucagon secretion in response to AVP. Data represent mean ± SEM; n = 4. *** P < 0.001.

(C) Alpha cell calcium activity in response to brief stimulation with increasing doses of AVP, indicated by arrows. Depolarization (30 mM KCl) serves as positive control. See also Movie S5.

(D) Coordinated beta cell calcium activity in response to continuous stimulation with 16.8 mM glucose. Arrow indicates the start of continuous stimulation. See also Movie S5.
(E) Consecutive stimulation with a brief pulse of AVP (arrow) followed by continuous stimulation with 16.8 mM glucose in an islet where GCaMP6 expression is restricted to the alpha cell lineage and nuclear mCherry marks current beta cells. See also Movie S6.
(F) Still images of key frames in at the indicated times in (E). Yellow ellipsoid contains conventional alpha cells; blue ellipsoid contains alpha-to-beta transdifferentiated cells.
(G) Beta-to-alpha transdifferentiated cells revealed by the absence of Ucn3 and expression of glucagon in Ucn3-Cre lineage cells at the islet periphery.
(H) Responses of two islets where expression of GCaMP6 is restricted to the alpha cell lineage (Gcg-Cre, top) or mature beta cell lineage (Ucn3-Cre; bottom). Stimulation protocol as in Figure 6E. Both islets were imaged simultaneously in a single recording. See also Movie S7.
(I) Still images of key frames in (H).

V. Chapter 5: Artemether Does Not Turn a Cells into b Cells

Talitha van der Meulen, Sharon Lee, Els Noordeloos, Cynthia J. Donaldson, Michael W. Adams, Glyn M. Noguchi, **Alex M. Mawla,** and Mark O. Huising

Cell Metabolism, 2018. 27(1):218-225. doi: 10.1016/j.cmet.2017.10.002

Contributions to Jointly Authored Works: As a contributing author to this manuscript, I was responsible for the entirety of formal re-analysis and interpretation of transcriptomic data, along with the re-analysis and visualization of single cell data from Li *et al,* 2017 incorporated into the paper. This manuscript was intended to highlight the lack of reproducibility that unfortunately does occur in science. A focal point in the discrepancies which stemmed this manuscript revolved around transcriptomic data results, interpretations, and visuals suggested by the original paper.

Significance of Research: In summary, this manuscript directly refuted a high-profile publication that suggested an FDA-approved drug for malaria could be utilized to regenerate beta cells in diabetic patients. The re-analysis of the data provided by the original paper conflicted with their suggested findings, as did all other experiments done by colleagues in the lab. In fact, the re-analysis of the original publication's data showed the opposite of what the authors suggested, as did all other experiments done by my colleagues. This paper has an Altemetric score of 138 and has been well cited.

Citations: 58

Published in final edited form as:

Cell Metab. 2018 January 09; 27(1): 218–225.e4. doi:10.1016/j.cmet.2017.10.002.

Artemether Does Not Turn Alpha Cells into Beta Cells

Talitha van der Meulen[1,4], **Sharon Lee**[1,4], **Els Noordeloos**[1], **Cynthia J. Donaldson**[2], **Michael W. Adams**[1], **Glyn M. Noguchi**[1], **Alex M. Mawla**[1], and **Mark O. Huising**[1,3,*]

[1]Department of Neurobiology, Physiology & Behavior, College of Biological Sciences, University of California, Davis, CA 95616, USA

[2]Clayton Foundation Laboratories for Peptide Biology, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA

[3]Department of Physiology and Membrane Biology, School of Medicine, University of California, Davis, CA 95616, USA

Summary

Pancreatic alpha cells retain considerable plasticity and can – under the right circumstances – transdifferentiate into functionally mature beta cells. In search of a targetable mechanistic basis, a recent paper suggested that the widely used antimalarial drug artemether suppresses the alpha cell transcription factor Arx to promote transdifferentiation into beta cells. However, key initial experiments in this paper were carried out in islet cell lines and most subsequent validation experiments implied transdifferentiation without direct demonstration of alpha to beta cell conversion. Indeed, we find no evidence that artemether promotes transdifferentiation of primary alpha cells into beta cells. Moreover, artemether reduces *Ins2* expression in primary beta cells >100-fold, suppresses glucose uptake, and abrogates beta cell calcium responses and insulin secretion in response to glucose. Our observations suggest that artemether induces general islet endocrine cell dedifferentiation and call into question the utility of artemisinins to promote alpha to beta cell transdifferentiation in treating diabetes.

eTOC Blurb

The antimalaria drug artemether has been recently shown to promote transdifferentiation of alpha cells into beta cells. Van der Meulen et al. now report loss of beta cell gene expression, glucose uptake, calcium responses and insulin secretion following stimulation in intact islets treated with a high dose of artemether.

[4]Both authors contributed equally.

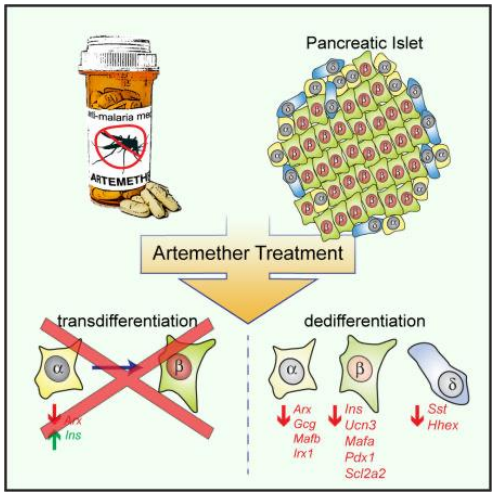
[*]Lead Contact: mhuising@ucdavis.edu

Author Contributions

Conceptualization, TvdM, SL, MOH; Methodology, SL, MOH; Software, TvdM, MWA; Formal Analysis and Data Curation, TvdM, SL, AMM, MOH; Investigation, TvdM, SL, EN, CJD, GMN, MOH; Writing, TvdM, MOH; Review and Editing, TvdM, SL, EN, CJD, GMN, AMM, MOH; Visualization, MOH; Supervision, Project Administration and Funding Acquisition, MOH.

Supplemental Information

Supplemental Experimental Procedures, 2 Figures, 1 Table, and 3 Movies can be found online.



Keywords

artemether; artemisinins; transdifferentiation; dedifferentiation; maturation; diabetes; GCaMP6; beta cell; alpha cell; pancreatic islet; neogenesis

Introduction

Type 1 and Type 2 diabetes are fundamentally different diseases, yet both are associated with a deficiency in functional beta cells. This has generated intense interest in any strategy that could regenerate functional beta cells towards a cure for diabetes. Self-replication has long been viewed as the principal mechanism of beta cell maintenance and regeneration. However, beta cell replication declines rapidly with age and no methods to promote clinically meaningful restoration of beta cell mass via proliferation have been achieved to date. Recent years have seen a surge of interest into the innate plasticity of islet endocrine cells. Following near-complete ablation of beta cells, functional beta cell mass partially recovers by transdifferentiation from alpha or delta cells, depending on the age at which beta cell ablation occurred (Chera et al., 2014; Thorel et al., 2010). Conversely, beta cells in both types of diabetes dedifferentiate into a non-functional state that no longer contributes to the functional beta cell pool, but may escape death by autoimmune destruction or exhaustion (Rui et al., 2017; Talchai et al., 2012). These observations have changed the perspective on pancreatic islets that were traditionally considered to consist of terminally differentiated cells. It is now apparent that islet endocrine cells maintain considerable plasticity and can change fate from one cell type to another. In further support of such plasticity, we demonstrated that the periphery of healthy, adult mouse islets constitutes a privileged ‘neogenic niche’ that supports the conversion of alpha cells into functionally mature beta cells (van der Meulen et al., 2017).

The notion that beta cells can be generated via transdifferentiation of alpha cells has spurred considerable interest into the underlying mechanisms that establish and control islet cell fate

that could therefore be targeted to turn alpha cells into beta cells. Understanding these mechanisms to target the transdifferentiation of alpha to beta cells would serve a dual purpose by not only regenerating beta cell mass, but simultaneously reducing the number of alpha cells, which are complicit in diabetic hyperglycemia via excess glucagon secretion (Unger and Cherrington, 2012). The successive waves of transcription factors that establish alpha and beta cell identity during embryonic development are well-described (reviewed by Shih et al., 2013). From these studies, the transcription factors Arx and Pax4 have emerged as critical determinants of alpha and beta cell identity, respectively. Whole body Arx loss-of-function leads to a lack of alpha cells that is accompanied by an increase in delta and beta cells, while Pax4-deficient mice present with the opposite phenotype: a gain of alpha cells at the expense of beta and delta cells (Collombat et al., 2003; Sosa-Pineda et al., 1997). This has led to a model where Arx and Pax4 drive opposing transcriptional programs towards alpha and beta cell identity. Indeed, overexpression of Pax4 (Collombat et al., 2009) or deletion of Arx in alpha cells (Chakravarthy et al., 2017; Courtney et al., 2013; Wilcox et al., 2013) both result in alpha to beta cell transdifferentiation. This has made Arx a central target in efforts to generate beta cells from alpha cells.

The notion that inhibition of Arx suffices to promote beta cell identity was the premise for one recent study that screened 280 clinically approved drugs and discovered that the anti-malaria drug artemether suppressed Arx function and promoted the conversion of alpha into beta cells (Li et al., 2017). However, many of the key observations in this paper were obtained in experiments that used αTC-1 alpha or Min6 beta cell lines. This is potentially problematic, as islet tumor cell lines lose properties of the primary cell type they model, while gaining traits associated with other primary islet cells (Huising et al., 2011; Oie et al., 1983). Particularly for studies that address the mechanisms underlying the establishment and maintenance of mature islet cell fate and identity, primary tissues are preferable over cell lines. Because the ramifications of this study – the repurposing of an approved drug to promote restoration of beta cells in T1D – are significant, we conducted the pertinent experiments at the basis of the conclusion that artemether drives transdifferentiation of alpha to beta cells on primary islets. Following treatment with artemether, we also observed significant inhibition of *Arx* expression, as previously reported in human islets (Li et al., 2017). However, we observed no alpha to beta cell transdifferentiation upon artemether treatment of intact mouse primary islets. Moreover, the application of artemether at the dose and duration originally used (Li et al., 2017) reduces *Ins2* expression by 100-fold, downregulates *Slc2a2* mRNA (which encodes the beta cell Glut2 glucose transporter), inhibits glucose uptake, and abolishes glucose-induced Ca^{2+} responses and insulin secretion. Our observations that artemether 1) does not turn primary alpha cells into beta cells and 2) severely affects beta cell identity and function cast doubt on the original suggestion that artemisinins could turn alpha cells into functional beta cells.

Results and discussion

The main finding behind the idea that artemisinins could drive transdifferentiation of alpha to beta cells was the observation that artemether suppressed glucagon protein content or otherwise antagonized the effects of Arx (Li et al., 2017). However, these observations were largely made in αTC-1 alpha or Min6 beta cell lines. Furthermore, artemether was suggested

to promote restoration of beta cell mass following beta cell ablation in zebrafish or rat and increase beta cell function in human islets, but none of these experiments offered direct evidence that alpha to beta transdifferentiation contributed to the observed effect. The direct evidence that was offered for alpha to beta transdifferentiation – based on lineage tracing using *Gcg*-Cre – consisted of an experiment using isolated islets of *Gcg*-Cre x *Rosa26*-stop-RFP reporter mice that were dissociated after treatment (Li et al., 2017). Slightly over 1% of all cells in control cultures was an RFP+ alpha lineage positive cell that stained for insulin, in line with the frequency of alpha to beta transdifferentiation that accumulates at a specialized neogenic niche at the periphery of adult islets (van der Meulen et al., 2017). This fraction of beta cells with an alpha cell lineage marker increased to 4% following 24 hr of culture with artemether (Li et al., 2017), although the absolute number of beta cells and beta cells with an alpha cell lineage mark in this experiment was not reported. As the local micro-environment within the islet is important in establishing and maintaining islet cell fate (van der Meulen et al., 2017), we treated intact C57BL/6 wild type mouse islets with 10 μM artemether for 72 hr as described by (Li et al., 2017) and assessed whole islet gene expression by quantitative PCR (qPCR). We observed loss of Arx mRNA expression in artemether-treated islets compared to untreated control islets cultured in parallel (Figure 1A), consistent with the original observations on human islets (Li et al., 2017). Moreover, expression of other alpha cell genes, including *Gcg*, *Mafb*, and *Irx1* was also downregulated, suggesting a general loss of alpha cell identity (Figure 1A).

Artemether inhibits Arx but fails to promote transdifferentiation of primary alpha cells

To determine whether this inhibition of Arx initiated a fate switch of alpha to beta cells in intact islets, we turned to a lineage labelling strategy combining the *Ins1*-H2b-mCherry beta cell reporter (Benner et al., 2014) with alpha lineage labelling using *Gcg*-Cre (Herrera, 2000) and *Rosa26*-stop-YFP, similar to what was done in the original paper. In triple transgenic offspring of these mice all current beta cells express nuclear mCherry, while the alpha lineage will be indelibly labelled with YFP, as previously reported (van der Meulen et al., 2017). Any alpha to beta transdifferentiation events induced by artemether in these islets should therefore present as cells that retain YFP and acquire a red nucleus following artemether exposure. These mice are thus well-suited to study transdifferentiation from primary alpha to beta cells within intact islets. We imaged islets of five different mice (3 female, 2 male) in 3D over the course of 72 hr treatment with 10 μM artemether. DMSO-treated controls cultured in parallel were included. The fact that the islets attach to the glass coverslip enabled us to re-image the same islets at multiple time points. This allowed us to track the fate of a large number of alpha cells to detect any that transdifferentiated into beta cells in response to artemether. We observed alpha to beta transdifferentiated cells at the islet periphery prior to treatment (Figure 1B), in agreement with our recent description of a 'neogenic niche' in the periphery of mouse islets (van der Meulen et al., 2017). However, of the 2344 conventional alpha cells we observed prior to artemether treatment (n = 5 mice, 8–17 islets per treatment for each animal), not a single alpha cell acquired a red nucleus that would indicate induction of *Ins1* expression and thus alpha to beta cell transdifferentiation during the course of 72 hr treatment (Figure 1C–F; movies S1). We verified on the islets we imaged of two mice (both female) that *Arx* was inhibited at the conclusion of the experiment (Supplemental Figure 1).

Artemether effectively suppresses beta cell identity

Artemether-treated islets showed an obvious pattern of speckles or fragmentation in the red channel after 72 hr, which was absent prior to treatment or in control islets at 72 hr (compare Figure 1D, E). We suspected this pattern to reflect a decline in beta cell health. Indeed, expression of *Ins1* and *Ins2* was downregulated >10-fold and >100-fold, respectively. Many mature beta cell markers, including *Ucn3*, *Mafa*, *Pdx1*, and *Slc2a2* are also significantly inhibited by 72 hr of artemether treatment (Figure 1G). Moreover, two delta cell markers, somatostatin (*Sst*) and *Hhex*, a delta cell-specific transcription factor that is important for delta cell development and function (Zhang et al., 2014), were both downregulated as well (Figure 1H). These results establish that artemether does not selectively inhibit *Arx*, but instead causes broad inhibition of alpha, beta, and delta cell-specific transcription factors, in addition to key beta cell maturity genes. These observations suggested the possibility that the continued presence of artemether had prevented the transdifferentiation of alpha cells into beta cells following *Arx* downregulation (Figure 1F). Therefore, we performed a 48 hr washout after stimulating with 10 μM artemether for 24 or 72 hr, but still did not observe marked transdifferentiation of alpha cells into beta cells (Supplemental Figure 1).

Li et al. reported significant inhibition of ARX expression by artemether in human islets, but did not show the effect of artemether treatment on the expression of insulin or any other key beta cell markers in the same experiment. We therefore reanalyzed their human single islet cell RNAseq data, which revealed no differences in *INS* expression between control and artemether-treated beta cells. However, *ARX* expression between control and artemether-treated alpha cells was also not different (Supplemental Figure 2), which is internally inconsistent with the robust inhibition of *ARX* in human islets reported by quantitative PCR in the same paper (Li et al., 2017).

Inhibition of Ins2 by artemether occurs in excess of its normal therapeutic concentration

Our observations that artemether inhibits expression of key beta cell genes would suggest that a widely used class of anti-malaria drugs impairs beta cell function. Therefore, we compared the 10 μM dose of artemether that was chosen by Li et al. and thus adopted in our study, to a 50-fold lower dose of artemether that is representative of the plasma artemether concentration in patients on a standard Artemether-lumefantrine oral anti-malarial drug regimen (four or six doses within a 48 hr period) (Lefevre et al., 2001). While artemether applied directly at islets in vitro at both doses inhibits key beta cell genes, the effects of artemether at 200 nM are significantly attenuated (Figure 1I) and 72 hr stimulation exceeds the 48 hr exposure that is common in artemether-based malaria therapies. Therefore, we do not believe that our observations of the adverse consequences of 72 hr treatment with 10 μM artemether on isolated mouse islets in vitro should give reason for pause for the safety and efficacy of artemether for the treatment of malaria, its primary indication. Artemisinins save lives and have been safely prescribed to millions of malaria patients for years (Miller and Su, 2011).

Artemether does not induce beta cell death

To determine if the robust inhibition of beta cell genes by artemether was attributable to beta cell death, we assessed the expression of a small panel of pro-apoptosis (*Bad*, *Bax* and

Txnip), and anti-apoptosis markers (*Bcl2*, *Bclx1* and *Bip*) (Danial, 2007; Minn et al., 2005) and observed no consistent changes (Figure 2A). We also did not detect an increase in the amount of cleaved caspase 3 (Figure 2B) or the number of Sytox Blue-positive dead beta cells (Figure 2C–H) that would indicate increased beta cell death following 72 hr artemether treatment.

Artemether suppresses beta cell glucose uptake

Prompted by the consistent downregulation of key beta cell genes in our experiments, we next set out to establish how 72 hr of artemether treatment would affect beta cell function. The downregulation of *Slc2a2*, which encodes the beta cell surface Glut2 glucose transporter, suggested impairment of beta cell glucose sensing. We therefore compared the rate of glucose uptake using the fluorescent glucose marker 6-NBDG in islets from *Ins1*-H2b-mCherry beta cell reporter mice (van der Meulen et al., 2017). The rate of glucose uptake was similar in freshly isolated islets and after 72 hr of culture, but was significantly suppressed in beta cells after 72 hr exposure to artemether (Figure 3, Movie S2).

Artemether inhibits the normal calcium response

Artemisinins have several proposed mechanisms of action. One of those is inactivation of the *Plasmodium falciparum* homolog of SERCA (Eckstein-Ludwig et al., 2003), a Ca^{2+}-ATPase that normally transports cytosolic free Ca^{2+} back into the endoplasmic reticulum. As calcium is required for normal insulin secretion, we sought to determine the effect of 72 hr artemether exposure to the calcium response of beta cells within primary islets using triple transgenic offspring of a cross between *Ins1*-H2b-mCherry x *Ucn3*-Cre x *Rosa26*-stop-GCaMP6. These islets exhibit efficient expression of the genetically encoded green fluorescent calcium indicator GCaMP6 across the mature beta cell lineage and independent of current Ucn3 expression status in addition to expression of nuclear mCherry across all beta cells (van der Meulen et al., 2017). Islets imaged 24 hr after isolation invariably showed a strong calcium response to an increase in glucose concentration from 5.5 to 16.8 mM, with a majority of islets exhibiting robust pulsatile calcium behavior (Figure 4A, Movie S3). After 72 hr in culture, islets from the same animal continue to respond robustly to glucose, although the pulsatility of the second phase has been replaced with sustained, but uncoordinated calcium activity that returns to baseline when ambient glucose is lowered back to 5.5 mM (Figure 4B). In sharp contrast, 72 hr of continuous treatment with 10 μM artemether completely abolished any detectable calcium response to glucose, leaving only a muted response to 30 mM KCl-induced depolarization (Figure 4C). Li et al. reported increased excitability of αTC-1 cells using Fura2 following 72 hr artemether treatment in response to excitation with 15 or 30 mM KCl (Li et al., 2017), but did not measure the calcium responses to physiologically relevant cues such as glucose on primary islet cells. Our observations that calcium responses of artemether-treated primary beta cells within intact islets to glucose and even KCl all but disappear illustrates the detrimental effect of prolonged stimulation with 10 μM artemether on beta cell function.

Artemether abrogates insulin secretion

As calcium is required for insulin secretion, we followed these functional calcium imaging experiments with studies to directly measure insulin secretion. Fresh islets secreted

significantly more insulin in response to 16.8 mM glucose, by itself or in combination with 1 nM Exendin 4, or depolarization with 30 mM KCl (Figure 4D). Islets remain responsive to these cues after 72 hr in culture (Figure 4E). However, islets from the same animals cultured in the presence of 10 μM artemether no longer responded to glucose, incretins or even depolarization by 30 mM KCl with increased insulin release (Figure 4F). These findings conflict with the reported effect of artemether on human islets, where 72 hr artemether treatment of human islets was reported to significantly increase glucose-stimulated insulin secretion (Li et al., 2017), although it was not explained how a relatively small contribution of transdifferentiating alpha cells to the existing beta cell mass promoted such a significant increase in normalized insulin secretion. We therefore repeated the artemether treatment on human islets and observed robust insulin secretion in response to 16.8 mM glucose and KCl shortly after receipt of the islets (Figure 4G) and following 72 hr of culture (Figure 4H). However, insulin secretion after 72 hr of culture in the presence of artemether completely abrogated normal insulin secretion (Figure 4I). These results are fully in line with the body of observations we reported here on mouse islets, but do not reproduce the stimulation of insulin secretion in response to artemether treatment reported previously (Li et al., 2017).

The mechanism of artemether action

Li et al. proposed a model where GABA (from neighboring beta cells) inhibits glucagon secretion from alpha cells via the artemether-induced upregulation of alpha cell GABA receptors (Li et al., 2017). Reduced extracellular glucagon concentration then would induce the loss of alpha cell identity – presumably via glucagon receptors on alpha cells. However, we did not detect expression of the glucagon receptor (*GCGR*) or most GABA receptors in these single alpha cell libraries (Supplemental figure 2), which are key parts of their model. Furthermore, it is well-known that alpha cells in *Gcgr* null mice undergo impressive hyperplasia with only limited evidence of insulin/glucagon co-positive cells (Solloway et al., 2015). This does not fit a model where loss of glucagon-mediated autocrine signaling causes alpha cells to transdifferentiate into beta cells. Li et al. also reported that expression of *ABCC8* (the sulfonylurea receptor subunit of the KATP channel) is undetectable in most single alpha cells, but is induced in alpha cells upon artemether treatment. Instead, upon reanalysis of their data we detect robust expression of *ABCC8* in all single alpha cells irrespective of artemether treatment (Supplemental figure 2). This is in keeping with the fact that KATP channel subunit expression by alpha cells is in fact well established and supported by transcriptomic analysis (Benner et al., 2014) and direct functional measurements (MacDonald et al., 2007; Zhang et al., 2013). While the mechanistic basis for its actions are unresolved, the actions of artemether at high doses could prove to be a useful tool for the experimental induction of beta cell dedifferentiation in vitro.

Summary and conclusion

The anti-malarial drug artemether was recently reported to inhibit the alpha cell-specific transcription factor Arx to promote alpha to beta transdifferentiation (Li et al., 2017). However, this study relied heavily on experiments conducted on immortalized islet cell lines and was limited in its ability to directly demonstrate transdifferentiation in primary islet cells. As we have a vested interest in transdifferentiation of islet cells and have developed state-of-the art tools to follow this process over time (van der Meulen et al., 2017), we set

out to reproduce the reported effects of artemether on alpha to beta cell transdifferentiation within intact, live islets. While we were able to reproduce the original observation that artemether treatment reduces *Arx* mRNA expression along with the expression of other alpha cell markers, we observed no evidence that this promotes alpha to beta transdifferentiation. This is at odds with the notion that Arx deletion in adult alpha cells leads to the formation of mature beta cells from alpha cells (Chakravarthy et al., 2017; Courtney et al., 2013), although others have also reported that ablation of Arx in adult alpha cells was insufficient to promote their transdifferentiation into beta cells (Wilcox et al., 2013). The lack of alpha to beta cell conversion in our experiments may be attributed to the fact that artemether inhibited *Arx* mRNA, but may not have fully ablated its expression. Moreover, 72 hr artemether treatment on primary islets caused sustained loss of identity across all islet endocrine cell types including the dedifferentiation of existing beta cells as evidenced by severe impairments in glucose uptake, calcium responses and insulin secretion. While artemether at high doses impacts islet cell function in vitro, we find no evidence to corroborate the key conclusion (Li et al., 2017) that this drug upregulates GABA receptor expression on alpha cells to inhibit glucagon secretion and thereby promote their conversion into beta cells. As sustained systemic GABA administration was reported in a separate paper to drive robust beta cell neogenesis by promoting alpha to beta cell conversions (Ben-Othman et al., 2017), it will be important to elucidate the mechanisms that explain GABA's potent beta cell neogenic actions.

STAR Methods

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Mark Huising (mhuising@ucdavis.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals—For in vitro stimulation of islets with artemether for quantitative PCR, commercial C57BL6/NHsd mice were obtained from Envigo (Indianapolis, IN). A number of transgenic mouse lines were employed. The *Rosa26*-stop-eYFP reporter mouse (B6.129X1-*Gt(ROSA)26Sor*$^{tm1(EYFP)Cos}$/J) (Srinivas et al., 2001) was used to label cells for lineage tracing. Calcium levels in the cell were visualized using the *Rosa26*-stop-GCaMP6 mouse line (B6;129S6-*Gt(ROSA)26Sortm96(CAG-GCaMP6s)Hze*/J) (Madisen et al., 2015). A Ucn3 BAC transgenic reporter mouse based on BAC clone RP23-332L13, which contains the *Ucn3* gene flanked by more than 197 kb of genomic context was used: the *Ucn3*-Cre line (B6.FVB(Cg)-Tg(*Ucn3*-cre)KF43Gsat/Mmucd (van der Meulen et al., 2017). For alpha cell lineage labeling we employed a *Gcg*-Cre mouse line (Herrera, 2000). To label beta cells we used the B6.Cg-Tg(*Ins1*-HIST1H2BB/mCherry)^{5091Mhsg}/J mouse line (Benner et al., 2014). All transgenic lines are maintained by back crossing to commercially obtained C57BL6/NHsd (Envigo). Isolated islets of mice between 3 and 13 months of age were used. Male islets were used for most experiments, with the following exceptions. The Western blot in Figure 2B was conducted on islets pooled from 8 animals of mixed sex. The 3D imaging experiments in Figure 1B–F were conducted on 3 female and 2 male mice. For 2 of these females, we verified at the conclusion of the imaging experiment that *Ins2*, *Arx* and *Ucn3*

were reduced (Figure S1A) in line with the rest of the observations we reported. We have therefore observed no indication that sex was associated with the effects of artemether, but did not formally test this. Animals were maintained in group-housing on a 12-h light/12-h dark cycle with free access to water and standard rodent chow. All animal procedures were approved by the UC Davis Institutional Animal Care and Use Committees and performed in compliance with the Animal Welfare Act and the ILAR Guide to the Care and Use of Laboratory Animals.

Primary cell cultures—Primary islets were cultured in RPMI (5.5 mM glucose, 10% FBS, pen/strep) under 5% CO_2 at 37° C in 10 cm petri dishes (*i.e.* not tissue culture treated). Islets for microscopy experiments were cultured overnight on uncoated #1.5 glass-bottom 35 mM culture dishes (MatTek Corporation, Ashland, MA).

METHOD DETAILS

Islet isolation—Islets were isolated by injecting collagenaseP (0.8 mg/mL in HBSS; Roche Diagnostics) (Invitrogen) via the common bile duct while the ampulla of Vater was clamped. The entire pancreas was collected following the injection of 2 mL collagenase solution and, after addition of an additional 2 mL of collagenase solution, was incubated at 37°C for 13 min. Pancreata were dissociated by gentle manual shaking followed by three washes with cold HBSS containing 5% NCS. The digested suspension was passed through a nylon mesh (pore size 425 μm; Small Parts Inc.), and islets were isolated by density gradient centrifugation on a Histopaque gradient (1.077 g/mL density; Sigma) for 20 min at 1400 × g without brake. Islets were collected from the interface, washed once with cold HBSS containing 5% NCS, and hand-picked several times under a dissecting microscope prior to culture in RPMI (5.5 mM glucose, 10% FBS, pen/strep).

Quantitative PCR—Islets were treated as indicated and then collected in Trizol reagent. RNA was isolated according to standard protocol and converted into cDNA using the High Capacity cDNA Archive Kit (Thermo Fisher Scientific, Waltham, MA) per the manufacturer's instructions. Primers for quantitative PCR are listed in Table S1.

Western blot—Islets were treated for 72 hr with 10 μM artemether, 750 nM thapsigargin, or DMSO (control). Thirty μl of sample treatment buffer (50 mM Tris (pH 7.5), 100 mM dithiothreitol, 2% (weight/volume) sodium dodecyl sulfate, 0.1% (weight/volume) bromophenol blue, and 10% (weight/volume) glycerol) was added to 50 islets. Antibody 9661 (Cell Signaling Technologies; diluted 1:6000) was used to detect cleaved caspase 3 and was visualized using SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher Scientific, Waltham, MA). Total protein in each lane was checked with Ponceau S solution.

3D Imaging of intact islets—To determine if alpha to beta cell transdifferentiation occurred during artemether treatment, we cultured *Ins1*-H2b-mCherry x *Gcg*-Cre x *Rosa26*-stop-YFP triple transgenic islets as described above on uncoated #1.5 glass-bottom 35 mm culture dishes (MatTek Corporation, Ashland, MA) in RPMI (10% FBS, 5.5 mM glucose, pen/strep) and followed them over a period of 72 hr in the presence or absence of 10 μM artemether. The live islets were allowed to attach onto the glass bottom overnight and

imaged in x, y, and z on a Nikon A1R+ confocal microscope using a 40x objective the next day before treatment (time = 0) and consecutively every 24 hr for 3 days after the addition of artemether. Media was also refreshed at 24 hr intervals with artemether or DMSO for control islets. For the 48 hr washout experiment, the islets were treated similarly, washed out 3x with fresh RPMI at the end of the 24 or 72 hr artemether treatments, and cultured for another 48 hr in RPMI before imaging. Z stacks of 50 micron thick were captured for each islet in resonant scanning mode. Artemether-induced beta cell death ex vivo was determined by the addition of 500 nM of the nuclear dead cell marker Sytox Blue (Thermo Fisher Scientific, Waltham, MA) 30 minutes prior to imaging. STZ-induced beta cell death was documented after 6 hr of stimulation with 5 mM Streptozotocin (EMD Millipore, Billerica, MA).

Glucose uptake—To measure glucose uptake, we incubated freshly isolated intact islets from *Ins1*-H2b-mCherry reporter mice overnight on uncoated #1.5 glass-bottom 35 mm culture dishes (MatTek Corporation, Ashland, MA) in RPMI (10% FBS, 5.5 mM glucose, pen/strep). The next day, Z-stacks of islets were continuously acquired as the non-hydrolysable glucose analog 6-NBDG (6-(*N*-(7-Nitrobenz-2-oxa-1,3-diazol-4-yl)amino)-6-Deoxyglucose; Thermo Fisher Scientific, Waltham, MA) was added at a final concentration of 0.3 mM, using a Nikon A1R+ confocal microscope in resonant scanning mode. The relative rate of glucose uptake was determined by drawing ROIs of individual *Ins1*-H2b-mCherry+ beta cells that either had or had not taken up glucose. In parallel, islets were cultured for 72 more hr in media with or without 10 μM artemether. Media and artemether were refreshed every 24 hr. Islets were allowed to attach to the glass-bottom culture dishes for the last 24 hr before applying 6-BNDG and imaging as above.

Insulin secretion—Static insulin secretion experiments were carried out on 10 wild type mouse islets per well in Krebs Ringer Buffer (KRB). Mouse islets were isolated the day prior to the secretion assay, cultured overnight in RPMI (5.5 mM glucose, 10% FBS, pen/strep). Human islets were cultured overnight in CMRL (Thermo Fisher Scientific, Waltham, MA) after receipt prior to the experiment. Islets were transferred to KRB containing 5.5 mM glucose an hr before the start of the assay. Islets were picked to a 24-wells plate for final secretion in 10% of the final assay volume. The other 90% of volume contained concentrated treatment compounds in KRB (glucose, exendin-4, KCl) as indicated. From the same batch of islets, two pools were cultured in the presence or absence of 10 μM artemether for 72 hr, with the media refreshed every 24 hr. They were then subjected to the same secretion assay as above and in the continued presence of artemether in the case of artemether-treated islets. Insulin was measured by radioimmunoassay (EMD Millipore).

Calcium responses in intact islets—We used islets from a triple transgenic offspring of a cross between *Ins1*-H2b-mCherry, *Rosa26*-stop-GCaMP6 and *Ucn3*-Cre to label the mature beta cell lineage (van der Meulen et al., 2017). Live islets were cultured overnight after the islet prep, placed on 35 mm dishes with glass bottom (#1.5; MatTek Corporation), allowed to attach overnight and imaged in x, y, z and t on a Nikon A1R+ confocal microscope using a 40x objective with a long working distance. Similar to the glucose uptake experiments, two additional sets of islets were cultured in the presence or absence of 10 μM artemether for 72 hr prior to measurement of calcium activity, with islets transferred

to fresh media and artemether every 24 hr. Treatments were continuously perfused over the islets using a Masterflex peristaltic pump at 2.5 mL per minute. Each protocol concluded with a 30 mM potassium chloride pulse to demonstrate viability and responsiveness of the islets throughout the treatment. Individual islets in individual z-planes were defined as regions of interest (ROI) and the green fluorescence intensity within the ROIs was plotted over time as a measure of calcium activity.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical Analysis—Data were analyzed by ANOVA followed by Holm-Sidak's multiple comparisons test or by t-test and are represented as mean ± SEM, with n defined in the corresponding figure legend. Differences were considered significant when $p<0.05$. Statistics were computed using Prism (GraphPad Software, La Jolla, CA).

Counting algorithm—We developed a Matlab algorithm to count the beta cells in *Ins1*-H2B-mCherry x *Gcg*-Cre x *Rosa26*-stop-eYFP islets. This takes the NIS Elements file of the islet imaged in 3D as its input and uses a combination of automated thresholding and size and shape exclusions to distinguish the red nuclei from the background. Recorded measurements are written to a csv file containing all measurements along with the sample information. We observed a close correlation between manual and automated counts of beta cells (<5% difference). YFP+ alpha cells and YFP+/mCherry co-positive transdifferentiated cells were counted manually.

Bioinformatics—Read SRA files were pulled from GEO Datasets GSE73727 and GSE84714, and converted into fastq format using the NCBI SRA toolkit. Sequence files were then aligned using STAR (Dobin et al., 2013) with default parameters to the Gencode GRCh37 24 human genome. Bigwigs were generated using STAR's wiggle output option and UCSC's Genome Utilities. Gene-level quantification was performed on all samples' sorted BAM files using featureCounts (Liao et al., 2014) default parameters, counted by Gencode defined exons, and aggregated to the gene level. Differential expression analyses were performed with edgeR (Robinson et al., 2010) using the generalized linear model approach and maximum likelihood method testing. RPKM values were generated using edgeR's *rpkm* function.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The research described in this paper was funded by an Individual Biomedical Research Award from the Hartwell Foundation and a Career Development Award from the Juvenile Diabetes Research Foundation (2-2013-54) to MOH. SL was supported by an NSF Bridge to Doctorate Program and the UC Davis Training Program in Molecular and Cellular Biology (funded in part by T32GM007377 from NIH-NIGMS). GMN was supported by a NIGMS-funded Pharmacology Training Program (T32GM099608). We thank Dr. Maike Sander for constructive feedback on the final draft of this manuscript, Dr. Aldrin Gomes for assistance with the caspase 3 Western Blot and Jessica Huang for assistance in the quantification of alpha cell numbers.

References

Ben-Othman N, Vieira A, Courtney M, Record F, Gjernes E, Avolio F, Hadzic B, Druelle N, Napolitano T, Navarro-Sanz S, et al. Long-Term GABA Administration Induces Alpha Cell-Mediated Beta-like Cell Neogenesis. Cell. 2017; 168:73–85 e11. [PubMed: 27916274]

Benner C, van der Meulen T, Caceres E, Tigyi K, Donaldson CJ, Huising MO. The transcriptional landscape of mouse beta cells compared to human beta cells reveals notable species differences in long non-coding RNA and protein-coding gene expression. BMC Genomics. 2014; 15:620. [PubMed: 25051960]

Chakravarthy H, Gu X, Enge M, Dai X, Wang Y, Damond N, Downie C, Liu K, Wang J, Xing Y, et al. Converting Adult Pancreatic Islet alpha Cells into beta Cells by Targeting Both Dnmt1 and Arx. Cell Metab. 2017

Chera S, Baronnier D, Ghila L, Cigliola V, Jensen JN, Gu G, Furuyama K, Thorel F, Gribble FM, Reimann F, et al. Diabetes recovery by age-dependent conversion of pancreatic delta-cells into insulin producers. Nature. 2014

Collombat P, Mansouri A, Hecksher-Sorensen J, Serup P, Krull J, Gradwohl G, Gruss P. Opposing actions of Arx and Pax4 in endocrine pancreas development. Genes Dev. 2003; 17:2591–2603. [PubMed: 14561778]

Collombat P, Xu X, Ravassard P, Sosa-Pineda B, Dussaud S, Billestrup N, Madsen OD, Serup P, Heimberg H, Mansouri A. The ectopic expression of Pax4 in the mouse pancreas converts progenitor cells into alpha and subsequently beta cells. Cell. 2009; 138:449–462. [PubMed: 19665969]

Courtney M, Gjernes E, Druelle N, Ravaud C, Vieira A, Ben-Othman N, Pfeifer A, Avolio F, Leuckx G, Lacas-Gervais S, et al. The inactivation of Arx in pancreatic alpha-cells triggers their neogenesis and conversion into functional beta-like cells. PLoS genetics. 2013; 9:e1003934. [PubMed: 24204325]

Danial NN. BCL-2 family proteins: critical checkpoints of apoptotic cell death. Clin Cancer Res. 2007; 13:7254–7263. [PubMed: 18094405]

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013; 29:15–21. [PubMed: 23104886]

Eckstein-Ludwig U, Webb RJ, Van Goethem ID, East JM, Lee AG, Kimura M, O'Neill PM, Bray PG, Ward SA, Krishna S. Artemisinins target the SERCA of Plasmodium falciparum. Nature. 2003; 424:957–961. [PubMed: 12931192]

Herrera PL. Adult insulin- and glucagon-producing cells differentiate from two independent cell lineages. Development. 2000; 127:2317–2322. [PubMed: 10804174]

Huising MO, Pilbrow AP, Matsumoto M, van der Meulen T, Park H, Vaughan JM, Lee S, Vale WW. Glucocorticoids differentially regulate the expression of CRFR1 and CRFR2alpha in MIN6 insulinoma cells and rodent islets. Endocrinology. 2011; 152:138–150. [PubMed: 21106875]

Lefevre G, Looareesuwan S, Treeprasertsuk S, Krudsood S, Silachamroon U, Gathmann I, Mull R, Bakshi R. A clinical and pharmacokinetic trial of six doses of artemether-lumefantrine for multidrug-resistant Plasmodium falciparum malaria in Thailand. Am J Trop Med Hyg. 2001; 64:247–256. [PubMed: 11463111]

Li J, Casteels T, Frogne T, Ingvorsen C, Honore C, Courtney M, Huber KV, Schmitner N, Kimmel RA, Romanov RA, et al. Artemisinins Target GABAA Receptor Signaling and Impair alpha Cell Identity. Cell. 2017; 168:86–100 e115. [PubMed: 27916275]

Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014; 30:923–930. [PubMed: 24227677]

MacDonald PE, De Marinis YZ, Ramracheya R, Salehi A, Ma X, Johnson PR, Cox R, Eliasson L, Rorsman P. A K ATP channel-dependent pathway within alpha cells regulates glucagon release from both rodent and human islets of Langerhans. PLoS Biol. 2007; 5:e143. [PubMed: 17503968]

Madisen L, Garner AR, Shimaoka D, Chuong AS, Klapoetke NC, Li L, van der Bourg A, Niino Y, Egolf L, Monetti C, et al. Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. Neuron. 2015; 85:942–958. [PubMed: 25741722]

Miller LH, Su X. Artemisinin: discovery from the Chinese herbal garden. Cell. 2011; 146:855–858. [PubMed: 21907397]

Minn AH, Hafele C, Shalev A. Thioredoxin-interacting protein is stimulated by glucose through a carbohydrate response element and induces beta-cell apoptosis. Endocrinology. 2005; 146:2397–2405. [PubMed: 15705778]

Oie HK, Gazdar AF, Minna JD, Weir GC, Baylin SB. Clonal analysis of insulin and somatostatin secretion and L-dopa decarboxylase expression by a rat islet cell tumor. Endocrinology. 1983; 112:1070–1075. [PubMed: 6129963]

Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26:139–140. [PubMed: 19910308]

Rui J, Deng S, Arazi A, Perdigoto AL, Liu Z, Herold KC. beta Cells that Resist Immunological Attack Develop during Progression of Autoimmune Diabetes in NOD Mice. Cell Metab. 2017; 25:727–738. [PubMed: 28190773]

Shih HP, Wang A, Sander M. Pancreas organogenesis: from lineage determination to morphogenesis. Annu Rev Cell Dev Biol. 2013; 29:81–105. [PubMed: 23909279]

Solloway MJ, Madjidi A, Gu C, Eastham-Anderson J, Clarke HJ, Kljavin N, Zavala-Solorio J, Kates L, Friedman B, Brauer M, et al. Glucagon Couples Hepatic Amino Acid Catabolism to mTOR-Dependent Regulation of alpha-Cell Mass. Cell reports. 2015; 12:495–510. [PubMed: 26166562]

Sosa-Pineda B, Chowdhury K, Torres M, Oliver G, Gruss P. The Pax4 gene is essential for differentiation of insulin-producing beta cells in the mammalian pancreas. Nature. 1997; 386:399–402. [PubMed: 9121556]

Srinivas S, Watanabe T, Lin CS, William CM, Tanabe Y, Jessell TM, Costantini F. Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. BMC developmental biology. 2001; 1:4. [PubMed: 11299042]

Talchai C, Xuan S, Lin HV, Sussel L, Accili D. Pancreatic beta cell dedifferentiation as a mechanism of diabetic beta cell failure. Cell. 2012; 150:1223–1234. [PubMed: 22980982]

Thorel F, Nepote V, Avril I, Kohno K, Desgraz R, Chera S, Herrera PL. Conversion of adult pancreatic alpha-cells to beta-cells after extreme beta-cell loss. Nature. 2010; 464:1149–1154. [PubMed: 20364121]

Unger RH, Cherrington AD. Glucagonocentric restructuring of diabetes: a pathophysiologic and therapeutic makeover. J Clin Invest. 2012; 122:4–12. [PubMed: 22214853]

van der Meulen T, Mawla AM, DiGruccio MR, Adams MW, Nies V, Dolleman S, Liu S, Ackermann AM, Caceres E, Hunter AE, et al. Virgin beta cells persist throughout life at a neogenic niche within pancreatic islets. Cell Metab. 2017; 25:911–926.e6. [PubMed: 28380380]

Wilcox CL, Terry NA, Walp ER, Lee RA, May CL. Pancreatic alpha-cell specific deletion of mouse Arx leads to alpha-cell identity loss. PLoS One. 2013; 8:e66214. [PubMed: 23785486]

Zhang J, McKenna LB, Bogue CW, Kaestner KH. The diabetes gene Hhex maintains delta-cell differentiation and islet function. Genes & development. 2014; 28:829–834. [PubMed: 24736842]

Zhang Q, Ramracheya R, Lahmann C, Tarasov A, Bengtsson M, Braha O, Braun M, Brereton M, Collins S, Galvanovskis J, et al. Role of KATP channels in glucose-regulated glucagon secretion and impaired counterregulation in type 2 diabetes. Cell metabolism. 2013; 18:871–882. [PubMed: 24315372]

Highlights

- Artemether does not induce the transdifferentiation of α cells into β cells
- High doses of artemether dedifferentiate islet cells without inducing death
- Artemether not only inhibits Arx and Gcg, but inhibits Ins2 > 100-fold
- Artemether suppresses glucose uptake and prevents insulin secretion

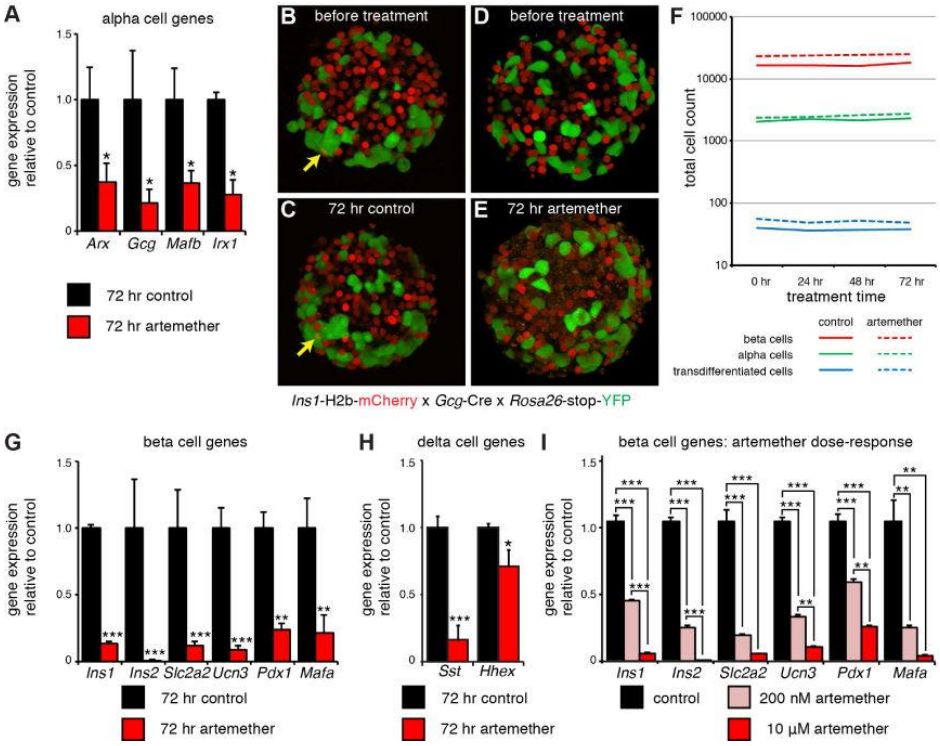


Figure 1. Artemether does not promote the transdifferentiation of alpha to beta cells but instead suppresses overall islet cell identity

(A) Real time quantitative PCR analysis of *Arx*, *Gcg*, *Mafb*, and *Irx* gene expression in artemether treated islets (n=4 replicates). *p<0.05.

(B) 3D reconstruction of a representative image of an islet from an *Ins1*-H2B-mCherry x *Gcg*-Cre x *Rosa26*-stop-YFP triple transgenic reporter mouse at the onset of the experiment. Arrow indicates an mCherry/YFP co-positive cell that represents a spontaneous alpha to beta transdifferentiated cell.

(C) Islet in (B), re-imaged after 72 hr of incubation. Arrow indicates the same alpha to beta transdifferentiated cell that was present prior to the 72 hr culture window.

(D) 3D reconstruction of a representative image of another islet from the same mouse as in (B) prior to artemether treatment.

(E) Islet in (D), re-imaged after 72 hr of culture in the presence of 10 μM artemether. No alpha to beta transdifferentiated cells are present, note the appearance of a 'speckled' background across the entire islet. See also Movie S1.

(F) Quantification of the total number of alpha, beta and alpha to beta transdifferentiated cells. The same islets were imaged repeatedly in 3D at 24 hr intervals. n = 5 mice, 8–17 islets per animal for each treatment.

(G) Real time quantitative PCR analysis of the expression of a panel of beta cell genes in artemether treated islets (n=4 replicates). *p<0.05, **p<0.01, ***p<0.001.

(H) Real time quantitative PCR analysis of the expression of a panel of delta cell genes in artemether treated islets (n=4 replicates). *p<0.05, **p<0.01, ***p<0.001.
(I) Real time quantitative PCR analysis of a panel of beta cell genes in islets treated for 72 hr with 200 nM or 10 μM artemether (n=4 replicates). *p<0.05, **p<0.01, ***p<0.001.

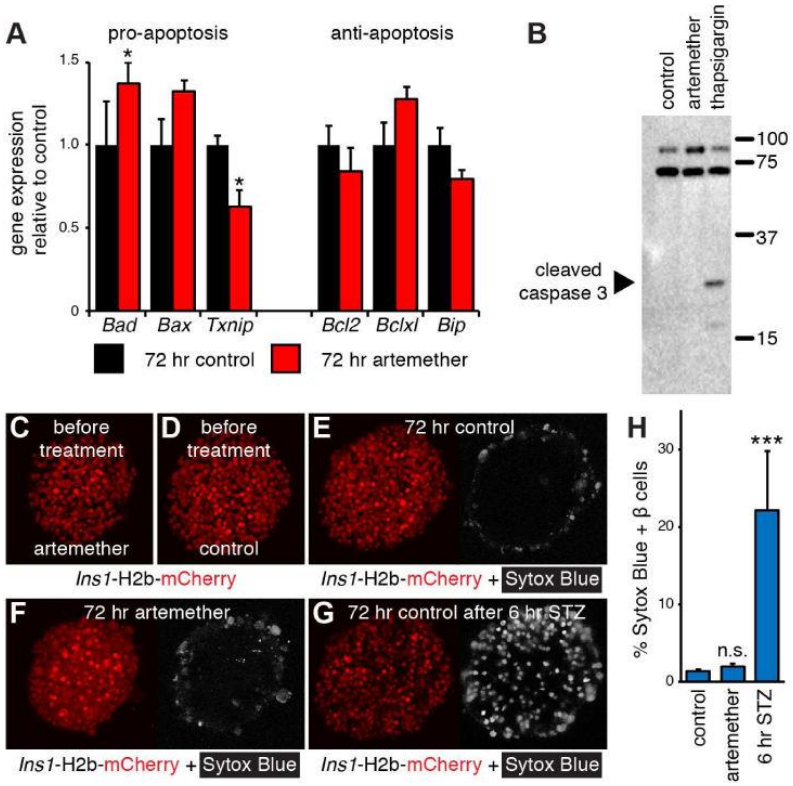


Figure 2. Artemether does not induce beta cell death

(A) Real time quantitative PCR analysis of the expression of a panel of pro- and anti-apoptotic genes in artemether treated islets (n=4 replicates). *p<0.05.

(B) Western blot analysis of the apoptosis marker cleaved caspase 3 in islets treated for 72 hr with 10 μM artemether. Thapsigargin (750 nM) was used as a positive control.

(C) 3D reconstruction of a representative *Ins1*-H2b-mCherry islet before artemether treatment.

(D) 3D reconstruction of a representative *Ins1*-H2b-mCherry control islet before culture.

(E) Islet in (D) re-imaged after 72 hr in culture. The nuclear dead cell marker Sytox Blue was added at 72 hr.

(F) Islet in (C) re-imaged after 72 hr in culture with 10 μM artemether. The nuclear dead cell marker Sytox Blue was added at 72 hr.

(G) 3D reconstruction of the same islet as in (E) after 6 hr in the presence of STZ to induce beta cell death.

(H) Quantification of the fraction of Sytox Blue positive beta cells after 72 hr in control (12 islets), and artemether-treated cultures (16 islets) compared to control islets exposed to STZ for an additional 6 hr (5 islets). ***p<0.001 compared to control.

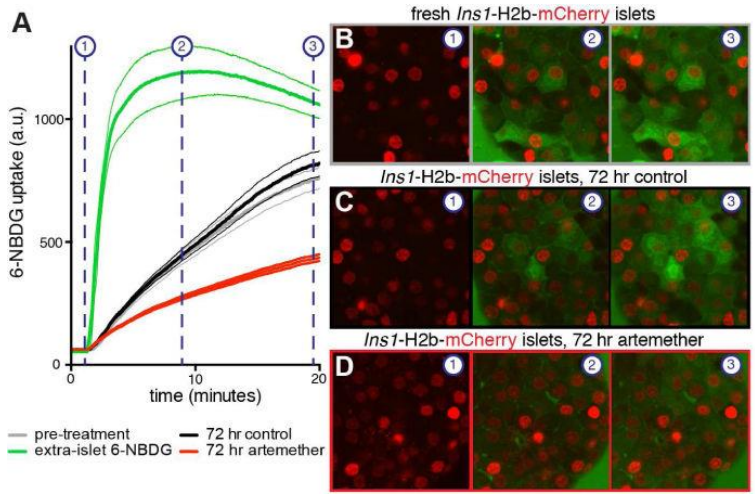


Figure 3. Artemether inhibits glucose uptake by beta cells

(A) Comparison of the rate of uptake of the glucose analog 6-NBDG in islets of *Ins1*-H2b-mCherry beta cell reporter mice in isolated islets after isolation (black), after 72 hr in culture in the presence of DMSO (gray) or artemether (red). 6-NBDG signal in the media outside of the islet is quantified as a reference (green). Averages and 95% confidence intervals are given for 19–26 individual beta cells of 2–3 islets for each treatment.

(B) Video stills of 6-NBDG uptake in a freshly isolated *Ins1*-H2b-mCherry islet. See also Movie S2.

(C) Video stills of 6-NBDG uptake in an *Ins1*-H2b-mCherry control islet after 72 hr of culture.

(D) Video stills of 6-NBDG uptake in an *Ins1*-H2b-mCherry islet after 72 hr of artemether treatment.

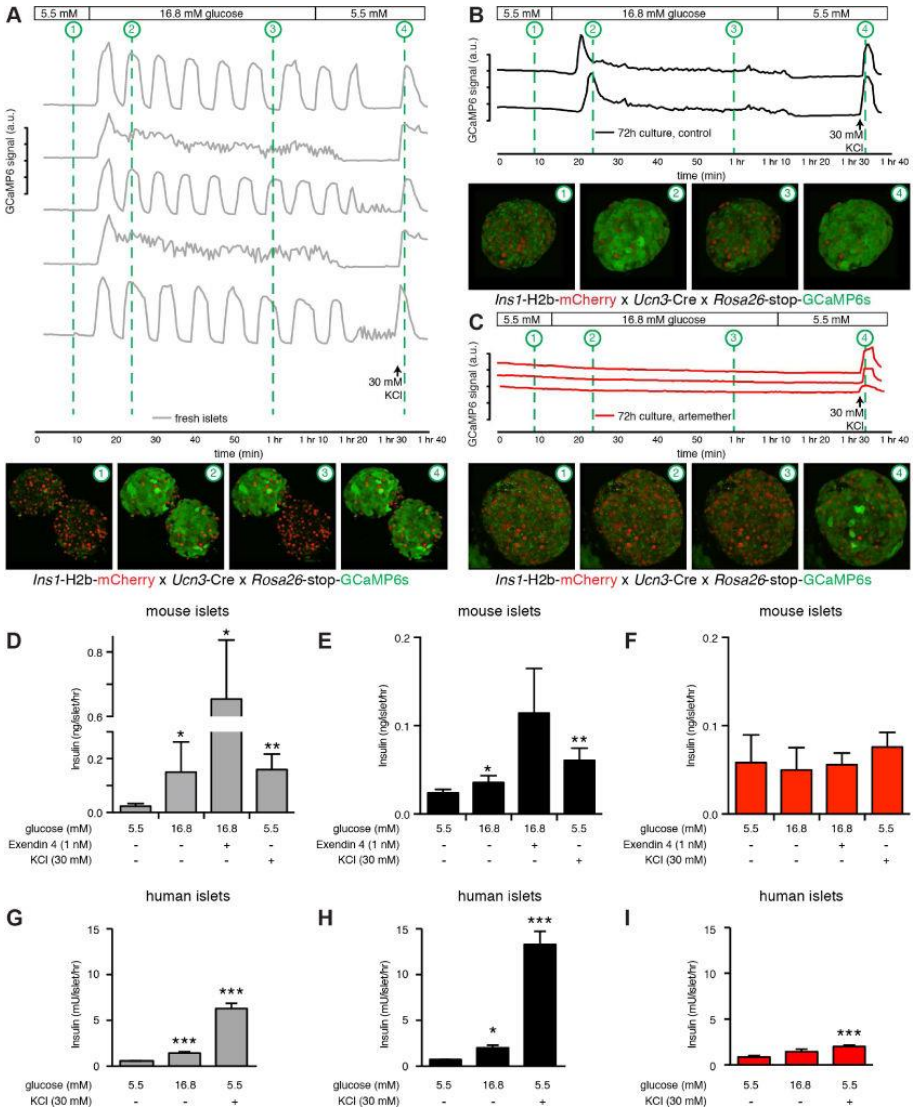


Figure 4. Artemether abrogates beta cell calcium responses and insulin secretion

(A) The calcium responses of freshly isolated islets from *Ins1*-H2B-Chy x *Ucn3*-Cre x *Rosa26*-stop-GCaMP6 triple transgenic mice were imaged in 3D during a standard glucose stimulation protocol from 5.5 to 16.8 to 5.5 mM glucose, followed by a brief depolarization by 30 mM KCl to demonstrate responsiveness of the islets throughout the experiment. Each trace represents the combined response of a single islet. Thumbnails provide snapshots of the calcium response at key points during the trace for 2 of the 5 islets imaged. See also Movie S3.

(B) Islets from the same animal and subjected to the same stimulation protocol as in (A), but after 72 hr in culture. First phase secretion is intact, but the second phase is muted and pulsatility is lost.
(C) Islets from the same animal and subjected to the same stimulation protocol as in (A), but after 72 hr in culture in the presence of 10 μM artemether. The glucose response is completely abolished and only a modest response to forced depolarization with 30 mM KCl can be detected.
(D) Insulin secretion from freshly isolated mouse islets in response to glucose, glucose + Exendin 4, or 30 mM KCl (n=4 replicate treatments). *p<0.05, **p<0.01 compared to 5.5 mM glucose. (E) Insulin secretion on islets from the same batch as in (D), but after 72 hr in culture (n=4 replicate treatments). *p<0.05, **p<0.01 compared to 5.5 mM glucose.
(F) Insulin secretion on islets from the same batch as in (D), but after 72 hr of treatment with 10 μM artemether (n=4 replicate treatments). *p<0.05, **p<0.01 compared to 5.5 mM glucose.
(G) Insulin secretion from freshly isolated human islets in response to glucose or 30 mM KCl (n=8 replicate treatments). *p<0.05, **p<0.01, ***p<0.001 compared to 5.5 mM glucose.
(H) Insulin secretion on islets from the same batch as in (G), but after 72 hr in culture (n=6 replicate treatments). *p<0.05, **p<0.01, ***p<0.001 compared to 5.5 mM glucose.
(I) Insulin secretion on islets from the same batch as in (G), but after 72 hr of treatment with 10 μM artemether (n=6 replicate treatments). *p<0.05, **p<0.001, ***p<0.001 compared to 5.5 mM glucose.

VI. Chapter 6: Navigating the Depths and Avoiding the Shallows of Pancreatic Islet Cell Transcriptomes

Alex M. Mawla and Mark O. Huising

Diabetes, 2019. 68(7):1380-1393. doi: 10.2337/dbi18-0019

Contributions to Jointly Authored Works: As first author of this manuscript, I was responsible for the entirety of the formal meta-analysis of all data, and downstream computational approaches. I was responsible for creating all figures, and co-wrote the manuscript with my professor, Dr. Mark O. Huising. Furthermore, I also created an interactive page on our lab website for other colleagues in the field to query these data and generate visuals.

Significance of Research: This review was intended to both highlight the abilities of single-cell sequencing, but to also caution colleagues in the field from liberally interpreting results. This paper was well-received and was the basis for my being the first graduate student invited to present, and then receive the Student Research Award at the AAAS Pacific Division Conference in June 2019. Furthermore, it was the basis of my Sigma Xi Membership nomination, later leading me to become Chapter Officer at the University of California, Davis branch.

Citations: 38

Navigating the Depths and Avoiding the Shallows of Pancreatic Islet Cell Transcriptomes

Alex M. Mawla[1] and Mark O. Huising[1,2]

Diabetes 2019;68:1380–1393 | https://doi.org/10.2337/dbi18-0019

Islet gene expression has been widely studied to better understand the transcriptional features that define a healthy β-cell. Transcriptomes of FACS-purified α-, β-, and δ-cells using bulk RNA-sequencing have facilitated our understanding of the complex network of cross talk between islet cells and its effects on β-cell function. However, these approaches were by design not intended to resolve heterogeneity between individual cells. Several recent studies used single-cell RNA sequencing (scRNA-Seq) to report considerable heterogeneity within mouse and human β-cells. In this Perspective, we assess how this newfound ability to assess gene expression at single-cell resolution has enhanced our understanding of β-cell heterogeneity. We conduct a comprehensive assessment of several single human β-cell transcriptome data sets and ask if the heterogeneity reported by these studies showed overlap and concurred with previously known examples of β-cell heterogeneity. We also illustrate the impact of the inevitable limitations of working at or below the limit of detection of gene expression at single cell resolution and their consequences for the quality of single–islet cell transcriptome data. Finally, we offer some guidance on when to opt for scRNA-Seq and when bulk sequencing approaches may be better suited.

Type 1 diabetes (T1D) and type 2 diabetes (T2D) affect roughly 14% of the population and are the seventh leading causes of death in the U.S. (1). T1D is characterized by autoimmune-mediated β-cell destruction within the pancreas. T2D is characterized by increased peripheral insulin resistance, which eventually unmasks and/or precipitates β-cell dysfunction (2). Consequently, the field has mostly focused on β-cells, despite the fact that pancreatic islets of Langerhans contain at least five different hormone-secreting endocrine cell types, supported by a constellation of auxiliary cells, whose clustering supports coordinated secretion of insulin and glucagon to maintain nutrient homeostasis (3–5). The spatial distribution of these cells within islets varies between human and mouse models, but β-cells are the most abundant endocrine cell type in both species, followed by α-cells, δ-cells, and a lower number of γ-/pancreatic polypeptide cells and ε-cells (6,7).

While islet isolation is a routine procedure, the close association of all of these endocrine and auxiliary cell types within the islet has long complicated the isolation and purification of homogeneous populations of each islet cell type. Consequently, changes in gene and protein expression within intact isolated islets were often attributed to β-cells, as they are numerically the most abundant islet cell type within the islet. Clearly, this ignores the fact that multiple additional endocrine cells, as well as endothelial cells, macrophages, glia, fibroblasts, and pericytes collectively make up the pancreatic islet (8–11). β-Cell dysregulation and dysfunction are a prominent factor in disrupted insulin secretion and blood glucose control, but major functional and transcriptional changes also occur in α-cells (12,13), as well as vasculature (14), that are difficult to detect or distinguish from changes to β-cells at the level of the intact islet.

RESOLVING DIFFERENCES BETWEEN ISLET ENDOCRINE CELLS

Purification of β-cells had initially been achieved on the basis of autofluorescence (15), an approach that works reasonably well. Subsequent strategies have improved this approach by generating transgenic reporter lines that express fluorescent markers such as GFP or mCherry

[1]Department of Neurobiology, Physiology and Behavior, College of Biological Sciences, University of California, Davis, Davis, CA
[2]Department of Physiology and Membrane Biology, School of Medicine, University of California, Davis, Davis, CA

Corresponding author: Mark O. Huising, mhuising@ucdavis.edu

Received 27 October 2018 and accepted 29 April 2019

This article contains Supplementary Data online at http://diabetes.diabetesjournals.org/lookup/suppl/doi:10.2337/dbi18-0019/-/DC1.

specifically in β-cells (16,17). However, neither strategy can copurify pure α- or δ-cells. Several groups have recently resolved this limitation by generating combinations of transgenic reporter mice that made it possible to isolate pure populations of α-, β-, and δ-cells from the same islet by FACS. This has enabled the generation of comprehensive transcriptomes of FACS-purified pools of mouse α-, β-, and δ-cells with >99% purity (17–19). For human islets, the problem of purifying α- and β-cells was resolved independently by the generation of a panel of antibodies that enabled the purification of human α- and β-cells with approximately 90% purity (20–22). The ability to purify human islet cell types has allowed for further exploration in human islet transcriptomics and the subsequent identification of genes that encode proteins exclusively expressed in β-cells (23,24). However, cell-surface markers are currently unable to isolate human δ-cells or other, more rare islet endocrine cells with reasonable purity by flow cytometry.

PREVIOUSLY ESTABLISHED HETEROGENEITY

In addition to the heterogeneity that results from the clustering of many different cell types within a functional islet, it has long been evident that considerable heterogeneity exists within the β-cell population (21,25–29), and likely within non-β populations of islet cells as well. Functional heterogeneity among β-cells occurs with regard to the glucose threshold and insulin secretory response of individual β-cells (25,26,30). Heterogeneity in the expression of a number of markers, such as the peptide hormone neuropeptide Y (NPY), tyrosine hydroxylase (TH), and Dickkopf-3, by individual β-cells has also been reported (31–34).

More recently, a series of articles have rekindled interest in β-cell heterogeneity, with the description of Flattop (Fltp)-expressing β-cells (27), ST8SIA1/CD9-positive β-cells (21), Ucn3/Glut2-negative "virgin" β-cells (35,36), "bottom" β-cells (named for the bottom of two FACS gates used to isolate them [37]), and senescent β-cells (38). This paints a landscape of β-cell heterogeneity that features changes in marker expression over the life span of the β-cell and/or in relation to the functional state of the β-cell in health and disease. Understanding of this heterogeneity would benefit greatly from transcriptional read-outs at single-cell resolution. Indeed, a number of recent articles have reported on single-cell transcriptomes of mouse and human primary islet cells (39–49).

The great promise of sequencing at single-cell resolution is that this should resolve the considerable heterogeneity that exists among the individual β-cells that come together in the islet. Here, we take stock of what these recent single-cell studies have added to our understanding of islet cell biology. We do so by asking two basic questions: *1*) Have individual single-cell sequencing studies that are similar in design resulted in comparable outcomes? *2*) Did single-cell approaches recapitulate well-known and validated examples of β-cell heterogeneity? In addressing these two straightforward questions, we discuss areas where single-cell approaches have made clear and tangible contributions to our field. However, we also document examples where single-cell sequencing approaches may fall short of the unrealistically high expectations that exist for this approach. We review and clarify some of the underlying reasons that may have contributed to this disconnect. Finally, we offer some guidance on when a single-cell approach is preferred and what question may be better resolved using a bulk RNA sequencing (RNA-Seq) approach.

VALIDATION OF NOVEL β-CELL HETEROGENEITY IDENTIFIED IN SINGLE-CELL RNA-Seq STUDIES

As a first step in assessing the reported heterogeneity by recent single-cell (sc)RNA-Seq studies of islets, we compared β-cell heterogeneity that was highlighted by the authors of several recent scRNA-Seq studies of human pancreatic islets (43,44,46–48). The overall design for each of these studies was to sequence dissociated islet cells of human subjects at single-cell resolution, even though each of these studies inevitably differed in the technical details and the sequencing methodologies that were used (Supplementary Table 1). Nevertheless, given the agreement in the overall design, we reasoned that true heterogeneity should emerge despite the inevitable variations in methodologies and should be reproducible across individual human donors in each of these studies. After all, if this were not true, all observations that have emerged from scRNA-Seq studies of human islets to date would be limited only to the deceased islet donors who were the subject of these studies and would not extend to the general population.

To our surprise, not a single gene was highlighted after manual annotation by the authors as heterogeneously expressed across all five studies, and only a few genes were highlighted independently by up to three scRNA-Seq studies of human β-cells (Supplementary Fig. 1*A* and Supplementary Table 2). This observation can be interpreted in two possible ways. It may be that the extent of β-cell heterogeneity is so great that the human β-cell scRNA-Seq studies to date have effectively undersampled this heterogeneity. The alternative explanation is that the detection of variation in gene expression across single β-cells is dominated by noise resulting from operating at or below the limit of detection of gene expression in single-cell expression, causing false negatives to dominate the list of heterogeneously detected β-cell genes. Moreover, the short list of heterogeneously detected genes in β-cells was notably lacking genes encoding proteins known to demonstrate heterogeneous expression patterns among β-cells (e.g., NPY, TH, UCN3, DKK3). This raises the question whether scRNA-Seq approaches were able to accurately detect expression of established markers of heterogeneity among β-cells. While many of these studies have taken their analyses beyond single-gene transcriptomes, i.e., gene set enrichment and multiparametric pathway

analyses, our primarily focus was to evaluate whether heterogeneity could accurately be recapitulated.

VALIDATION OF NOVEL HETEROGENEITY IN β-CELLS

To determine the degree with which different scRNA-Seq studies detect overlap in β-cell heterogeneity, we conducted a meta-analysis of the five human scRNA-Seq studies. Given that the overall design of each of these studies was essentially the same and that differences between these studies were limited to the inevitable variation across human donors and variations in the sequencing methods and analyses pipelines, we expected to observe considerable overlap between each data set (Supplementary Table 1). To further reduce any differences, we downloaded and reanalyzed the raw data from each of the studies, generating an integrated analysis that resolved each of the major pancreas populations (Fig. 1*A–L* and Approach & Tools in Supplementary Data). We also verified that clustering was not driven by the platform used or by donor (Supplementary Fig. 2*A* and *B*).

We identified two subclusters within the β-cell population (Fig. 1*A*). By differential testing we identified 52 genes that drove variation between the two subpopulations ($P \leq 0.05$; Supplementary Table 3). Notably, *G6PC2*, *MAFA*, and *NPY* were detected within this list, as were several non-β endocrine and acinar cell markers, such as *GCG*, *SST*, *PPY*, *PRSS*, *SOD2*, and *PDK4*. Among the 52 genes were three genes—retinol-binding protein 4 (*RBP4*) (46–48), delta-like noncanonical Notch ligand 1 (*DLK1*) (43,44,46), and homocysteine-responsive endoplasmic reticulum-resident ubiquitin-like domain member 1 (*HERPUD1*) (44)—that had previously been self-reported as heterogeneously expressed by one of the original human scRNA-Seq studies. While multiparameter signature analyses, such as gene set pathway testing, can be a powerful tool to make meaning out of subtle changes across varying genes, our small list of 52 did not suffice for further downstream analysis.

Surprised by the fact that only a limited number of genes drive variation between these two β-cell subpopulations, and the fact that non-β markers featured prominently in this list, we limited our analysis on only the β-cells from healthy donors (Supplementary Fig. 3*A*). Because these β-cells are more closely related to each other than, for example, to α- and δ-cells, clustering is confounded significantly by study-related confounders such as sequencing platform, genetic variation among donors, and variations in islet collection and culture parameters, suggesting that these contributions outweighed the contributions of true biological heterogeneity to clustering of β-cells (Supplementary Fig. 3*B* and *C*). Indeed, in a Venn diagram of the 2,000 genes that drove clustering of β-cell subpopulations from healthy donors for each individual study, only a distinct minority of 24 genes (1.2%) emerged as common drivers of heterogeneous expression among β-cells across all five human β-cell scRNA-Seq studies (Supplementary Table 4 and Supplementary Fig. 3*D–F*). Moreover, approximately half of the genes that drove clustering of β-cells were unique to that particular data set and did not contribute to β-cell subpopulation clustering in each of the other human β-cell scRNA-Seq data sets (Supplementary Fig. 3*D*). *NPY* was the only gene encoding a known β-cell heterogeneity marker on this list.

To confirm these results, we selected 10 of these 24 genes that had a low to high range of abundance to evaluate how the expression of these genes compared across the same five studies. We observed varying fractions of β-cells with detectable expression (counts per million [CPM] >1) (Supplementary Fig. 3*E*) and comparable distribution of gene expression in violin plots (Supplementary Fig. 3*F*) for the majority of these 10 genes. Overrepresentation of *INS* (Supplementary Fig. 4*A*) may have caused poor library complexity by reducing the detection of other, less abundant genes below the detection limit. This is a general drawback of scRNA-Seq and explains in part why the number of detectable genes in each single cell is several-fold lower than the number of detectable genes in the same sample processed for bulk RNA-Seq (50,51). Two of the studies in our meta-analysis had the foresight to include in their experimental design the parallel processing of bulk samples from the same donors that were used to generate scRNA-Seq libraries, although cold-ischemic and postisolation culture times, as well as processing and dissociation methods, varied between them (43,47). This revealed that the average number of genes detectably expressed in whole-islet bulk islet samples (CPM >1) approximates 15,000, while the number of genes that are detectably expressed in each single β-cell ranges from 2,000 to 6,000 (Supplementary Table 1 and Supplementary Fig. 4*B* and *C*), with the subset of genes that is detected in each single β-cell in large part determined by chance (52). To illustrate this heterogeneous detection, we plotted the fraction of single human β-cells with detectable expression for all genes ranked in descending order of abundance (Fig. 2 and Supplementary Fig. 5*A–E*). This revealed a clear correlation between the average level of gene expression and the rate of detection in single human β-cells across all five studies, with more abundant genes detected in a larger fraction of β-cells. However, only an exceedingly small number of 86 genes (0.46% of all detectable genes; range 1 [0.005%]–153 [0.83%]) on average was detectable across all single β-cells in any given study. This is an obvious concern, as even the most conservative estimates place the number of housekeeping genes—genes required at all times in each cell—at several hundred (53). Stated differently, for the large majority of genes (>99.5%) heterogeneous detection in single β-cells is the norm (Fig. 2 and Supplementary Fig. 5*A–E*). It is highly unlikely that all of these genes are truly heterogeneously expressed in β-cells. Instead, this observation indicates that heterogeneous detection of expression in single cells

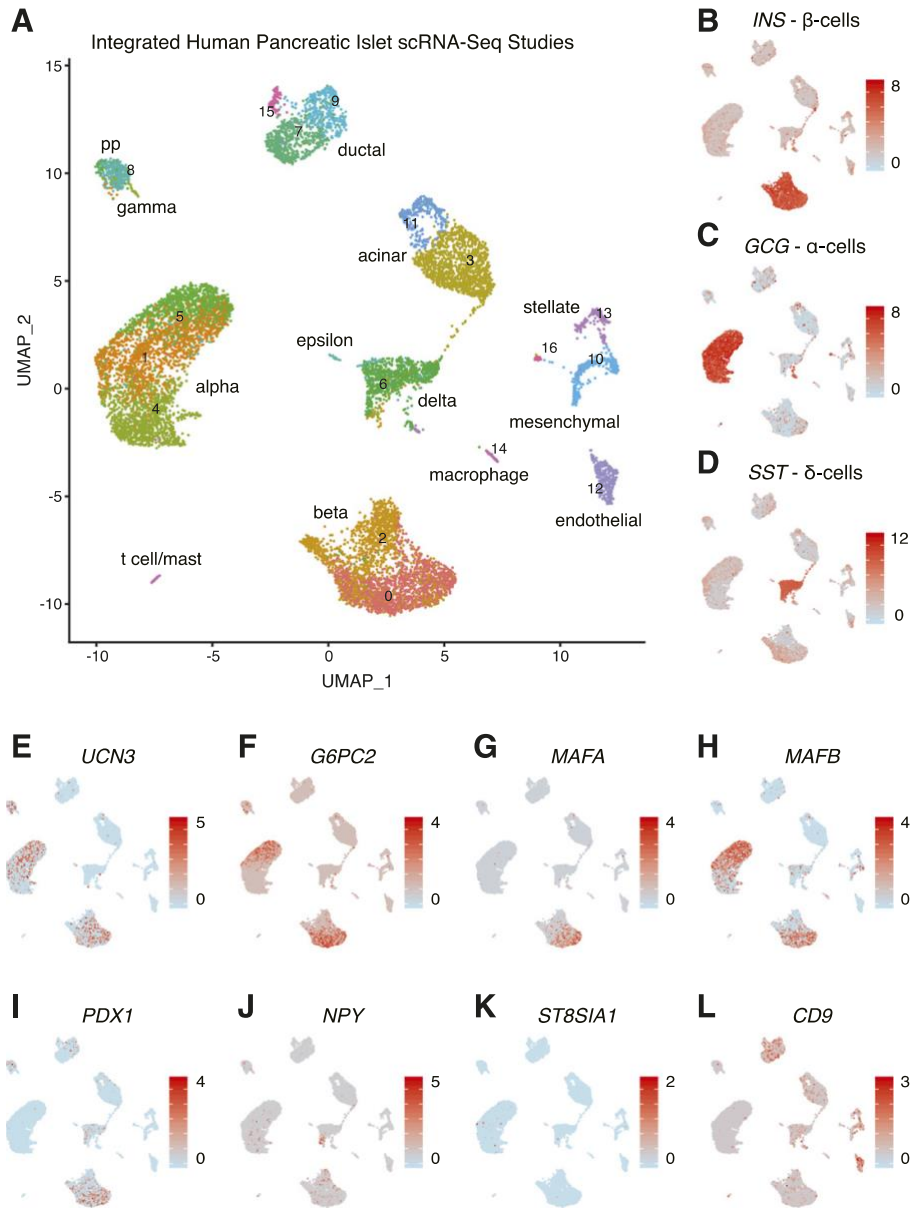


Figure 1—Integrated analysis of five human pancreatic islet scRNA-Seq studies across healthy donors. *A*: Dimensional reduction through uniform manifold approximation and projection across the five human studies identified clear clusters based on cell identity (see Supplementary Data, Approach & Tools for details). Two subpopulation clusters emerged from this analysis within the β-cell group, prompting for differential expression testing that identified 52 genes as significantly driving heterogeneity between the two (Supplementary Fig. 3*A*) ($P \leq 0.05$). *B–D*: Dimensional reduction of our clustered population to confirm β-, α-, and δ-cell clusters based on hallmark hormone expression. *E–L*: Exploring gene presence across selected, established markers identified in β-cell heterogeneity. PP, pancreatic polypeptide. Data from 43,44,46–48.

may be a poor predictor of actual single β-cell expression. Collectively, these observations suggest that heterogeneity of detection that is observed across single human β-cells may largely reflect the low fidelity of detection that is a consequence of operating at or below the limit of detection for a majority of transcripts. This may also have driven the limited overlap among the shared set of genes that emerged as common contributors to β-cell clustering across the five human scRNA-Seq studies we assessed in our analysis. Nevertheless, accumulated across all cells in a pool of single

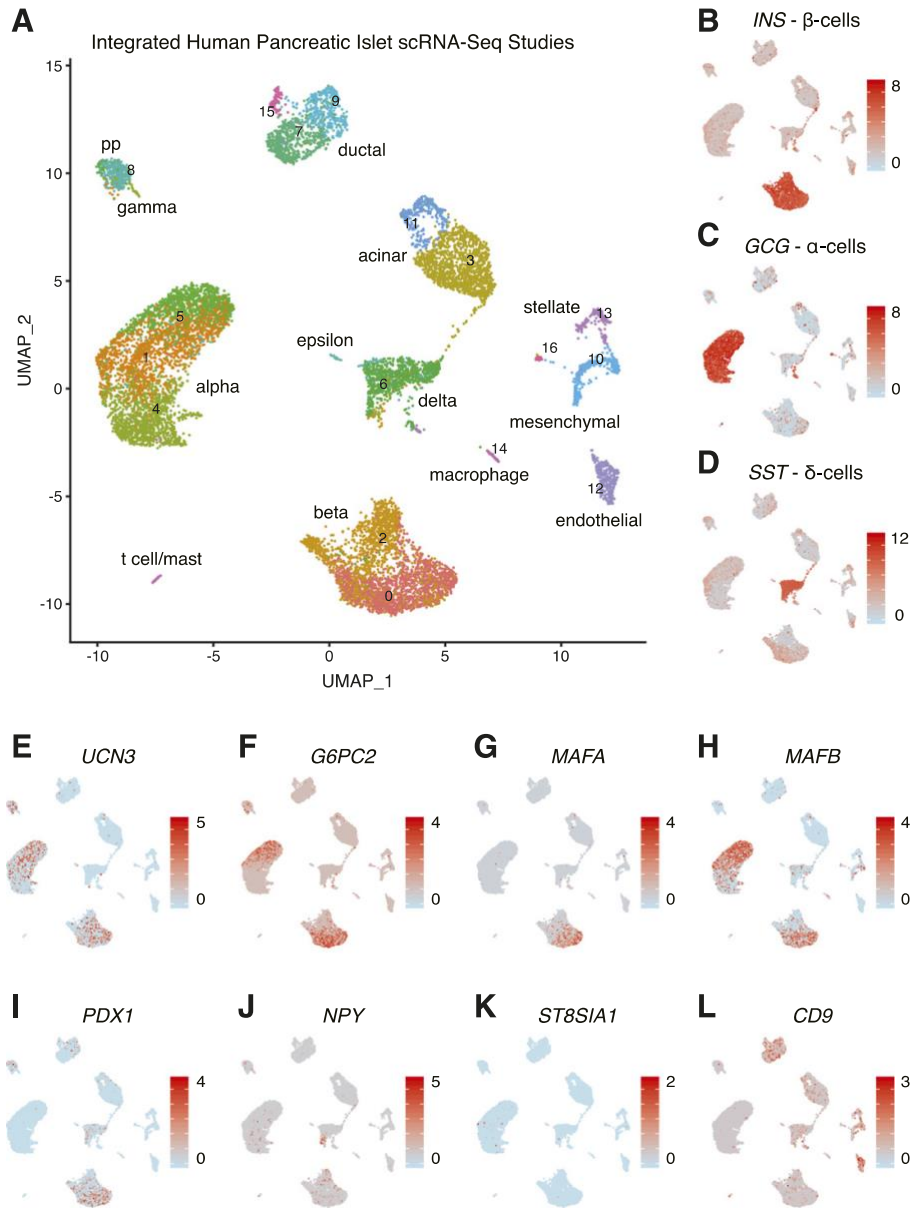


Figure 1—Integrated analysis of five human pancreatic islet scRNA-Seq studies across healthy donors. *A*: Dimensional reduction through uniform manifold approximation and projection across the five human studies identified clear clusters based on cell identity (see Supplementary Data, Approach & Tools for details). Two subpopulation clusters emerged from this analysis within the β-cell group, prompting for differential expression testing that identified 52 genes as significantly driving heterogeneity between the two (Supplementary Fig. 3*A*) ($P \leq 0.05$). *B–D*: Dimensional reduction of our clustered population to confirm β-, α-, and δ-cell clusters based on hallmark hormone expression. *E–L*: Exploring gene presence across selected, established markers identified in β-cell heterogeneity. PP, pancreatic polypeptide. Data from 43,44,46–48.

may be a poor predictor of actual single β-cell expression. Collectively, these observations suggest that heterogeneity of detection that is observed across single human β-cells may largely reflect the low fidelity of detection that is a consequence of operating at or below the limit of detection for a majority of transcripts. This may also have driven the limited overlap among the shared set of genes that emerged as common contributors to β-cell clustering across the five human scRNA-Seq studies we assessed in our analysis. Nevertheless, accumulated across all cells in a pool of single

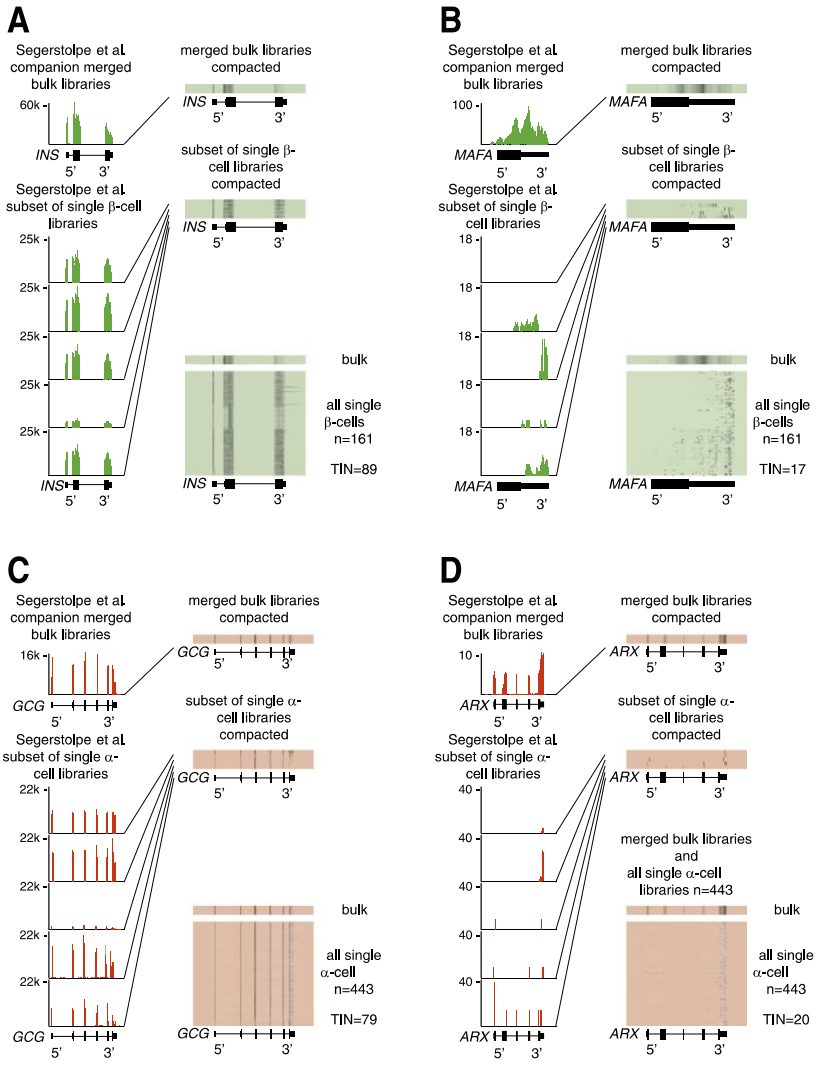


Figure 3—Visualizing the detection across the gene model between single and bulk sequencing. *A*: UCSC Genome Browser comparing the read coverage of insulin (INS) between the Segerstolpe et al. (47) healthy single β-cell population (n = 161) and their companion bulk islet libraries. *B*: MafA is an established β-cell marker that is poorly detected in single human β-cells. *C*: The highly abundant expression of GCG uniformly covered in single human α-cells (n = 443). *D*: The important α-cell transcription factor ARX, in contrast, is captured poorly in healthy single α-cells.

MAFA capture across all five studies, less than 50% of cells in the β-cell cluster had detectable expression as determined through Seurat (56) (Fig. 1*G*). Using the same approach across the study's 443 single α-cell libraries revealed uniform coverage of the abundant *GCG* transcript in each single α-cell library (Fig. 3*C*). However, the transcription factor *ARX*, which is required for α-cell identity (57,58), was not detected at all in 18% of α-cells with evidence of significant 3′ bias in incomplete coverage in the α-cells with detectable *ARX* expression (Fig. 3*D*).

REPRODUCING KNOWN β-CELL GENE EXPRESSION

These observations raise the question of whether the heterogeneous detection of mRNA expression in single β-cells reflects true biological heterogeneity in gene expression or instead is a product of the inherent limitations of scRNA-Seq. Therefore, we queried if single β-cell transcriptomes accurately detected genes encoding for proteins that are required by every single β-cell, as well as genes that encode for proteins with well-documented and validated heterogeneous expression across the β-cell population

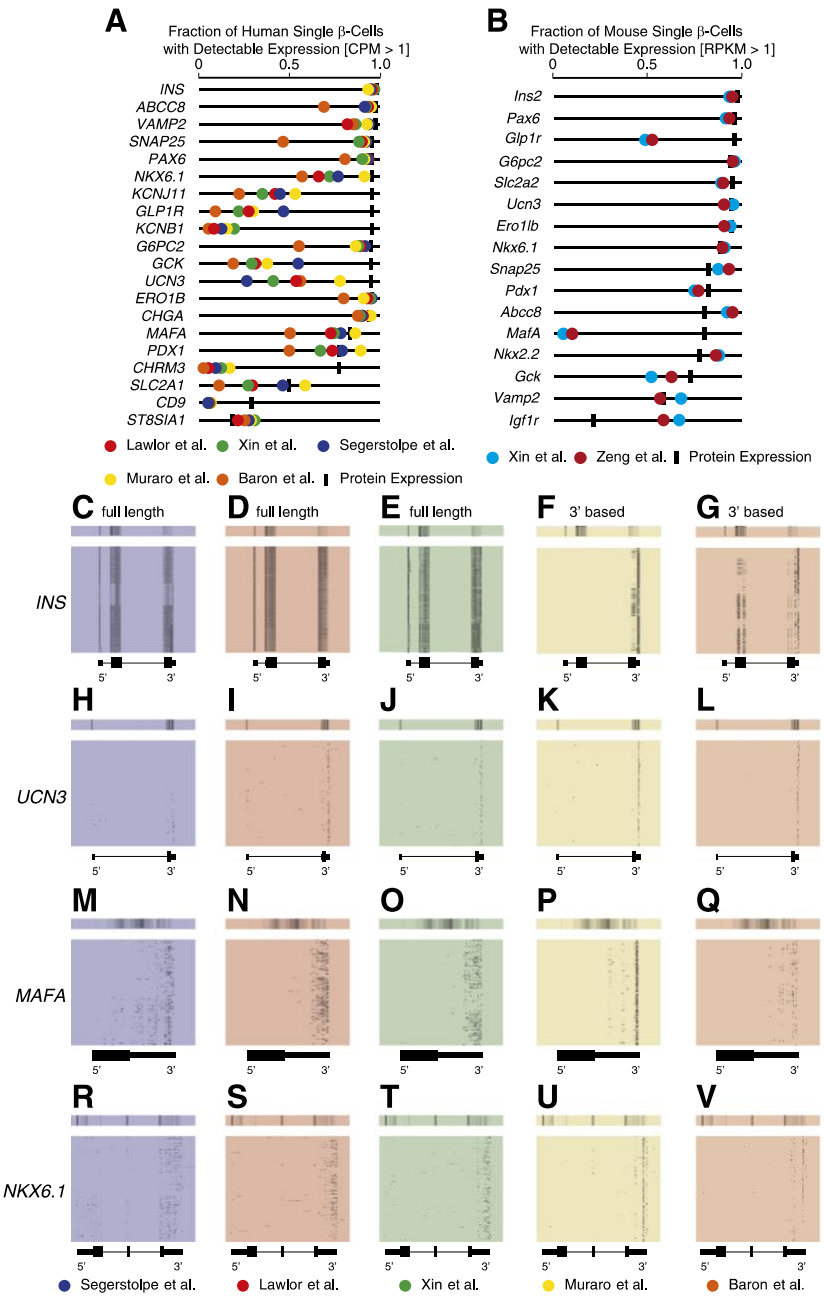


Figure 4—Comparison of detection of established β-cell markers in single β-cells comparing RNA-Seq and immunohistochemistry and coverage across the gene body of established β-cell markers detected by scRNA-Seq. *A*: Fraction of single human β-cells with detectable expression (CPM >1) across five human β-cell scRNA-Seq studies, compared with the fraction of human β-cells reported to express the corresponding protein (Supplementary Table 5*A*). The majority of genes are underdetected in human scRNA-Seq compared with the fraction of human β-cells that express the corresponding protein. *B*: The approach used in Fig. 4*A*, applied to established β-cell markers comparing two mouse scRNA-Seq studies and expression of the proteins encoded by these genes (Supplementary Table 5*B*). *C*–*V*: UCSC Genome Browser plots of read coverage across these genes, using single–β-cell libraries and companion bulk sequencing across all studies and platforms, highlighting differences between full-length and 3′-based capture (43,44,46–48).

(Supplementary Table 5). In addition to INS, examples include transcription factors such as PDX1 (57), NKX6.1 (59), PAX6 (60), and MAFA (61), as well as proteins required for normal stimulus-secretion coupling, insulin processing, and exocytosis such as SLC2A1 (62), GLP1R (63), ABCC8 (64), KCNJ11 (65), GCK (66), G6PC2 (67), KCNB1 (65), ERO1B (68), VAMP2 (69), SNAP25 (69), and UCN3 (70). With the exception of PDX1, MAFA, and SLC2A1, all of the proteins encoded by these genes are detected in more than an estimated 95% of human β-cells in healthy islets by immunohistochemical techniques. However, mRNA for all but the most abundantly expressed of these genes is consistently detected in a decidedly smaller fraction of β-cells than stain positive for the protein product they encode (Fig. 4*A*). *NKX6.1*, *UCN3*, *KCNJ11*, and *KCNB1* transcripts are detected in a particularly low fraction of β-cells. One possible explanation for this is intermittent transcription, where transcription occurs in discrete bursts that underlies stable protein expression (71). However, if this is the case, one would expect uniform coverage gene body capture for the subset of β-cells that would have been captured during the burst phase of expression for that gene. Instead, UCSC Genome Browser plots for these genes indicate widespread 3′ bias and underrepresentation of many known β-cell genes, even those that are expressed at medium to high transcript levels such as *UCN3*, *MAFA,* and *NKX6*-1 (Fig. 4*H–V*). This is a likely consequence of working at or below the level of detection of scRNA-Seq approaches. One uncommon example with read coverage across the full gene model was observed in a distinct subset of β-cells for *DLK1*, which reflects a pattern in line with burst transcription (Supplementary Fig. 6*A–E*). *ST8SIA1* and *CD9*, two genes that encode protein markers recently used to distinguish four distinct human β-cell types (21), are also consistently underdetected in single–human β-cell transcriptomes. A similar set of ubiquitous β-cell genes that are expressed at medium to high levels in mouse β-cell transcriptomes are detected in a higher fraction of β-cells, although large discrepancies remain for *Glp1r* and *Mafa* (Fig. 4*B*).

ASSESSING SINGLE-CELL SEQUENCING QUALITY

Until this point, we have largely used the fraction of β-cells with detectable expression (CPM >1) of a given gene as a metric of the fidelity of scRNA-Seq (Fig. 2, Supplementary Fig. 1*B*, and Supplementary Fig. 3*E*). This revealed that heterogeneous detection and significant 3′ bias is the norm for single–human β-cell transcriptomes, irrespective of investigator, approach, or platform (Supplementary Table 1). To better quantify the gap in quality of gene coverage between single-cell and bulk sequencing approaches, we adopted the transcript integrity number (TIN) score (72). This metric ranges between 0 and 100 and is calculated after library preparation and sequencing to reflect the quality and uniformity of read coverage across the gene model. A high TIN score for a gene reflects uniform read coverage across the gene model, while a low TIN score reflects uneven coverage across the gene model owing to 3′ bias, GC bias, or transcript degradation (Fig. 3 and Supplementary Fig. 7). TIN scores strongly correlate with the RNA integrity number, a measure of RNA quality used to assess input RNA quality before library preparation.

To visualize the relationship between gene expression and quality of its representation in single-cell versus bulk RNA-Seq approaches, we compared the correlation of TIN scores and gene expression among five human (43,44,46–48) and two mouse single-cell studies (42,45) with two mouse (18,19) and three human bulk islet RNA-Seq data sets (43,47,73). For bulk RNA-Seq approaches, there is essentially no drop-off in TIN score with lower gene expression (expressed as CPM) until CPM values are <5 (Fig. 5*A*). In other words, in bulk RNA-Seq approaches, the quality of the coverage of gene expression across the gene model from 5′ to 3′ is both high and independent of transcript abundance unless gene expression is quite low. In sharp contrast, in scRNA-Seq, there is a very clear effect of the abundance of gene expression on TIN score across the full range of transcript abundance values. Even at highly abundant transcripts with CPM values >100, TIN scores remain well below those of similarly abundant genes detected via bulk RNA-Seq. This reflects the drop-off in the quality of sequence coverage that is the consequence of working at or below the level of detection in scRNA-Seq approaches. Limiting analysis to only genes with a consistently high TIN score would yield more reliable and reproducible results but would also drastically undercut the number of genes that are included in the analysis, as over half of the genes detected in human β-cell scRNA-Seq have TIN scores <20. A comparison of TIN score cutoff versus CPM cutoff to the fraction of remaining genes suggests that TIN score cutoffs are a better metric than CPM cutoffs to separate high- and medium-quality read data (Fig. 5*B*).

Conversely, for bulk RNA-Seq samples, significant numbers of genes are excluded from the analysis only when the TIN quality threshold is raised over 50 (Fig. 5*C*).

CROSS-CONTAMINATION IN SINGLE–ISLET CELL TRANSCRIPTOMES

One question that continues to stir debate in the field is whether healthy β-cells transcribe GCG at low abundance and conversely if α-cells transcribe INS. Indeed, the α-cell cluster in our study clearly contains lower but detectable levels of *INS*, and β-cells had detectable levels of *GCG* (Fig. 1*B* and *C*). While cells that coexpress insulin and glucagon protein are regularly observed during embryonic development and in stem cell–derived β-cell–like cultures (32,74,75), they are exceedingly rare in healthy adult islets (76,77). However, this does not rule out translational inhibition of *GCG* in β-cells and *INS* in α-cells. Indeed, bulk RNA-Seq data of FACS-purified mouse α-cells detect *Ins2* expression at 80- to 170-fold lower than *Ins2* in β-cells from the same islets (Fig. 6*A*) (18). Similarly, *Gcg* is detected

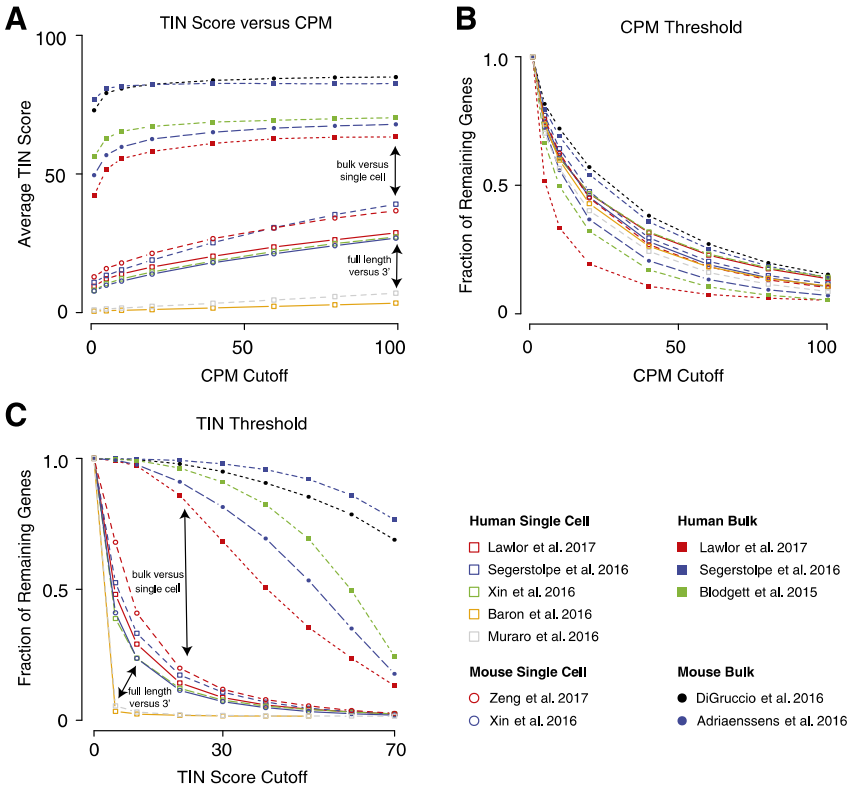


Figure 5—Visualization of the difference in sequence quality between single-cell and bulk RNA-Seq. *A*: Average TIN quality score plotted against transcript abundance in CPM across five human and two mouse single-cell and bulk RNA-Seq data sets. Note the large difference in quality between single-cell and bulk RNA-Seq data and that TIN scores in bulk are uniformly high irrespective of transcript abundance. *B*: Progressive application of a CPM threshold leads to the exclusion of similar relative numbers of genes between single-cell and bulk data sets (after correcting for the fact that single-cell libraries detect between 2,000 and 6,000 genes, while bulk libraries detect around 16,000 genes). *C*: Progressive application of a TIN threshold leads to the rapid exclusion of genes from single-cell data sets but only effects bulk RNA-Seq data sets at much more stringent TIN score cutoffs. Data from human (43,44,46–48) and mouse (42,45) scRNA-Seq and human (43,47,73) and mouse (18,19) bulk RNA-Seq.

in FACS-purified β-cells at 220-fold lower levels than its expression in α-cells (Fig. 6*B*) (18). This relatively low level of detectable reads could be caused by cross-contamination during FACS purification. While doublets, including those consisting of an α-cell and a β-cell, are normally gated out before collection, a well-calibrated FACS running at a conservative speed has an error rate less than 1%. In the context of FACS purification of dissociated islet suspensions, this means that fewer than 1% of the events that are sorted as β-cells are in fact a non–β-cell, possibly an α-cell. Since *GCG* accounts for up to 20% of all reads in the α-cell pool (17,18), a couple of contaminating α-cells could suffice to explain the detection of *Gcg* in transcriptomes of bulk FACS-purified β-cells.

Single-cell approaches ostensibly do not suffer from this confounder as they assess transcription in individual cells. Indeed, 0.2–1.5% of all reads in single human β-cells map to *GCG* and 0.001–1.109% of reads in single human α-cells map to *INS*. These observations at face value have been suggested as definitive proof that β-cells express *GCG* and α-cells express *INS*. However, Macosko et al. (78), in their original article describing the Drop-Seq approach, conducted a key control experiment that is often overlooked but is of direct relevance in this discussion. They approached the question of contamination at the single-cell level by mixing human HEK cells and mouse 3T3 cells prior to droplet formation and single-cell sequencing. They observed that an average of 0.26–2.44% of the reads in each and every single cell mapped uniquely to the genome of the other species (Fig. 6*C*). As they demonstrate, this can only be explained by the integration of free-floating or naked mRNA derived from cells that were disrupted by generating cell suspensions into libraries constructed from single cells that did not actually express the message (78). This problem is not unique to the Drop-Seq approach but will affect any procedure where tissues are dissociated into a single-cell suspension in preparation of single-cell sequencing or FACS sorting in bulk RNA-Seq approaches

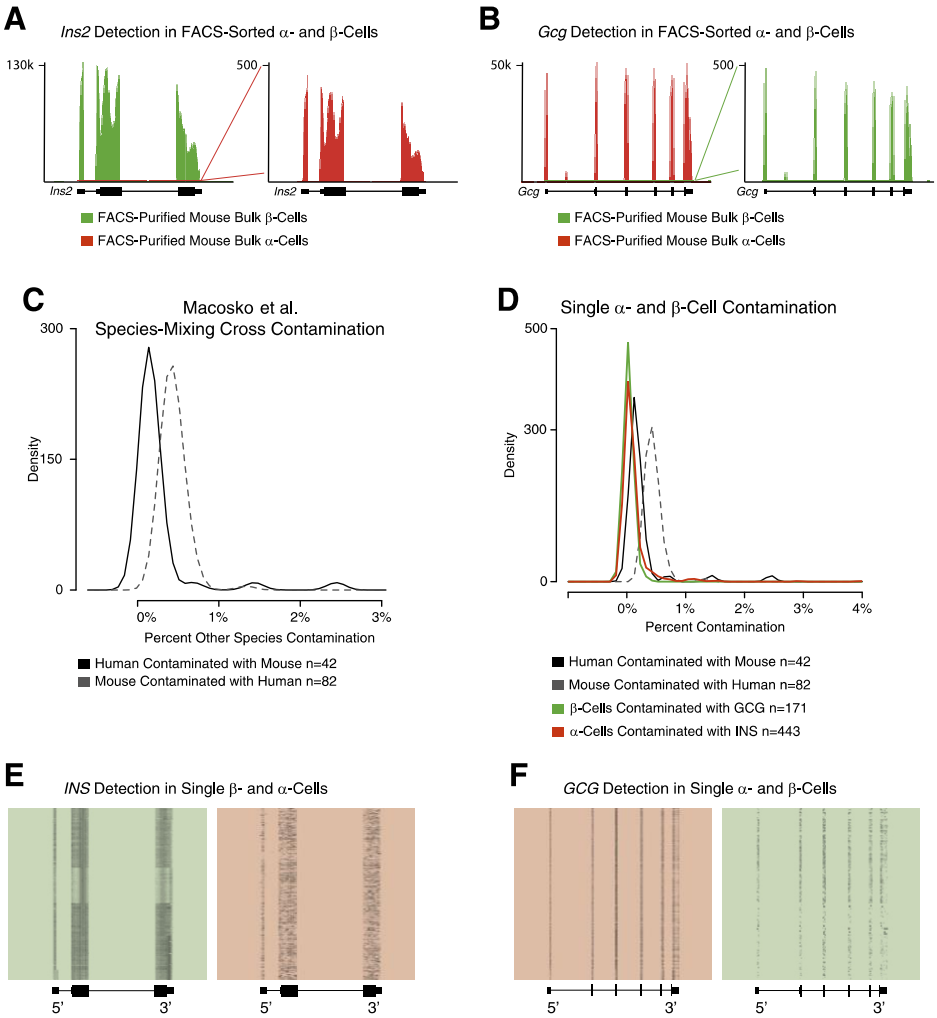


Figure 6—Detection of GCG in β-cells and INS in α-cells in bulk and scRNA-Seq. *A*: Detection of *Ins2* in FACS-purified mouse α-cells at 260-fold lower levels compared with β-cells from the same islets. Data from DiGruccio et al. (18). *B*: Detection of *Gcg* in FACS-purified mouse β-cells at 100-fold lower levels compared with α-cells from the same islets. Data from DiGruccio et al. (18). *C*: Species mixing experiments between cell lines of mouse and human origin have illustrated that 0.26–2.44% of all reads detected in single-cell libraries uniquely mapped to the other species. Data from Macosko et al. (78). *D*: Fraction of reads within 161 single–β-cell libraries that maps to GCG, and fraction of reads within 443 single α-cell libraries that maps to INS. Data from Segerstolpe et al. (47). The INS and GCG cross-detection is below the threshold of species contamination established by Macosko et al. (78). Given the extremely high abundance of INS and GCG, their detection in single α- and β-cells, respectively, may not be evidence of actual gene expression. *E*: Coverage of INS across all single β- and α-cell libraries. Data from Segerstolpe et al. (47). *F*: Coverage of GCG across all single α- and β-cell libraries. Data from Segerstolpe et al. (47).

(78). This relatively low level of cross-contamination will likely not meaningfully affect detection of the large majority of genes. However, *INS* and *GCG* are expressed so abundantly in β- and α-cells, respectively, that their cross-detection could be explained entirely by contamination of free-floating mRNA (Fig. 6*D–F*). These observations do not rule out true *GCG* expression by β-cells. However, the detection of *GCG* in single–β-cell transcriptomes at levels below those estimated through the species cross-contamination paradigm established by Macosko et al. (78) cannot be taken as proof that β-cells actually express *GCG* mRNA.

FUTURE OUTLOOK

The fact that we can now detect and attempt to quantify gene expression in single cells is in itself a remarkable achievement. A survey of β-cell gene expression at single-cell resolution across hundreds or even thousands of

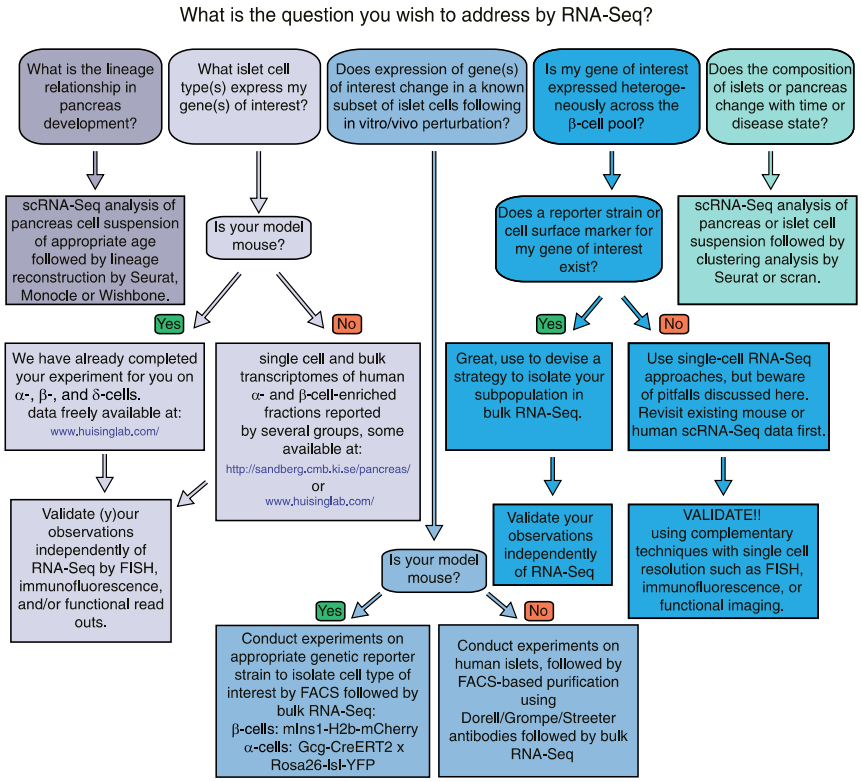


Figure 7—Workflow to align the appropriate sequencing approaches with the stated experimental goal**.** Starting from five common experimental scenarios, this flowchart offers guidance to the reader to facilitate the choice for the different experimental strategies available, considering their benefits and drawbacks**.** FISH, fluorescence in situ hybridization**.**

individual cells is a very enticing prospect that would resolve some of the long-known heterogeneity among β-cells with regard to their functional state or proliferative status. However, in attempting to detect gene expression in single β-cells, it has become obvious that we are operating at or below the limit of reliable detection for a large majority of genes. This comes at a steep price with regard to the quality of the single-cell sequence data that is obtained, irrespective of the investigating laboratory or the chosen single-cell approach.

In this Perspective, we have illustrated these inherent limitations of scRNA-Seq applied to adult human islet cells by pointing out the underestimation of the number of detected genes per single cell and by applying TIN scores as a quantitative measure of the incomplete coverage and 3′ bias that affects all genes, from rare to highly abundant. By comparison, the quality of the gene coverage in bulk RNA-Seq samples is so much better that it is quite possible that the coverage and data quality of scRNA-Seq may not approach that of bulk RNA-Seq for some time. Therefore, for each experiment investigators need to determine if transcript detection at single-cell resolution is worth these inevitable drawbacks (Fig. 7). Given the large quality gap between single-cell versus bulk transcriptome data, we would advocate for a bulk transcriptome approach, if compatible with your experimental question, in spite of the perceived novelty of single-cell sequencing. Evidently, if transcriptional heterogeneity among β- or α-cells is the central focus of a study, scRNA-Seq experiments may be the only choice, unless a known marker for these subpopulations can be leveraged to isolate these cells by FACS for bulk sequencing. Nevertheless, our illustration that—with the exception of a handful of the most highly abundant transcripts—every single gene is detected in only a fraction of β-cells questions the ability of scRNA-Seq to discern true heterogeneous expression amid widespread heterogeneous detection. Case in point is the fact that none of the many markers of known heterogeneity were independently identified by any of the "unbiased" scRNA-Seq approaches, with some acknowledging their inability to do so (43). Therefore, any observation derived from single-cell or bulk RNA-Seq experiments should—wherever possible—be subject to rigorous validation using independent approaches that can achieve single-cell resolution, such

as RNA fluorescence in situ hybridization to detect the message, immunofluorescence to detect the protein encoded by that mRNA, and/or live cell functional imaging to correlate gene expression with functional readout that indicates the presence of the corresponding protein.

Our intent in drawing attention to the limitations of scRNA-Seq approaches applied to islet cells is certainly not to dissuade our colleagues from relying on observations obtained by scRNA-Seq approaches in their studies of islet function. It should not be a surprise that working at the extreme limits of our technical capabilities comes at a price. Ongoing improvements in library preparation, including the generation of protocols that no longer rely on multiple rounds of PCR amplification, should constitute a significant improvement (79,80). As tissue collection and processing time will influence gene expression and mRNA stability, standardization of the collection of human islets to the extent possible will increase the conformity of gene representation across both bulk and scRNA-Seq studies. New methods, such as split-pool ligation-based transcriptome sequencing (SPLiT-Seq) (81), may be able to overcome some of the limitations of current scRNA-Seq protocols. While SPLiT-Seq still requires tissue dissociation, it instead compartmentalizes the RNA into single-cell libraries within the native cell rather than relying on droplets or wells. This may help mitigate issues such as poor library complexity and 3′ bias and may reduce contamination with naked mRNA (78) (Fig. 6). Spatial transcriptomics may also provide more reliable avenues for scRNA-Seq, as they avoid confounders associated with islet dissociation and would allow the field an unbiased perspective to determine whether heterogeneity of gene expression is spatially driven (82), as was recently suggested (35,83). Newer 3′-based methods that allow for higher throughput, and greater sample sizes at reasonable cost, have allowed the identification of rarer populations, such as ε-cells (84). Targeted sequencing approaches, such as droplet-assisted RNA targeting by single-cell sequencing (DART-Seq), significantly improve coverage by targeting the limited depth of scRNA-Seq to a subset of preselected transcripts of interest (85). Computational methods are being developed to take into account and correct for confounding factors, such as donor genetic variation, dropout, and technical noise, although avoiding confounders will always be preferable to correcting for them through bioinformatic means (86,87).

Despite the current limitation of the approach, scRNA-Seq experiments have successfully resolved gene expression in human δ-cells (47,88), for which purification methods to obtain bulk samples do not exist. Moreover, scRNA-Seq has recapitulated known differentiation trajectories in the development of many organs and tissues (89–93). This includes the pancreas, where scRNA-Seq has been able to trace Ngn3+ progenitor populations at different embryonic ages to preferentially differentiate into α- or β-cells (94), thus recapitulating and validating a phenomenon that had previously been independently described by careful developmental biology experiments (95,96). *Gjd2*, *Scg2*, *Ociad2*, and *Fev*, novel genes whose contribution to embryonic pancreas development had not been known, have also emerged from scRNA-Seq efforts (80,94). Moreover, pseudo-time strategies, where single cells are placed on a lineage based on their transcriptional stage instead of their chronological age, have successfully resolved aspects of postnatal β-cell maturation (42).

In summary, our goal with this Perspective has been to raise awareness among a general audience of diabetes researchers of some of the limitations of scRNA-Seq and discuss potential solutions to overcome the current limitations. It is amazing that we are now capable of detecting islet cell gene expression at single-cell resolution. It therefore should not come as a surprise that there is inevitably a price to pay for the benefit of single-cell resolution. The limitations we discussed should be well known to investigators who have been at the forefront of single-cell sequencing. However, they are likely less appreciated by a general audience of diabetes researchers not as well versed in bioinformatics, who nevertheless use scRNA-Seq data generated by others or are adopting scRNA-Seq for their own future experiments. Next-generation sequencing at single-cell resolution has the potential to reveal unprecedented insight into biological processes that until recently had remained out of reach. We hope that the considerations discussed in this Perspective will help our colleagues align their sequencing approaches with realistic experimental goals.

Acknowledgments. The authors thank Dr. Talitha van der Meulen, University of California, Davis, for constructive comments on the manuscript.

Funding. This work was supported by grants from the National Institutes of Health National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK110276) and JDRF (CDA-2-2013-54) to M.O.H. A.M.M. was supported by the Stephen F. and Bettina A. Sims Immunology Fellowship.

Duality of Interest. No potential conflicts of interest relevant to this article were reported.

References

1. Menke A, Casagrande S, Geiss L, Cowie CC. Prevalence of and trends in diabetes among adults in the United States, 1988-2012. JAMA 2015;314:1021–1029

2. Halban PA, Polonsky KS, Bowden DW, et al. β-Cell failure in type 2 diabetes: postulated mechanisms and prospects for prevention and treatment. Diabetes Care 2014;37:1751–1758

3. Schuit FC, Huypens P, Heimberg H, Pipeleers DG. Glucose sensing in pancreatic β-cells: a model for the study of other glucose-regulated cells in gut, pancreas, and hypothalamus. Diabetes 2001;50:1–11

4. Röder PV, Wu B, Liu Y, Han W. Pancreatic regulation of glucose homeostasis. Exp Mol Med 2016;48:e219

5. Johnston NR, Mitchell RK, Haythorne E, et al. Beta cell hubs dictate pancreatic islet responses to glucose. Cell Metab 2016;24:389–401

6. Brissova M, Fowler MJ, Nicholson WE, et al. Assessment of human pancreatic islet architecture and composition by laser scanning confocal microscopy. J Histochem Cytochem 2005;53:1087–1097

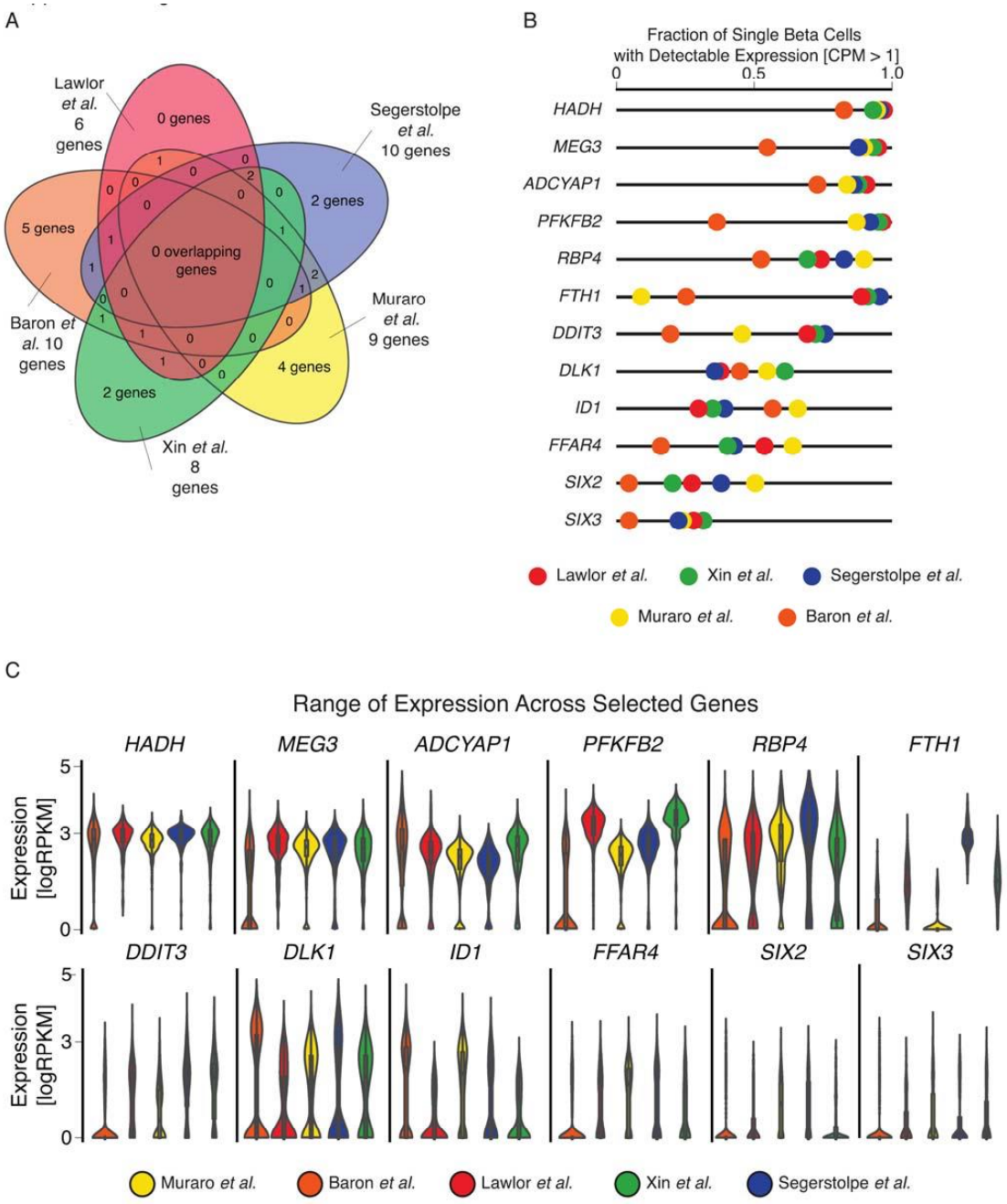
7. Brereton MF, Vergari E, Zhang Q, Clark A. Alpha-, delta- and PP-cells: are they the architectural cornerstones of islet structure and co-ordination? J Histochem Cytochem 2015;63:575–591

8. Cabrera O, Berman DM, Kenyon NS, Ricordi C, Berggren PO, Caicedo A. The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. Proc Natl Acad Sci U S A 2006;103:2334–2339
9. Arrojo e Drigo R, Ali Y, Diez J, Srinivasan DK, Berggren PO, Boehm BO. New insights into the architecture of the islet of Langerhans: a focused cross-species assessment. Diabetologia 2015;58:2218–2228
10. Ranjan AK, Joglekar MV, Hardikar AA. Endothelial cells in pancreatic islet development and function. Islets 2009;1:2–9
11. Unanue ER. Macrophages in endocrine glands, with emphasis on pancreatic islets. Microbiol Spectr 2016;4:1–11
12. Brissova M, Haliyur R, Saunders D, et al. α Cell function and gene expression are compromised in type 1 diabetes. Cell Reports 2018;22:2667–2676
13. Lam CJ, Cox AR, Jacobson DR, Rankin MM, Kushner JA. Highly proliferative α-cell–related islet endocrine cells in human pancreata. Diabetes 2018;67:674–686
14. Dai C, Brissova M, Reinert RB, et al. Pancreatic islet vasculature adapts to insulin resistance through dilation and not angiogenesis. Diabetes 2013;62:4144–4153
15. Pipeleers D. The biosociology of pancreatic B cells. Diabetologia 1987;30:277–291
16. Hara M, Wang X, Kawamura T, et al. Transgenic mice with green fluorescent protein-labeled pancreatic β-cells. Am J Physiol Endocrinol Metab 2003;284:E177–E183
17. Benner C, van der Meulen T, Cacéres E, Tigyi K, Donaldson CJ, Huising MO. The transcriptional landscape of mouse beta cells compared to human beta cells reveals notable species differences in long non-coding RNA and protein-coding gene expression. BMC Genomics 2014;15:620
18. DiGruccio MR, Mawla AM, Donaldson CJ, et al. Comprehensive alpha, beta and delta cell transcriptomes reveal that ghrelin selectively activates delta cells and promotes somatostatin release from pancreatic islets. Mol Metab 2016;5:449–458
19. Adriaenssens AE, Svendsen B, Lam BY, et al. Transcriptomic profiling of pancreatic alpha, beta and delta cell populations identifies delta cells as a principal target for ghrelin in mouse islets. Diabetologia 2016;59:2156–2165
20. Dorrell C, Abraham SL, Lanxon-Cookson KM, Canaday PS, Streeter PR, Grompe M. Isolation of major pancreatic cell types and long-term culture-initiating cells using novel human surface markers. Stem Cell Res (Amst) 2008;1:183–194
21. Dorrell C, Schug J, Canaday PS, et al. Human islets contain four distinct subtypes of β cells. Nat Commun 2016;7:11756
22. Banerjee M, Otonkoski T. A simple two-step protocol for the purification of human pancreatic beta cells. Diabetologia 2009;52:621–625
23. Wang YJ, Golson ML, Schug J, et al. Single-cell mass cytometry analysis of the human endocrine pancreas. Cell Metab 2016;24:616–626
24. Clardy SM, Mohan JF, Vinegoni C, et al. Rapid, high efficiency isolation of pancreatic β-cells. Sci Rep 2015;5:13681
25. Pipeleers D, Kiekens R, Ling Z, Wilikens A, Schuit F. Physiologic relevance of heterogeneity in the pancreatic beta-cell population. Diabetologia 1994;37(Suppl. 2):S57–S64
26. Van Schravendijk CF, Kiekens R, Pipeleers DG. Pancreatic beta cell heterogeneity in glucose-induced insulin secretion. J Biol Chem 1992;267:21344–21348
27. Bader E, Migliorini A, Gegg M, et al. Identification of proliferative and mature β-cells in the islets of Langerhans. Nature 2016;535:430–434
28. Benninger RK, Piston DW. Cellular communication and heterogeneity in pancreatic islet insulin secretion dynamics. Trends Endocrinol Metab 2014;25:399–406
29. Olsson R, Carlsson PO. A low-oxygenated subpopulation of pancreatic islets constitutes a functional reserve of endocrine cells. Diabetes 2011;60:2068–2075
30. Kiekens R, In 't Veld P, Mahler T, Schuit F, Van De Winkel M, Pipeleers D. Differences in glucose recognition by individual rat pancreatic B cells are associated with intercellular differences in glucose-induced biosynthetic activity. J Clin Invest 1992;89:117–125
31. Bouwens L, Klöppel G. Islet cell neogenesis in the pancreas. Virchows Arch 1996;427:553–560
32. Teitelman G, Alpert S, Polak JM, Martinez A, Hanahan D. Precursor cells of mouse endocrine pancreas coexpress insulin, glucagon and the neuronal proteins tyrosine hydroxylase and neuropeptide Y, but not pancreatic polypeptide. Development 1993;118:1031–1039
33. Hermann M, Pirkebner D, Draxl A, et al. Dickkopf-3 is expressed in a subset of adult human pancreatic beta cells. Histochem Cell Biol 2007;127:513–521
34. Rodnoi P, Rajkumar M, Moin ASM, Georgia SK, Butler AE, Dhawan S. Neuropeptide Y expression marks partially differentiated β cells in mice and humans. JCI Insight 2017;2:1–13
35. van der Meulen T, Mawla AM, DiGruccio MR, et al. Virgin beta cells persist throughout life at a neogenic niche within pancreatic islets. Cell Metab 2017;25:911–926.e6
36. Beamish CA, Strutt BJ, Arany EJ, Hill DJ. Insulin-positive, Glut2-low cells present within mouse pancreas exhibit lineage plasticity and are enriched within extra-islet endocrine cell clusters. Islets 2016;8:65–82
37. Rui J, Deng S, Arazi A, Perdigoto AL, Liu Z, Herold KC. β cells that resist immunological attack develop during progression of autoimmune diabetes in NOD mice. Cell Metab 2017;25:727–738
38. Thompson PJ, Shah A, Ntranos V, Van Gool F, Atkinson M, Bhushan A. Targeted elimination of senescent beta cells prevents type 1 diabetes. Cell Metab 2019;29:1045–1060.e10
39. Enge M, Arda HE, Mignardi M, et al. Single-cell analysis of human pancreas reveals transcriptional signatures of aging and somatic mutation patterns. Cell 2017;171:321–330.e14
40. Qiu W-L, Zhang Y-W, Feng Y, Li L-C, Yang L, Xu C-R, et al. Deciphering pancreatic islet beta cell and alpha cell maturation pathways and characteristic features at the single-cell level. Cell Metab 2017;25:1194–1205.e4
41. Ackermann AM, Wang Z, Schug J, Naji A, Kaestner KH. Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. Mol Metab 2016;5:233–244
42. Zeng C, Mulas F, Sui Y, et al. Pseudotemporal ordering of single cells reveals metabolic control of postnatal β cell proliferation. Cell Metab 2017;25:1160–1175.e11
43. Lawlor N, George J, Bolisetty M, et al. Single-cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. Genome Res 2017;27:208–222
44. Baron M, Veres A, Wolock SL, et al. A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. Cell Syst 2016;3:346–360.e4
45. Xin Y, Kim J, Ni M, et al. Use of the Fluidigm C1 platform for RNA sequencing of single mouse pancreatic islet cells. Proc Natl Acad Sci U S A 2016;113:3293–3298
46. Xin Y, Kim J, Okamoto H, et al. RNA sequencing of single human islet cells reveals type 2 diabetes genes. Cell Metab 2016;24:608–615
47. Segerstolpe Å, Palasantza A, Eliasson P, et al. Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. Cell Metab 2016;24:593–607
48. Muraro MJ, Dharmadhikari G, Grün D, et al. A single-cell transcriptome atlas of the human pancreas. Cell Syst 2016;3:385–394.e3
49. Wang YJ, Schug J, Won KJ, et al. Single-cell transcriptomics of the human endocrine pancreas. Diabetes 2016;65:3028–3038
50. Daley T, Smith AD. Predicting the molecular complexity of sequencing libraries. Nat Methods 2013;10:325–327
51. Daley T, Smith AD. Modeling genome coverage in single-cell sequencing. Bioinformatics 2014;30:3159–3165
52. Stegle O, Teichmann SA, Marioni JC. Computational and analytical challenges in single-cell transcriptomics. Nat Rev Genet 2015;16:133–145
53. Eisenberg E, Levanon EY. Human housekeeping genes, revisited. Trends Genet 2013;29:569–574

54. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. Genome Res 2002;12:996–1006
55. Guo S, Dai C, Guo M, et al. Inactivation of specific β cell transcription factors in type 2 diabetes. J Clin Invest 2013;123:3305–3316
56. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 2018;36:411–420
57. Dai C, Brissova M, Hang Y, et al. Islet-enriched gene expression and glucose-induced insulin secretion in human and mouse islets. Diabetologia 2012;55:707–718
58. Collombat P, Mansouri A, Hecksher-Sorensen J, et al. Opposing actions of Arx and Pax4 in endocrine pancreas development. Genes Dev 2003;17:2591–2603
59. Taylor BL, Liu FF, Sander M. Nkx6.1 is essential for maintaining the functional state of pancreatic beta cells. Cell Reports 2013;4:1262–1275
60. Ahmad Z, Rafeeq M, Collombat P, Mansouri A. Pax6 inactivation in the adult pancreas reveals ghrelin as endocrine cell maturation marker. PLoS One 2015;10: e0144597
61. Bonnavion R, Jaafar R, Kerr-Conte J, et al. Both PAX4 and MAFA are expressed in a substantial proportion of normal human pancreatic alpha cells and deregulated in patients with type 2 diabetes. PLoS One 2013;8:e72194
62. McCulloch LJ, van de Bunt M, Braun M, Frayn KN, Clark A, Gloyn AL. GLUT2 (SLC2A2) is not the principal glucose transporter in human pancreatic beta cells: implications for understanding genetic association signals at this locus. Mol Genet Metab 2011;104:648–653
63. Tornehave D, Kristensen P, Rømer J, Knudsen LB, Heller RS. Expression of the GLP-1 receptor in mouse, rat, and human pancreas. J Histochem Cytochem 2008;56:841–851
64. Guiot Y, Stevens M, Marhfour I, et al. Morphological localisation of sulfonylurea receptor 1 in endocrine cells of human, mouse and rat pancreas. Diabetologia 2007;50:1889–1899
65. Yan L, Figueroa DJ, Austin CP, et al. Expression of voltage-gated potassium channels in human and rhesus pancreatic islets. Diabetes 2004;53:597–607
66. Arden C, Harbottle A, Baltrusch S, Tiedge M, Agius L. Glucokinase is an integral component of the insulin granules in glucose-responsive insulin secretory cells and does not translocate during glucose stimulation. Diabetes 2004;53: 2346–2352
67. Hutton JC, Eisenbarth GS. A pancreatic β-cell-specific homolog of glucose-6-phosphatase emerges as a major target of cell-mediated autoimmunity in diabetes. Proc Natl Acad Sci U S A 2003;100:8626–8628
68. Zito E, Chin KT, Blais J, Harding HP, Ron D. ERO1-β, a pancreas-specific disulfide oxidase, promotes insulin biogenesis and glucose homeostasis. J Cell Biol 2010;188:821–832
69. Lena G, Reale P, Maggiora Vergano T, Anastasio M. [Fetal distress and spastic paralysis]. Quad Clin Ostet Ginecol 1967;22:755–766 [in Italian]
70. Blum B, Hrvatin S, Schuetz C, Bonal C, Rezania A, Melton DA. Functional beta-cell maturation is marked by an increased glucose threshold and by expression of urocortin 3. Nat Biotechnol 2012;30:261–264
71. Depken M, Parrondo JM, Grill SW. Intermittent transcription dynamics for the rapid production of long transcripts of high fidelity. Cell Reports 2013;5:521–530
72. Wang L, Nie J, Sicotte H, et al. Measure transcript integrity using RNA-seq data. BMC Bioinformatics 2016;17:58
73. Blodgett DM, Nowosielska A, Afik S, et al. Novel observations from next-generation RNA sequencing of highly purified human adult and fetal islet cell subsets. Diabetes 2015;64:3172–3181
74. Chiang MK, Melton DA. Single-cell transcript analysis of pancreas development. Dev Cell 2003;4:383–393
75. De Krijger RR, Aanstoot HJ, Kranenburg G, Reinhard M, Visser WJ, Bruining GJ. The midgestational human fetal pancreas contains cells coexpressing islet hormones. Dev Biol 1992;153:368–375
76. Katsuta H, Akashi T, Katsuta R, et al. Single pancreatic beta cells co-express multiple islet hormone genes in mice. Diabetologia 2010;53:128–138
77. Herrera PL. Adult insulin- and glucagon-producing cells differentiate from two independent cell lineages. Development 2000;127:2317–2322
78. Macosko EZ, Basu A, Satija R, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 2015;161:1202–1214
79. Tabula Muris Consortium; Overall coordination; Logistical coordination; Organ collection and processing; Library preparation and sequencing; Computational data analysis; Cell type annotation; Writing group; Supplemental text writing group; Principal investigators. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature 2018;562:367–372
80. Byrnes LE, Wong DM, Subramaniam M, et al. Lineage dynamics of murine pancreatic development at single-cell resolution. Nat Commun 2018;9:3922
81. Rosenberg AB, Roco CM, Muscat RA, et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science 2018; 360:176–182
82. Rodriques SG, Stickels RR, Goeva A, et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. Science 2019; 363:1463–1467
83. Cui Y, Hu D, Markillie LM, et al. Fluctuation localization imaging-based fluorescence in situ hybridization (fliFISH) for accurate detection and counting of RNA copies in single cells. Nucleic Acids Res 2018;46:e7
84. Dominguez Gutierrez G, Kim J, Lee AH, et al. Gene signature of the human pancreatic ε cell. Endocrinology 2018;159:4023–4032
85. Saikia M, Burnham P, Keshavjee SH, et al. Simultaneous multiplexed amplicon sequencing and transcriptome profiling in single cells. Nat Methods 2019;16:59–62
86. Liu P, Song R, Elison GL, Peng W, Acar M. Noise reduction as an emergent property of single-cell aging. Nat Commun 2017;8:680
87. Büttner M, Miao Z, Wolf FA, Teichmann SA, Theis FJ. A test metric for assessing single-cell RNA-seq batch correction. Nat Methods 2019;16:43–49
88. Lawlor N, Khetan S, Ucar D, Stitzel ML. Genomics of islet (dys)function and type 2 diabetes. Trends Genet 2017;33:244–255
89. Setty M, Tadmor MD, Reich-Zeliger S, et al. Wishbone identifies bifurcating developmental trajectories from single-cell data. Nat Biotechnol 2016;34:637–645
90. Schwalie PC, Ordóñez-Morán P, Huelsken J, Deplancke B. Cross-tissue identification of somatic stem and progenitor cells using a single-cell RNA-sequencing derived gene signature. Stem Cells 2017;35:2390–2402
91. Farrell JA, Wang Y, Riesenfeld SJ, et al. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. Science 2018;360: eaar3131
92. Semrau S, Goldmann JE, Soumillon M, Mikkelsen TS, Jaenisch R, van Oudenaarden A. Dynamics of lineage commitment revealed by single-cell transcriptomics of differentiating embryonic stem cells. Nat Commun 2017;8:1096
93. Baron CS, Kester L, Klaus A, et al. Single-cell transcriptomics reveal the dynamic of haematopoietic stem cell production in the aorta. Nat Commun 2018;9: 2517
94. Scavuzzo MA, Hill MC, Chmielowiec J, et al. Endocrine lineage biases arise in temporally distinct endocrine progenitors during pancreatic morphogenesis. Nat Commun 2018;9:3356
95. Gasa R, Mrejen C, Leachman N, et al. Proendocrine genes coordinate the pancreatic islet differentiation program in vitro. Proc Natl Acad Sci U S A 2004;101: 13245–13250
96. Sommer L, Ma Q, Anderson DJ. Neurogenins, a novel family of atonal-related bHLH transcription factors, are putative mammalian neuronal determination genes that reveal progenitor cell heterogeneity in the developing CNS and PNS. Mol Cell Neurosci 1996;8:221–241

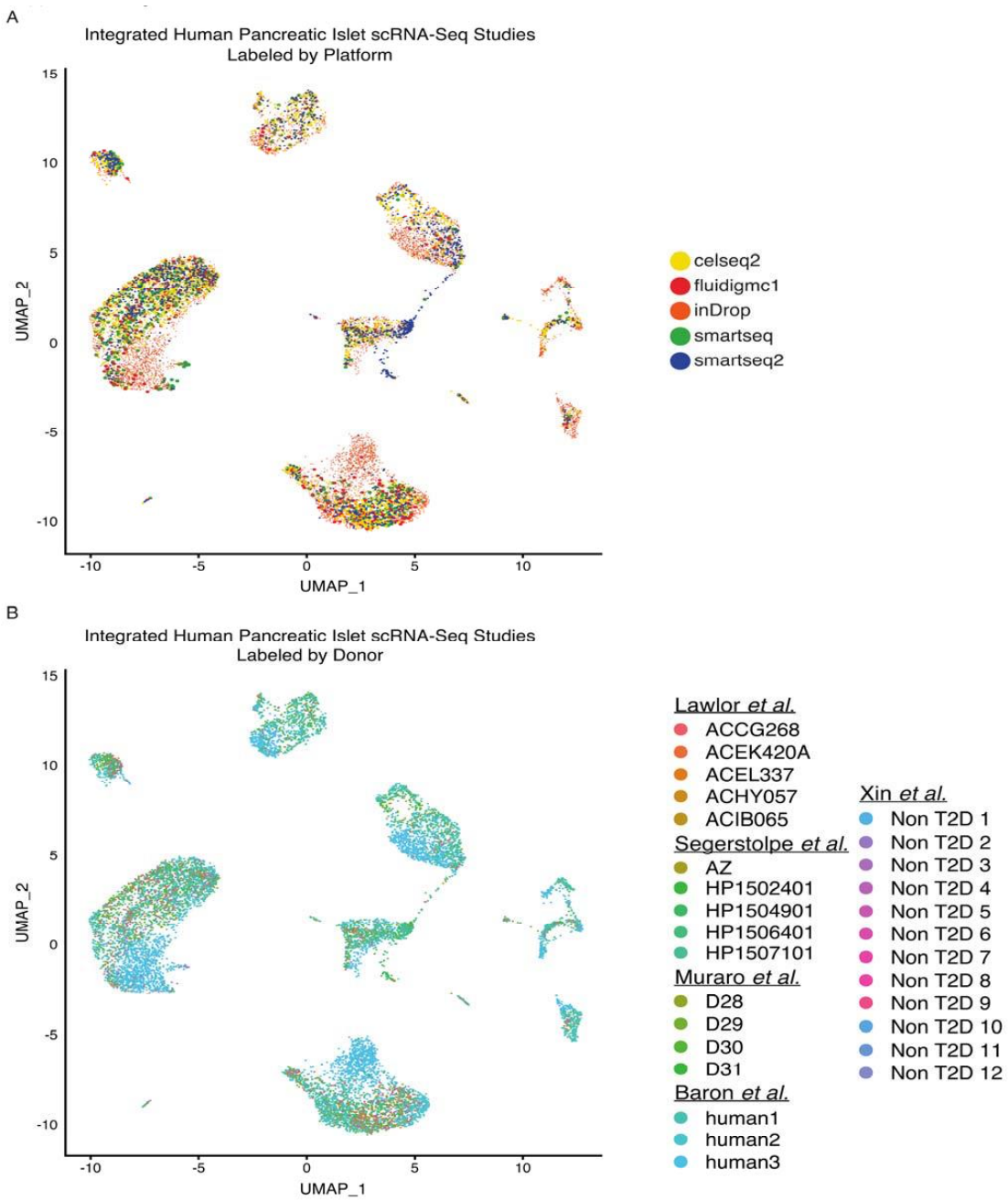
SUPPLEMENTARY DATA

Supplementary Figure S1. Human beta cell heterogeneity as self-identified across five human scRNA-Seq studies (Table S2). A: Venn diagram illustrating the overlap of genes highlighted for novel heterogeneity across studies. B: Fraction of single beta cells with detectable (CPM > 1) expression across twelve genes from Figure S1A. C: Violin plots of log transformed CPM abundance across these same twelve genes across all five studies.



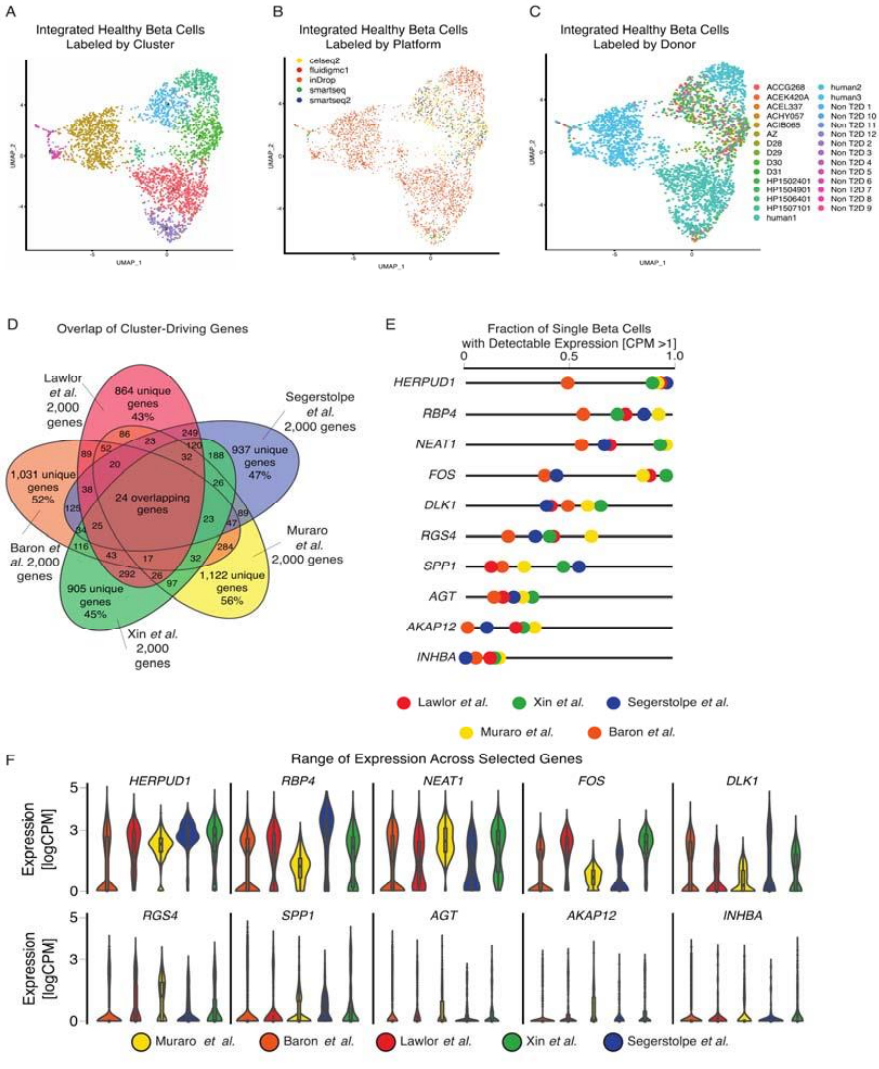
SUPPLEMENTARY DATA

Supplementary Figure S2. Dimensional reduction plots from our integrative analysis to confirm the absence of confounding factors such as donor variation or method of sequencing. A: Dimensional reduction highlighting the clusters derived based on platform of origin. Each cluster has an intermixed distribution of cells from all five studies. B: Dimensional reduction highlighting the clusters based on donor of origin. Again, all clusters contained cells across donors, confirming that clustering was not driven by these confounding factors.



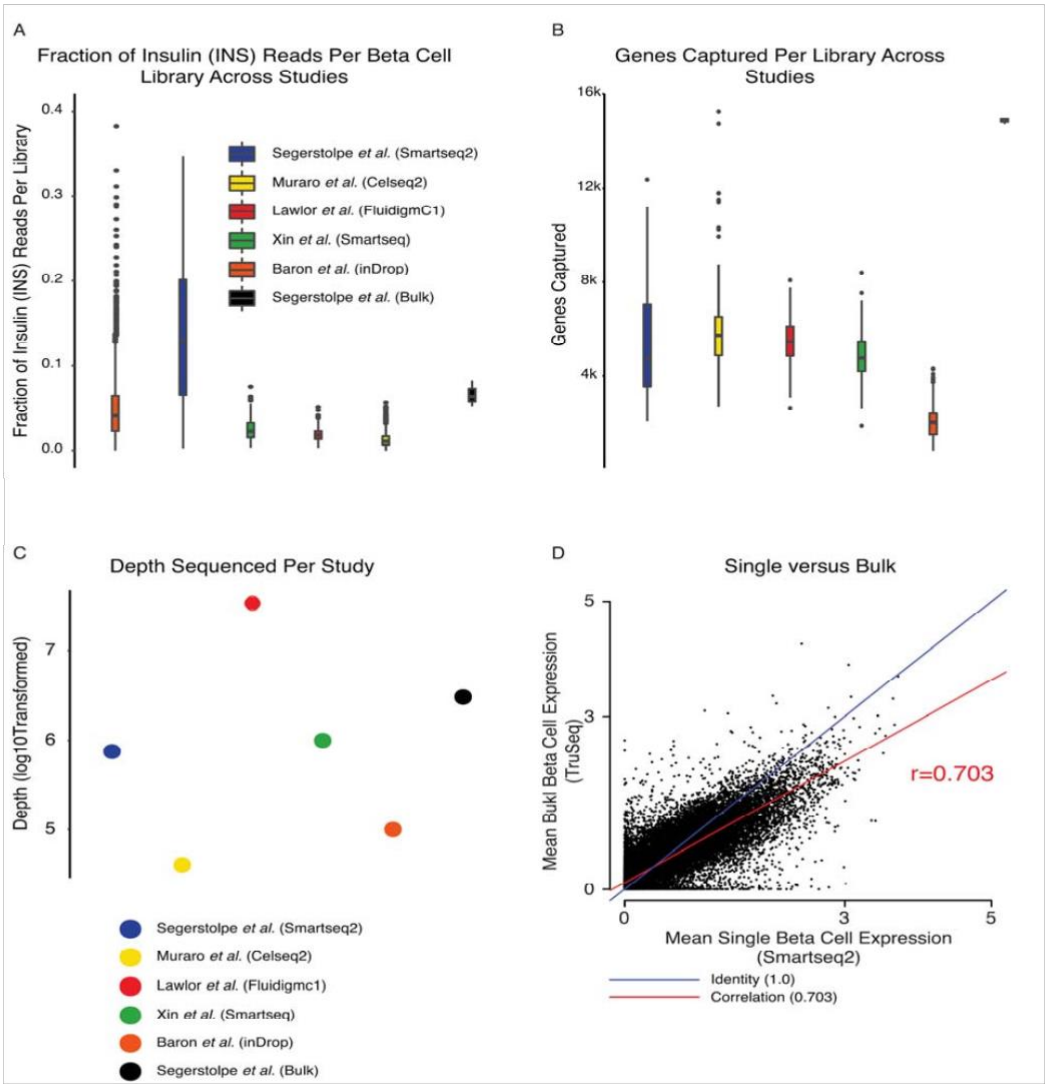
SUPPLEMENTARY DATA

Supplementary Figure S3. Integrative analysis as performed earlier (Figure 1A), but now on the subsetted beta cells from healthy donors. A: Dimensional reduction plots evaluating subpopulations within the healthy beta cells from all studies. B: Dimensional reduction exploring whether or not platform used drove subpopulations. C: Dimensional reduction exploring whether or not donor variation was influencing subpopulation formation. D: Unbiased clustering reveals surprisingly little overlap in the beta cell heterogeneity detected across five human single beta cell RNA-Seq studies. Venn diagram of the overlap in the genes that drove unbiased clustering between beta cells in five independent scRNA-Seq studies on healthy human islet donors (See Supplemental Approach & Tools for details). Between 43% and 56% of genes in each study were not detected as driving beta cell cluster in any of the four other studies; only 24 genes overlapped across all five studies (Table S4A). E: Fraction of single beta cells with detectable (CPM > 1) expression across ten genes selected from the list of 24 from (A). F: Violin plots of log transformed CPM abundance across these same ten genes across all five studies.



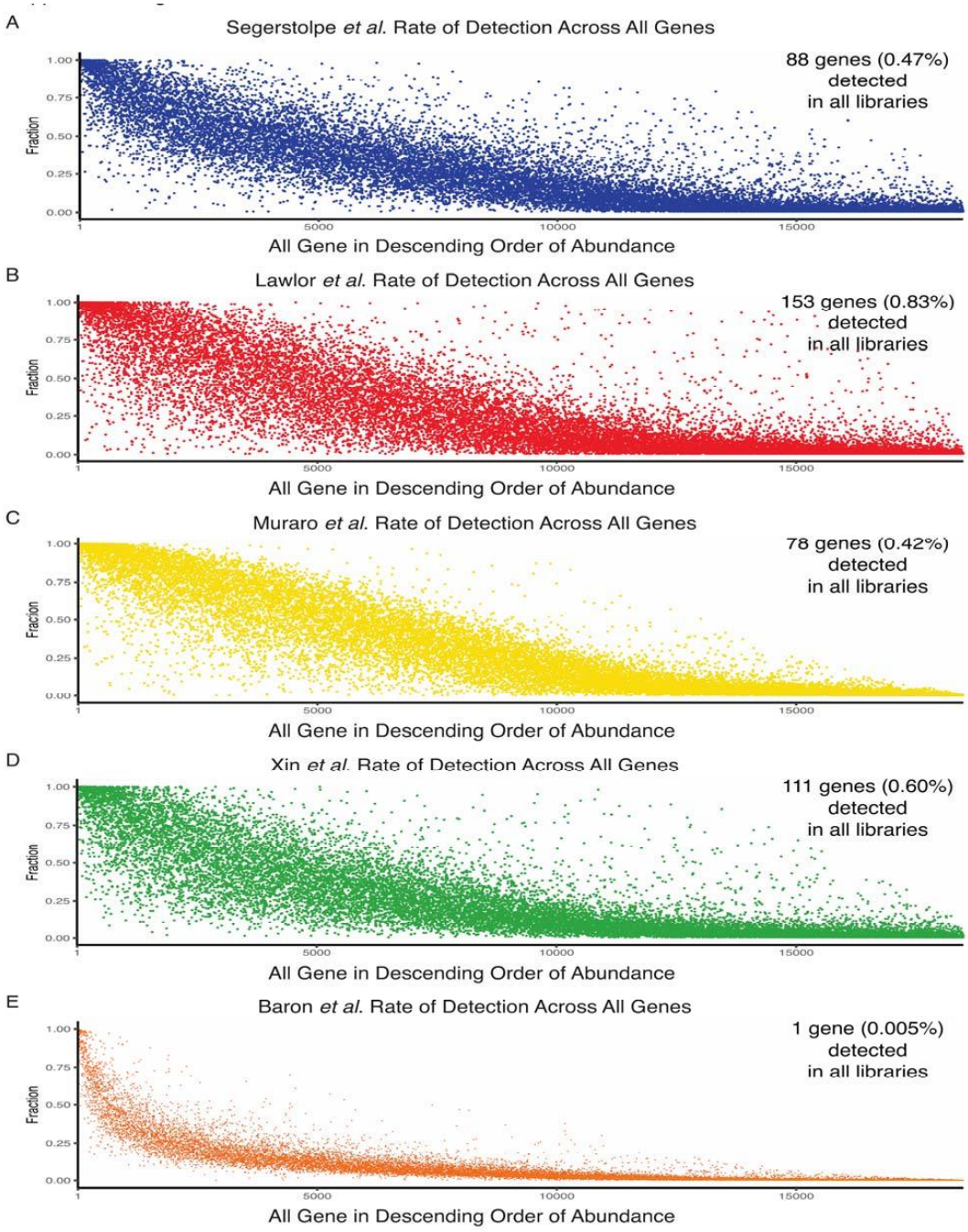
SUPPLEMENTARY DATA

Supplementary Figure S4. Range of genes detected in single cell libraries across five human single cell studies and one human bulk data set correlated with depth of sequencing and representation of INS. A: Fraction of insulin (INS) reads per beta cell library across five human single cell studies and one human bulk data set. Evaluation of the fraction of reads within a single or bulk beta cell libraries mapping to INS reveals that up to half of all reads in some scRNA-Seq libraries map to INS. No trend is apparent between platform used (Table S1) and INS over-representation. B: Evaluation of number of genes within a single or bulk beta cell libraries captured (CPM >=1).C: Depth of sequencing used across studies (Table S1), transformed by log10. No trend is apparent between platform used (Table S1) and number of genes captured. D: Correlation of gene expression between bulk (TruSeq) and pooled (SmartSeq2) single cell libraries from the same sample. Gene expression is represented as log transformed RPKM, data from (1). Blue line indicates identity, a hypothetical perfect correlation between the two. Red line indicates actual correlation, which fairs well despite different chemistries and methods used.



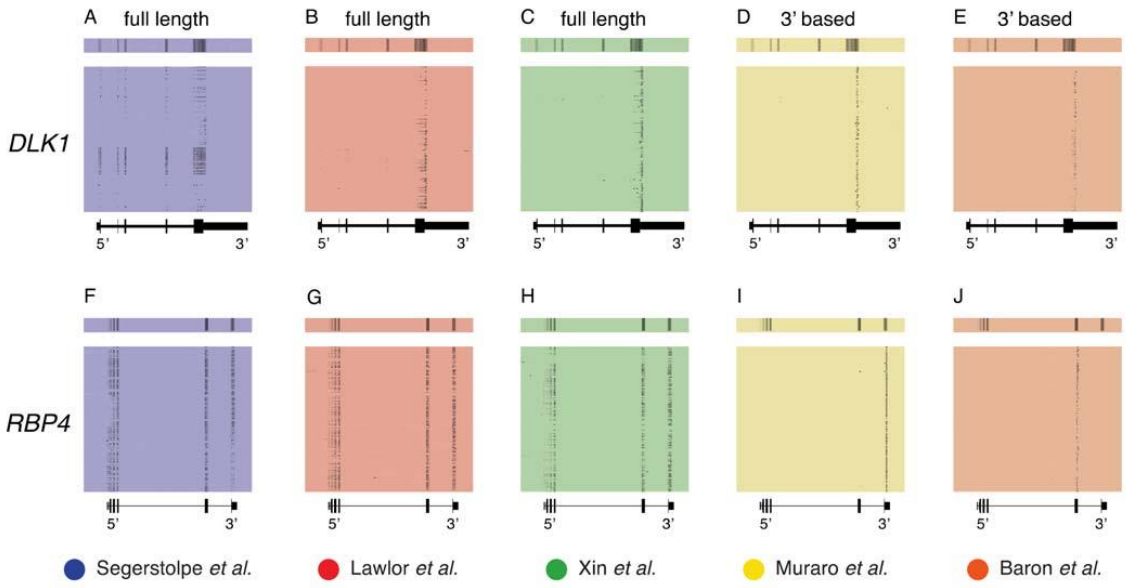
SUPPLEMENTARY DATA

Supplementary Figure S5. Transcriptome-wide representation of the fraction of single human beta cells where each gene is detected. A-E: Genes are ranked in descending order of average expression level; CPM > 0 was used as the detection threshold. The large majority of genes are heterogeneously detected, between 0.005% and 0.83% of the 18,485 genes are detected in all beta cells in any given study.



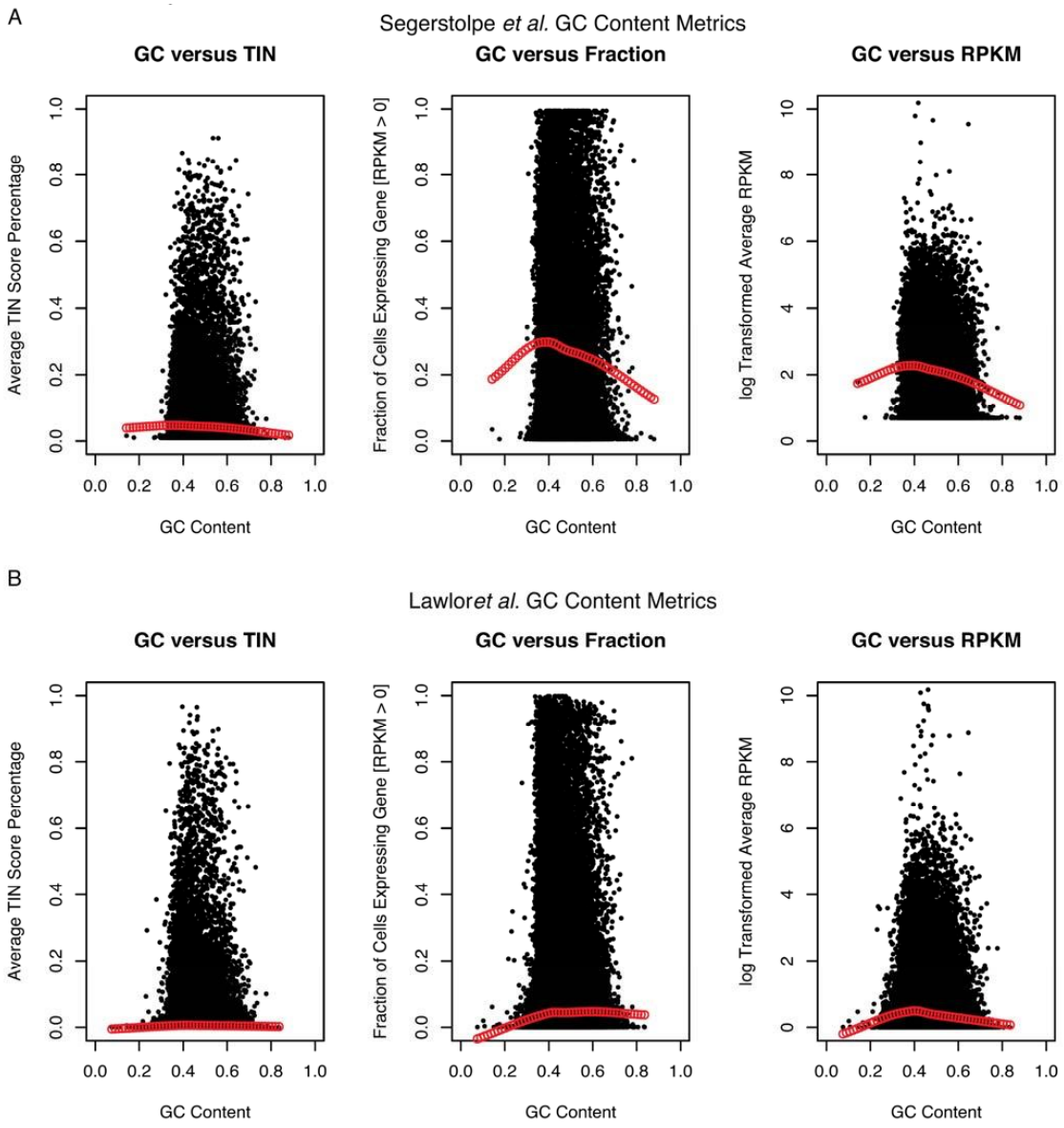
SUPPLEMENTARY DATA

Supplementary Figure S6. Coverage across the gene body of two genes that were self-identified in human beta cell heterogeneity by some of the studies (Table S2). Coverage is shown for each study to compare full-length and 3' based scRNA-Seq. A-E: DLK1 capture across the five studies. A rare and notable example of burst transcription. F-J: RBP4 capture across all studies.



SUPPLEMENTARY DATA

Supplementary Figure S7. Metrics evaluating GC content against fraction of detection, abundance, and TIN scores for two human single cell studies. A-B: Loess regression was applied to determine whether a pattern or trend exists between the three metrics. Kolmogorov-Smirnov testing was applied to evaluate whether GC content negatively impacted detection and abundance. Genes with a GC content below 50% performed better than those with a GC content higher.



SUPPLEMENTARY DATA

Supplementary Table S1. Publication protocol and metrics. A: Details for protocol and metrics across scRNA-Seq or bulk sequencing studies used, noting the platform, library prep, and depth of sequencing applied, as well as the number of donors, beta cell libraries generated, and genes captured.

A

Breakdown of Publication Protocol & Metrics												
Study	Baron *et al.* 2016	Blodgett *et al.* 2015	Lawlor *et al.* 2017	Lawlor *et al.* 2017	Muraro *et al.* 2016	Segerstolpe *et al.* 2016	Segerstolpe *et al.* 2016	Xin *et al.* 2016	Adriaenssens *et al.* 2016	DiGruccio *et al.* 2016	Xin *et al.* 2016	Zeng *et al.* 2017
Approach	scRNA	Bulk RNA	scRNA	Bulk RNA	scRNA	scRNA	Bulk RNA	scRNA	Bulk RNA	Bulk RNA	scRNA	scRNA
Organism	Human	Human	Human	Human	Human	Human	Human	Human	Mouse	Mouse	Mouse	Mouse
Library Prep	CEL-Seq	Tru-Seq	SMART-Seq	SMART-Seq	CEL-Seq2	SMART-Seq2	Tru-Seq	SMART-Seq	NuGen	Tru-Seq	SMART-Seq	SMART-Seq
Platform	inDrop	Intracellular Stain/FACS	Fluidigm C1	Whole Islet	FACS	FACS	Whole Islet	Fluidigm C1	FACS	FACS	Fluidigm C1	FACS & Fluidigm C1
Average Number of Genes [CPM >=1]	~2,000	~13,000	~5,000	~12,000	~4,500	~5,500	~15,000	~5,000	~10,000	~12,000	~4,000	~6,000
Depth	100 k	40 M	3 M	34 M	41 k	750 k	30 M	1 M	40 M	3M	1 M	4.3 M
# Beta Cell Libraries	2,525	6	168	3	448	171	3	207	4	5	400	387
Donors	4	6	5	3	4	6	3	12	4	5	.	15, 14, 10, 4, 4
Donor Ages	17, 51, 38, 59	18, 19, 24, 30, 53, 60	22, 29, 30, 53, 56	22,30,56	54, 23, 48, 59	43, 25, 48, 22, 27, 23	48, 22, 23	23, 32, 23, 56, 27, 68, 29, 60, 24, 43, 31, 56	.	.	3-7 Month	P1, P7, P14, P21, P28
Males	2	5	3	2	3	5	2	7	All Male	.	All Male	.
Females	2	1	2	1	1	1	1	5	.	.	.	.

SUPPLEMENTARY DATA

Supplementary Table S2. Genes highlighted in beta cell heterogeneity within the five human single cell sequencing studies. A: Table of genes mentioned within each of the five studies. B: Genes highlighted in more than one publication, and whether they were detected in the five study overlap of our unbiased meta-analysis.

A

Genes Highlighted in Human Novel Heterogeneity within Single Cell Studies				
Baron et al. 2016	Muraro et al. 2016	Segerstolpe et al. 2016	Lawlor et al. 2017	Xin et al. 2016
DDIT3 (Chop)	FTH1	ADCYAP1 (PACAP)	ADCYAP1 (PACAP)	ADCYAP1 (PACAP)
DLK1	FTH1P3	FFAR4 (GPR120)	DLK1	BMI1
ESR1	FTL	FTH1	HADH	DDIT3 (Chop)
FFAR4 (GPR120)	ID1	ID1	MEG3	DLK1
HERPUD1	PFKFB2	ID2	PFKFB2	HADH
HSPA5	RBP4	ID3	SIX3	MEG3
OLIG1	SIX2	MEG3		RBP4
RXRG	SQSTM1 (p26)	RBP4		RPL7P19
SIX2	SRXN1	SIX2		
SIX3		SIX3		

B

Highlighted Genes Detected by Multiple Studies or Within Meta-Analysis		
Genes with Overlap	# of Studies	Detected in Unbiased Overlap List of 78 Genes?
MEG3	3	No
SIX3	3	No
SIX2	3	No
DLK1	3	Yes
ADCYAP1	3	No
RBP4	3	Yes
PFKFB2	2	No
FTH1	2	No
FFAR4	2	No
DDIT3	2	No
ID1	2	No
HADH	2	No

SUPPLEMENTARY DATA

Supplementary Table S3. Differentially expressed genes between the two identified beta subpopulations. A: Table of genes derived from the meta-analysis, a total of 52 identified as significant (adjusted p-val <= 0.05).

A

Differentially Expressed Genes Between Two Beta Subpopulations					
Gene	p_val	avg_logFC	pct.1	pct.2	p_val_adj
G6PC2	5.96E-166	0.90525112	0.903	0.47	1.19E-162
SYT16	4.82E-159	0.510769	0.751	0.273	9.64E-156
MAFA	2.67E-132	0.68732461	0.86	0.487	5.35E-129
TSPAN1	3.43E-88	0.3537818	0.796	0.418	6.86E-85
ASB9	7.72E-88	0.2960037	0.548	0.163	1.54E-84
EDN3	9.08E-88	0.35234138	0.67	0.26	1.82E-84
IGSF1	1.41E-81	0.25147901	0.595	0.225	2.82E-78
PCSK1	1.31E-79	0.41436648	0.969	0.76	2.61E-76
IAPP	8.65E-75	0.60308077	0.998	0.98	1.73E-71
CDKN1C	1.22E-70	0.38336292	0.911	0.599	2.44E-67
HADH	2.21E-66	0.30462967	0.984	0.834	4.42E-63
SORL1	4.36E-64	0.25603155	0.836	0.483	8.71E-61
HSPA6	3.92E-51	-0.2939855	0.492	0.168	7.85E-48
KLHDC8A	9.65E-47	0.27021187	0.482	0.151	1.93E-43
ACTG1	1.37E-41	-0.9587862	0.961	0.931	2.73E-38
IL11	6.44E-41	-0.3609107	0.143	0.111	1.29E-37
CECR1	5.08E-40	0.25916099	0.506	0.184	1.02E-36
REG1A	2.96E-34	-0.9130366	0.784	0.818	5.92E-31
PRSS2	2.44E-33	-1.0200428	0.573	0.619	4.88E-30
NPY	5.04E-22	-0.2627741	0.582	0.32	1.01E-18
TRIB3	2.53E-19	-0.5622524	0.326	0.288	5.05E-16
PTGS2	4.70E-19	-0.6123507	0.581	0.313	9.40E-16
BIRC3	1.28E-18	-0.2921015	0.215	0.181	2.56E-15
HS3ST2	2.97E-18	-0.2749683	0.139	0.064	5.94E-15
CTRB2	9.09E-17	-0.6023506	0.633	0.606	1.82E-13
SOD2	2.20E-15	-1.7492848	0.579	0.542	4.41E-12
SERPINA1	1.67E-14	-0.5799237	0.495	0.257	3.34E-11
CPA1	6.31E-14	-0.4104326	0.442	0.418	1.26E-10
CELA3A	7.75E-14	-0.7120062	0.598	0.538	1.55E-10
NAMPT	2.08E-13	-0.4521167	0.425	0.41	4.16E-10
SETD7	2.10E-13	-0.262157	0.51	0.269	4.20E-10
TMSB4X	3.81E-13	-0.5192567	0.758	0.684	7.61E-10
PDK4	9.58E-13	-0.4383716	0.713	0.453	1.92E-09
SYNGR4	1.47E-12	-0.5886256	0.776	0.555	2.94E-09
NFKBIA	1.49E-12	-0.7210247	0.473	0.445	2.98E-09
CTRB1	1.83E-11	-0.4127868	0.447	0.401	3.67E-08
PRG4	4.06E-11	-0.5320836	0.245	0.172	8.12E-08
DNAJB1	7.82E-11	-0.2973403	0.644	0.413	1.56E-07
PPY	1.40E-10	-1.1106356	0.628	0.384	2.81E-07
S100A11	1.43E-10	-0.6617551	0.805	0.715	2.86E-07
AQP3	2.08E-10	-0.3636973	0.659	0.422	4.16E-07
HMOX1	2.75E-10	-0.4557159	0.298	0.217	5.50E-07
C10orf10	1.60E-09	-0.312835	0.831	0.617	3.20E-06
CPB1	6.81E-09	-0.4639012	0.355	0.3	1.36E-05
TIMP1	4.17E-08	-0.9600852	0.918	0.831	8.35E-05
GADD45B	4.92E-08	-0.548217	0.579	0.37	9.83E-05
PRSS1	1.38E-07	-0.5119934	0.509	0.441	0.00027512
PTPRH	1.74E-07	-0.2929226	0.539	0.313	0.00034756
SST	1.82E-07	-1.5593211	0.985	0.866	0.00036366
SPP1	1.38E-06	-0.6511131	0.492	0.29	0.00276214
PDGFA	1.50E-06	-0.292609	0.269	0.175	0.00300385
SAT1	6.54E-06	-0.4977158	0.524	0.46	0.01307116

SUPPLEMENTARY DATA

Supplementary Table S4. Overlapping genes identified when re-analyzing only the healthy beta cells. A: Only 24 genes were identified by all studies as implicated in heterogeneity.

A

Gene
INHBA
NEAT1
IGFBP5
SPP1
RBP4
TM4SF4
AGT
RGS4
C10orf10
DLK1
NPY
PCDHB11
AKAP12
ITGB8
GCG
TRIB3
CLU
ANXA3
FOS
PEX5L
EXPH5
HERPUD1
ZNF605
EFR3B

SUPPLEMENTARY DATA

Supplementary Table S5. Protein quantification markers for human and mouse beta cells. A: References used to quantify protein presence for known human beta cell gene markers. B: References used to quantify protein presence for known mouse beta cell gene markers.

A

Human Immunostaining References to Estimate Fraction of Detection for Protein of Interest				
Protein	Journal	Author	Year	DOI
INS	Diabetes	Okamoto *et al.* ...Tokunaga	2012	10.2337/db11-1201
ABCC8	Diabetologia	Guiot *et al.* ...Sempoux	2007	10.1007/s00125-007-0731-z
VAMP2	Oncotarget	Peart *et al.* ...Wang	2017	10.18632/oncotarget.17969
SNAP25	Oncotarget	Peart *et al.* ...Wang	2017	10.18632/oncotarget.17969
PAX6	PLoS ONE	Ahmad *et al.* ...Mansouri	2015	10.1371/journal.pone.0144597
NKX6.1	Cell Reports	Taylor *et al.* ...Sander	2013	10.1016/j.celrep.2013.08.010
KCNJ11	Diabetes	Yan *et al.* ...Kohler	2004	10.2337/diabetes.53.3.597
GLP1R	Journal of Histochemistry & Cytochemistry	Tornehave *et al.* ...Heller	2008	10.1369/jhc.2008.951319
KCNB1	Diabetes	Yan *et al.* ...Kohler	2004	10.2337/diabetes.53.3.597
G6PC2	PNAS	Hutton and Eisenbarth	2003	10.1073/pnas.1633447100
GCK	Diabetes	Arden *et al.* ...Agius	2004	10.2337/diabetes.53.9.2346
UCN3	Nature Biotechnology	Blum *et al.* ...Melton	2012	10.1038/nbt.2141
ERO1B	Journal of Cell Biology	Zito *et al.* ...Ron	2010	10.1083/jcb.200911086
CHGA	Cell Reports	Taylor *et al.* ...Sander	2013	10.1016/j.celrep.2013.08.010
MAFA	PLoS ONE	Bonnavion *et al.* ...Zhang	2013	10.1371/journal.pone.0072194
PDX1	Diabetologia	Dai *et al.* ...Powers	2012	10.1007/s00125-011-2369-0
CHRM3	Diabetes	Molina *et al.* ...Caicedo	2014	10.2337/db13-1371
SLC2A1	Molecular Genetics & Metabolism	McCulloch *et al.* ...Gloyn	2011	10.1016/j.ymgme.2011.08.026
CD9	Nature Communications	Dorrell *et al.* ...Grompe	2016	10.1038/ncomms11756
ST8SIA1	Nature Communications	Dorrell *et al.* ...Grompe	2016	10.1038/ncomms11756

B

Mouse Immunostaining References to Estimate Fraction of Detection for Protein of Interest				
Protein	Journal	Author	Year	DOI
Ins2	Cell Metabolism	van der Meulen *et al.* ...Huising	2017	10.1016/j.cmet.2017.03.017
Pax6	Diabetologia	Dai *et al.* ...Powers	2012	10.1007/s00125-011-2369-0
Glp1r	Journal of Histochemistry & Cytochemistry	Tornehave *et al.* ...Heller	2008	10.1369/jhc.2008.951319
G6pc2	PNAS	Hutton and Eisenbarth	2003	10.1073/pnas.1633447100
Slc2a2	Nature Genetics	Guillam *et al.* ...Thorens	1997	10.1038/ng1197-327
Ucn3	Nature Biotechnology	Blum *et al.* ...Melton	2012	10.1038/nbt.2141
Ero1lb	Journal of Cell Biology	Zito *et al.* ...Ron	2010	10.1083/jcb.200911086
Nkx6.1	Oncotarget	Peart *et al.* ...Wang	2017	10.18632/oncotarget.17969
Snap25	Oncotarget	Peart *et al.* ...Wang	2017	10.18632/oncotarget.17969
Pdx1	Endocrinology	Szabat *et al.* ...Johnson	2009	10.1210/en.2008-1224
Abcc8	Diabetologia	Guiot *et al.* ...Sempoux	2007	10.1007/s00125-007-0731-z
MafA	Diabetes	Artner *et al.* ...Stein	2010	10.2337/db10-0190
Nkx2.2	Oncotarget	Peart *et al.* ...Wang	2017	10.18632/oncotarget.17969
Gck	Diabetes	Arden *et al.* ...Agius	2004	10.2337/diabetes.53.9.2346
Vamp2	Oncotarget	Peart *et al.* ...Wang	2017	10.18632/oncotarget.17969
Igf1r	Cell Metabolism	Aguayo-Mazzucato and Bonner-Weir	2017	10.1016/j.cmet.2017.08.007

SUPPLEMENTARY DATA

Supplementary Table S6. Accession information for all datasets analyzed. Separated by species and approach used.

Human Single Cell Accession Information	
Study	**Accession**
Baron *et al.* 2016	GSE84133
Lawlor *et al*. 2017	GSE86469
Muraro *et al.* 2016	GSE85241
Segerstolpe *et al.* 2016	E-MTAB-5061
Xin *et al.* 2016	GSE81608

Human Bulk Cell Accession Information	
Study	**Accession**
Blodgett et al. 2015	GSE67543
Lawlor *et al*. 2017	GSE86468
Segerstolpe *et al.* 2016	E-MTAB-5060

Mouse Single Cell Accession Information	
Study	**Accession**
Xin *et al.* 2016	GSE77980
Zeng *et al.* 2017	GSE86479

Mouse Bulk Cell Accession Information	
Study	**Accession**
Adriaenssens et al. 2016	GSE76017
DiGruccio *et al*. 2016	GSE80673

SUPPLEMENTARY DATA

Supplemental Computational Approach & Tools

Data Acquisition

Raw Fastq files were pulled from GEO accession numbers for all studies evaluated. A table of datasets used and where they were retrieved from is located in Table S4. For inDrop (2) and Cel-Seq2 (3), demultiplexing and barcode identification was required prior to alignment and quantification. Demultiplexing and barcode splitting was performed as previously described by each respective study.

Alignment & Quantification

All raw Fastq files were aligned using STAR's (4) default parameters, with the exception of a tighter mismatch restriction of no greater than 3% per read, to the Gencode GRCm38.p4 M8 (mm10) mouse genome or GRCh38.p5 (hg38) human genome (5). Bigwigs were generated using STAR's wiggle output option and UCSC's Genome Utilities (6). Gene-level quantification was performed on all samples' sorted BAM files using featureCounts (7) default parameters, counted by Gencode defined exon's, and aggregated to the gene level. TIN scores were derived from BAM files using the RSeQC's (8) TIN quantification tool.

Clustering

Unbiased clustering was initially performed on the integrated five study's' healthy donors using the R Bioconductor package Seurat v3 (9), and UMAP dimensional reduction. Clusters were identified for cell type using known gene markers and validated against each study's self-assigned labels for each cell. The two beta cell subpopulations were further analyzed using differential expression testing to identify genes driving heterogeneity between the two, using an adjusted p-value cutoff <= 0.05. 52 genes were derived from this list.

Unbiased clustering was performed a second time around on only the healthy beta cells from the same five studies, using the same approach. The top 2000 highly variable genes that were identified as clustering driving per study were retrieved through this method. All plots visualizing clustering were generated through Seurat.

Downstream Analyses

All Venn diagrams were generated using the R library package VennDiagram. All violin plots and abacus-style plots used to show fraction of genes detected within single cell beta libraries were generated using ggplot2. KS testing was performed using the R package coin (10). Density plots were generated using the R package sm.

References

1. Segerstolpe, A., et al., Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. Cell Metab, 2016. 24(4):593-607.
2. Baron, M., et al., A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. Cell Syst, 2016. 3(4):346-360 e4.
3. Muraro, M.J., et al., A single-cell transcriptome atlas of the human pancreas. Cell Syst, 2016. 3(4):385-394 e3.
4. Dobin, A., et al., Star: Ultrafast universal rna-seq aligner. Bioinformatics, 2013. 29(1):15-21.
5. Harrow, J., et al., Gencode: Producing a reference annotation for encode. Genome Biol, 2006. 7 Suppl 1:S4 1-9.
6. Kent, W.J., et al., Bigwig and bigbed: Enabling browsing of large distributed datasets. Bioinformatics, 2010. 26(17):2204-7.
7. Liao, Y., G.K. Smyth, and W. Shi, Featurecounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics, 2014. 30(7):923-30.

SUPPLEMENTARY DATA

8. Wang, L., et al., Measure transcript integrity using rna-seq data. BMC Bioinformatics, 2016. 17:58.
9. Butler, A., et al., Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol, 2018. 36(5):411-420.
10. Hothorn, T., et al., *Implementing a class of permutation tests: The coin package*. Journal of statistical software. Vol. 28. 2008. 1-23.

VII. Chapter 7: Chromatin accessibility differences between alpha, beta, and delta cells identifies common endocrine- and cell-specific enhancers

Alex M. Mawla, Talitha van der Meulen, and Mark O. Huising

Final draft, to be submitted for publication in BMC Genomics.

Contributions to Jointly Authored Works: As first author, I was responsible for the entirety of analysis of and interpretation of ATAC-Seq and transcriptomic data generated in-house and incorporated into this manuscript. I was also responsible for the review, selection, and uniform re-analysis of all other aggregated -omics data (histone and transcription factor) integrated into this manuscript in order to delineate murine pancreatic enhancers. I interpreted all results, created all figures, and wrote the manuscript. I also am responsible for creating an interactive resource of these data on our lab website so that colleagues in the field can query their data against our findings, and also integrate their data to quickly generate high-quality visuals.

Significance of Research: This manuscript confirmed previous findings in the field suggesting that alpha cells are epigenetically poised, but repressed, from becoming beta cells. This manuscript is the first to evaluated murine delta cell chromatin accessibility, and also goes on to suggest the novel finding that delta cells are also epigenetically poised, much like alpha cells, to become beta cells. This supports previous research showing that delta cells can spontaneously transdifferentiate and become beta cells in juveniles and supports previous findings in our lab.

Chromatin accessibility differences between alpha, beta, and delta cells identifies common and cell type-specific enhancers

Alex M. Mawla[1], Talitha van der Meulen[1], and Mark O. Huising[1,2*]

[1]Department of Neurobiology, Physiology & Behavior, College of Biological Sciences, University of California, Davis, CA 95616, USA

[2]Department of Physiology and Membrane Biology, School of Medicine, University of California, Davis, CA 95616, USA

* Corresponding author:

Dr. Mark O. Huising, Ph.D.

Professor

University of California, Davis

Department of Neurobiology, Physiology & Behavior, College of Biological Sciences

Department of Physiology and Membrane Biology, School of Medicine

One Shields Avenue,

Davis, CA, 95616

Ph: (530) 752-4670

Email: mhuising@ucdavis.edu

Abstract

High throughput sequencing has enabled the interrogation of the transcriptomic landscape of glucagon-secreting alpha cells, insulin-secreting beta cells, and somatostatin-secreting delta cells. These approaches have furthered our understanding of expression patterns that define healthy or diseased islet cell types and helped explicate some of the intricacies between major islet cell crosstalk and glucose regulation. All three endocrine cell types derive from a common pancreatic progenitor, yet alpha and beta cells have partially opposing functions, and delta cells modulate and control insulin and glucagon release. While gene signatures that define and maintain cellular identity have been widely explored, the underlying epigenetic components are incompletely characterized and understood. Chromatin accessibility and remodeling is a dynamic attribute that plays a critical role to determine and maintain cellular identity. Here, we compare and contrast the chromatin landscape between mouse alpha, beta, and delta cells using ATAC-Seq to evaluate the significant differences in chromatin accessibility. The similarities and differences in chromatin accessibility between these related islet endocrine cells help define their fate in support of their distinct functional roles. We identify patterns that suggest that both alpha and delta cells are poised, but repressed, from becoming beta-like. We also identify patterns in differentially enriched chromatin that have transcription factor motifs preferentially associated with different regions of the genome. Finally, we identify and visualize both novel and previously discovered common endocrine- and cell specific- enhancer regions across differentially enriched chromatin.

Introduction

The evaluation of the transcriptional landscape through high-throughput bulk sequencing (bulkSeq) in both mouse and human of major islet cell types has granted a deeper understanding of cellular identity and intercellular crosstalk within the pancreas. This has enabled the detection of distinct gene pattern signatures between major islet cell types in mouse and human [1-6]. However, gene expression represents the final outcome of a complex layer of genetic and epigenetic factors that determine islet cell fate [7-9] and identity [10, 11]. Previous studies have explored pancreatic islet cellular identity by evaluating epigenomic features such as methylation [12-14], histone modifications [15-18], and enhancer regulatory regions [19-24]. While each of these factors contributes to defining and maintaining cell fate and identity, connecting chromatin accessibility differences to epigenetic factors promises to provide further insight into outstanding questions within the field.

Chromatin remodeling is a central epigenetic regulator that can be surveyed in order to better understand cell states [20, 25-28]. The accessibility of chromatin via changes between euchromatin and heterochromatin, and nucleosome occupancy, plays a significant role in cell lineage, and in tissue- and cell-specific gene expression [11, 25, 29]. Epigenetic stability is required for the maintenance of islet cell identity, while changes in chromatin accessibility are associated with perturbations in gene expression due to disease [7, 22, 30]. Chromatin accessibility in tandem with other epigenetic factors at promoter-proximal regions [29, 31] of a gene allows for direct activation or repression of transcription. In contrast, open chromatin at exonic [32], intronic [33], or distal-intergenic regions [34] can be accessed by regulatory factors that act as nearby or distal enhancers that govern lineage branching and stable cell fate.

Assay for transposase-accessible chromatin using sequencing (ATAC-Seq) allows for the unbiased, modification-independent evaluation of chromatin accessibility within cell types and can be run with relatively small sample input [30, 35]. Previous studies have explored chromatin accessibility in healthy [5, 11, 36, 37] and T2D [22, 23, 38, 39] islets as well as pancreatic progenitors [9] using bulkSeq through

human antibody panels alongside FACS-purification or through single-cell sequencing (scATACSeq) [40-42]. However, none of these studies have explored pancreatic islet cell chromatin accessibility from mouse FACS-purified alpha, beta, and delta cells. Therefore, to better understand endocrine islet cell identity between mouse alpha, beta, and delta cells, we compared chromatin accessibility and transcriptome data for FACS-purified mouse alpha, beta, and delta cells sorted from pancreatic islets from triple transgenic reporter strains - mIns1-H2b-mCherry beta cells crossed to mice with alpha or delta cells marked by YFP in a Cre-dependent fashion - that we generated for this purpose [1, 6]. This approach allowed for the direct comparison between ATAC-Seq and RNA-Seq datasets from alpha, beta, and delta cells from these lines.

We integrated our ATAC-Seq data with high-quality transcription factor and histone binding data from other mouse pancreatic islet studies to evaluate how transcriptional activators and repressors may collectively regulate differential gene expression at promoter-proximal regions. To support the visualization and integration of our ATAC-Seq chromatin data and previously published transcriptome of the FACS-purified alpha, beta, and delta cells alongside select epigenomic datasets from histone marker and transcription factor Chromatin Immuno Precipitation (ChIP) data, we developed an R package, *epiRomics* [See: https://github.com/Huising-Lab/epiRomics]. This package is a novel, publicly available resource that is described in detail elsewhere [43]. *epiRomics* allows for the visualization of integrated epigenomic data and visualizes putative enhancer regions without the requirement for extensive bio-informatics experience, with the intent of enabling more of our colleagues to tease apart key regions that may drive cell fate switching and maintenance between the major islet endocrine cell types. Through this approach we identified putative enhancer regions at distal-intergenic regions common to all cell types as well as regions selectively accessible only in a single islet cell type and confirmed previously identified mouse pancreatic islet enhancers.

Methods

Islet isolation and FACS sorting

mIns1-H2b-mCherry [1] x Rosa-LSL-YFP crossed to either Sst-Cre [44] or Gcg-Cre [45] triple transgenic mice were pooled by sex, each sample yielding a median of 20,000 cells, with islet isolation and FACS-sorting as previously described (Supplemental Fig. 1) [1, 46-48].

Assay for transposase-accessible chromatin using sequencing

Single-end 50 bp reads were generated after library size selection yielded an average of 450 bp fragments and sequenced as previously described using NexteraDNA library protocol [30].

Alignment and differential peak calling

Reads from each replicate (Supplemental Table 1) were evaluated for quality control and trimmed using FastQC and Trimmomatic, respectively [49-51]. A modified index of Gencode GRCm38.p4 (mm10) was built to exclude mitochondrial DNA prior to aligning reads with Bowtie 2 [52, 53]. Post-alignment, duplicates were marked using Picard Tools, blacklist regions were removed, and BAM files were converted into tagAlign format for downstream use. Peak calling and bigwig generation was done using MACS2 [54]. Differential expression testing was performed using DiffBind's edgeR method [55, 56].

Quality control and validation

Quality control metrics were evaluated within raw reads as well as peak calls and compared against ENCODE standards for fraction of reads in peaks (FRiP), leading to the removal of one beta replicate with a FRiP score far below 0.3 (Supplemental Table 1; Fig. 1A-C) [57, 58].

Downstream analysis

Transcription factor footprinting analysis and validation against existing ChIP data was performed through a modified script utilizing chromVar [59], regioneR [60], GenomicRanges [61], and motifmatchr [62] using the HOCOMOCO database [63]. Pathways analysis on differential chromatin accessibility was performed using the R Bioconductor packages ChIPseeker [64], ReactomePA [65], and clusterProfiler [66].

Enhancer Identification

We developed a novel R package, *epiRomics*, to integrate our chromatin accessibility data alongside aggregated pancreatic islet ChIP and histone data to identify putative enhancer regions, as described [43]. The package, example data, and vignette can be found at: https://github.com/Huising-Lab/epiRomics and an interactive browser of the results from this manuscript is publicly available at: https://www.huisinglab.com/epiRomics_2021/index.html.

Integrated data

Mouse alpha, beta, and delta (GEO: GSE80673), alongside alpha- and delta- transdifferentiated beta (GEO: GSE88778) transcriptomes were integrated into this analysis [6, 67]. Aggregated ChIP datasets of transcription factors and histone marks were added to the analysis through *epiRomics* [43] to identify putative enhancer regions (Supplemental Table 2).

Results

ATAC-Seq validation

To determine whether chromatin accessibility patterns differed between islet endocrine cell types, principal component analysis (PCA) was applied to peak calls across all samples. This confirmed that replicates clustered by cell type (Fig. 1A), a finding that was further validated by heatmaps using all defined peaks across replicates (Fig. 1B). Alongside quality control applied through the generation and analysis of this dataset, the fraction of reads in peaks (FRiP) score was in excess of the commonly applied benchmark of 30% (Fig. 1C). Furthermore, the FRiP score was independent of variability in unique read depth, indicating that peak calls were reproducible across all replicates within cell types independent of read depth range.

Validation of islet cell chromatin accessibility data coupled to companion transcriptomes

After preliminary validation of our derived ATAC-Seq data, we checked for the presence of chromatin peak enrichment for alpha, beta, and delta marker genes that have been previously well-established and validated through complementary bench-lab or computational methods. We expected that if a gene is expressed within a cell type, its ATAC signal near the transcription start site (TSS) at promoter-proximal regions should reflect chromatin accessibility. Indeed, cell type-specific chromatin accessibility correlated with gene expression of *Ins2*, *Gcg*, and *Sst* genes for beta, alpha, and delta cells, respectively (Fig. 1D-F) [4, 6, 68]. After confirming chromatin accessibility in key cell-identity markers, we sought to compare and contrast select regions identified from prior groups that evaluated chromatin accessibility in

human islets [5, 12, 38], as well as to further query whether chromatin was always uniquely enriched on a panel of cell type-specific genes across alpha (*Arx, Ttr, Gc)*, beta (*Ucn3, MafA, Pdx1)*, and delta cells (*Pdx1, Hhex, Rbp4, Ghsr*) (Fig. 2; Supplemental Fig. 2). Each of these genes demonstrated overall strong concordance between cell type-enriched gene expression and cell type-specific enrichment of available chromatin. This validated the utility of ATAC-Seq data to detect epigenetic factors that determine gene expression.

The chromatin landscape of the annotated genome across cell types

As genes make up a small fraction of the entire genome, we determined the overall distribution of peaks across the annotated genome within each cell type. We defined five regions of interest to further explore – promoter-proximal, intronic, exonic, downstream, or distal-intergenic (Fig. 3A). We identified a consensus set of 124,494 peaks marking open chromatin through the R package DiffBind. This number is comparable to the number of open regions found in previous studies of pancreatic islet chromatin accessibility [11, 38-40] (Supplemental Dataset 1). We then evaluated the distribution of called peaks present in at least one replicate within 3kb upstream of the TSS and confirmed that a majority of genes enriched in each islet cell type were accompanied by promoter-proximal peaks (Fig. 3B-D). The distribution of ATAC-Seq peaks across different pre-defined genomic areas was overall similar across alpha, beta, and delta cells. For each endocrine cell type between 21.98-24.88% of open chromatin was promoter-proximal, whereas promoter-proximal areas account for 2.41% of the mouse genome. A further 34.92-38.33% of peaks for all cell types were found on distal-intergenic regions, which was proportional to the fraction of the genome that falls into this category (Fig. 3E-G). Finally, we noted that between 33.07- 33.65% of peaks occurred on intronic regions (first or other), relative to the 37.7% of the mouse genome classified as intronic [69].

Regional differences and characteristics of differentially enriched chromatin

As our overall distribution of ATAC-Seq peaks across different genomic regions was consistent across alpha, beta, and delta cells, we compared differential chromatin accessibility between these cell types in greater detail. To this end, we performed pairwise differential ATAC-Seq peak enrichment testing across alpha, beta, and delta cells. Out of 124,494 identified consensus regions of open chromatin across the three cell types, 18,409 (14.8%) differentially enriched peaks (p-value <= 0.05) were identified between alpha and beta (Fig. 4A), 12,722 (10.2%) between alpha and delta (Fig. 4B), and 16,913 (14.6%) between beta and delta cells (Fig. 4C).

After performing differential peak enrichment testing, we discovered that 22.89% of all differentially enriched peaks between alpha and beta cells were promoter-proximal (0-3kb) (Fig. 5A). A further 33.22% of differential peaks were linked to distal-intergenic regions and another 33.61% of differential peaks were intronic (first and other combined) (Fig. 5A). This assessment of differential peaks without considering the direction of enrichment revealed no major difference with overall peak distribution described earlier (Fig. 3). However, when factoring in the direction of enrichment we observed that 35.08% of alpha cell-enriched peaks was promoter-proximal. In contrast, only 12.5 % of beta cell enriched peaks occurred in promoter-proximal areas (Fig. 5B). Instead, a majority of ATAC peaks enriched in beta cells were located at distal-intergenic regions (45.41%) (Fig. 5B).

Between alpha and delta cells, we identified that 21.29% of differentially enriched peaks occurred promoter-proximally. Another 36.4% of peaks occurred on distal intergenic regions and 35.5% on intronic regions (Fig. 5C). A similar preference of alpha cell-enriched peaks in promotor-proximal regions was evident when comparing alpha to delta cells, with 30.33% of all enriched alpha peaks occurring promoter-proximally, but only 9.56% of delta cell peaks. Instead, 38.41% of delta cell enriched peaks were distal-intergenic (Fig. 5D).

Lastly, between beta and delta cells, 26.62% of all differentially enriched peaks were promoter-proximal, 32.2% distal intergenic, and 34.33% on intronic regions (Fig. 5E). Further break down revealed a bias towards distal-intergenic enriched peaks within beta cells (42.86%), as opposed to promoter-proximal peaks in delta cells (28.20%) (Fig. 5F).

Differential chromatin enrichment in the majority of cases correlates with gene expression

So far, we detected a disproportionate fraction of peaks associated with promoter-proximal regions in general (Fig. 3). Moreover, ATAC-Seq peaks that were differentially enriched in alpha and - to a lesser extent - delta cells were considerably more likely to occur at promoter-proximal sites. Instead, peaks enriched in beta cells more likely occurred at distal intergenic regions (Fig. 5). Therefore, we determined whether the enrichment of promoter-proximal peaks correlated with increased expression of the corresponding. Genes with increased expression in a cell type accompanied by a significantly enriched ATAC-Seq peak proximal to its TSS were considered 'congruent' genes (Fig. 6A). The underlying mechanism in such a scenario might be the presence of transcriptional activators at the promoter-proximal site that promote gene expression. Conversely, genes with a significantly enriched ATAC-Seq peak proximal to its TSS accompanying a reduction in corresponding gene expression were considered 'incongruent' genes (Fig. 6B). The underlying mechanism for these genes might be the presence of transcriptional repressors at the promoter that prevent gene expression (Supplemental Dataset 2) [70-73]. Finally, genes that had significantly enriched chromatin in either cell type, but no evidence of mRNA expression were considered 'unexpressed' (Fig. 6C).

When we compared differentially enriched TSS-associated chromatin against corresponding gene expression between alpha and beta cells, we found that in the majority of cases (86%), differential chromatin enrichment on TSS regions successfully captured the epigenetics of gene regulation. Exactly, 50% of genes with differentially enriched chromatin at the TSS had a corresponding increase in gene expression within

the same cell type (congruent genes). 36% showed TSS chromatin accessibility enrichment, but with a reduction in gene expression for each cell (incongruent genes - either alpha repressed (33%), or beta repressed (3%)). Strikingly, a substantial majority of the incongruent genes in this comparison were alpha repressed. Finally, only 14% of all genes with differentially enriched TSS chromatin showing no expression in either cell type (unexpressed) (Fig. 6D). A further visualization of select gene expression against TSS-associated chromatin accessibility indicated the majority as congruent, with a highlighted example of an incongruent (putatively alpha repressed or alpha cell poised) gene observed in the alpha cell TSS enrichment for the beta-specific genes *MafA* (Fig. 2E; Supplemental Fig. 3A). *MafA* is a key transcription factor enriched in beta cells yet shows abundant chromatin accessibility in alpha cells.

We observed a similar distribution between congruent (55%), incongruent (24%), and no expression genes (20%), between alpha and delta cells. We noted a more uniform distribution between alpha (14%) and delta (10%) repressed genes. (Fig. 6E). Upon visualizing gene expression and chromatin accessibility, we confirmed congruent gene expression and TSS chromatin accessibility of key transcription factors known to regulate both alpha – *MafB*, *Ttr*, and *Arx* – and delta – *Pdx1* and *Hhex* – cell fate (Supplemental Fig. 3B).

For our final pairwise comparison between beta and delta cells, we again found a similar fraction of congruent (57%), incongruent (32%), and no expression (11%) genes (Fig. 6F). We noted a minor fraction of repressed genes with open chromatin in beta cells (1.5%), with the overwhelming majority of repressed genes corresponding to delta cells (30.45%), similar to the pattern seen in alpha repressed genes between alpha and beta cells. Further visualization of select marker gene expression against chromatin accessibility showed generally good congruence between chromatin accessibility at the TSS and gene expression (Supplemental Fig. 3C).

Poised genes are enriched in beta cells with a non-beta cell lineage history

To further interrogate whether these alpha- or delta- repressed genes could be poised beta cell genes, we incorporated transcriptome data from beta cells with an alpha- or delta- cell lineage history [67] – also from our companion RNA-Seq experiment. These cells, termed "transdifferentiated," are functionally mature beta cells (defined by the presence of *Ucn3*), but have either a *Gcg*- or *Sst*-Cre lineage label, reflective of a lineage history as an alpha or delta cell, respectively. We reasoned that if alpha- or delta-repressed genes are poised beta cell genes, we should expect to observe a stepwise transition in gene expression levels, showing little or no expression in either alpha or delta cells, to intermediate expression in alpha- or delta- transdifferentiated cells, and full expression in beta cells. We confirmed that the majority (83.6%) of alpha-repressed genes showed intermediate expression in the alpha-to-beta-transdifferentiated population, and the highest expression in beta cells. A subset of genes (16.4%) showed the highest expression in the alpha-transdifferentiated population (Fig. 7A). We observed a similar pattern between delta, delta-transdifferentiated, and beta cells; however, only half (50.18%) of delta repressed genes demonstrated an intermediate expression in the delta-to-beta-transdifferentiated population and the highest in beta (Fig. 7B). The remainder of the genes showed the highest expression in delta-to-beta-transdifferentiated cells.

Differential meta-chromatin enrichment testing

Given that in a majority of cases, TSS-associated chromatin recapitulated the underlying regulation of gene expression, we inquired whether differentially enriched chromatin peaks were associated with genes concentrated in pathways or gene networks that would better reflect our understanding of the biology across these different islet endocrine cell types. Between alpha and beta cells, KEGG set pathway testing of differentially accessible chromatin identified pathways related to protein digestion and absorption and cell adhesion molecules unique to beta cells, Hippo, Wnt, and ubiquitin-mediated proteolysis unique to alpha cells, and MAPK, axon guidance, and cAMP pathways enriched within both (Supplemental Fig. 4A-B). Upon comparing the differentially accessibly chromatin between alpha and delta cells, adherens junctions

appeared selective to delta cells, while no pathways were enriched specifically in alpha cells. MAPK, axon guidance, and Ras signaling pathways showed general enrichment of associated peaks within both alpha and delta cells (Supplemental Fig. 5A-B). Lastly, in beta and delta cells, a pairwise analysis of differentially accessible chromatin identified the Glycosaminoglycan (GAG) biosynthesis pathway as unique to beta cells - where GAG metabolism and biosynthesis impairment has been linked to beta cell dysregulation [74], adherens junctions and Rap1 signaling pathways unique to delta cells, and MAPK, axon guidance, and cAMP signaling pathways enriched within both (Supplemental Fig. 6A-B).

Islet transcription factor ChIP-Seq binding correlates with open chromatin

After exploring the interrelationship between accessible chromatin and gene expression, we expanded our approach to include additional epigenetic controls to the regulation of islet cell gene expression. We therefore aggregated high-quality, mouse pancreatic islet transcription factor binding data via ChIP-Seq - Pdx1 [75], Nkx6-1 [76], Neurod1 [77], Insm1 [77], Foxa2 [77], Nkx2-2 [78], Rfx6 [79], MafA [24], Isl1 [80], Kat2b [81], Ldb1 [80], and Gata6 [82] - and asked what fraction of open chromatin – as defined by our consensus ATAC-Seq peak set – containing binding sites for each respective transcription factor. The transcription factors Foxa2 (29.07%), Insm1 (28.40%), and Neurod1 (20.09%) had the highest percentage of ChIP-confirmed binding site overlap with open chromatin. This provided further support that our ATAC-Seq data was of high quality and suggested that open chromatin is a reliable indicator of epigenetic regulation (Supplemental Table 3A). To further explore whether these aggregated transcription factor ChIP data convey epigenetic relevance, we queried what fraction of total ChIP binding sites overlapped with open chromatin. Indeed, we observed that in several cases, over 50% of transcription factor ChIP binding sites overlapped with open chromatin, with the transcription factors Nkx2.2 (63.79%), Neurod1 (55.07%), and Insm1 (51.49%) showing the greatest degree of overlap (Supplemental Table 3B). These results supported our findings that open chromatin reflects epigenetic regulation in alpha, beta, and delta cells, in part through the binding of transcription factors.

Transcription factor motif finding suggests genomic preferences at differentially enriched chromatin between cell types

After observing a strong degree of overlap of known islet transcription factor binding on open chromatin, we conducted an unbiased evaluation whether DNA motifs for their respective transcription factor proteins were differentially enriched on ATAC peaks in promoter, intronic, exonic, downstream, or distal regions. We included transcription factors with known DNA-binding motifs to determine if they were more likely to occur at specific areas of the genome. We required that the transcription factor associated with the DNA sequence motif considered is expressed (RPKM>0) in the cell type with chromatin-motif association.

Motifs for key transcription factors involved in beta cell identity, such as MafA, were present ubiquitously across most functional regions we defined (promoter-proximal, intronic, exonic, downstream, and distal intergenic) (Fig. 8A). In contrast, the motifs for cell-identity drivers Irx2 [4] were concentrated at the promoter-proximal regions of chromatin peaks associated with genes differentially expressed by alpha cells. Insm1 [77] motifs were concentrated at the promoter-proximal regions of chromatin peaks associated with genes differentially expressed by beta cells. In another example, DNA-binding motifs associated with the ubiquitous islet transcription factor Pax6 [83] were concentrated on intronic chromatin.

We performed the same transcription factor footprinting test between alpha and delta cells (Fig. 8B). Of note, the motif for Pbx3, a transcription factor driving *Sst* expression in delta cells [84], was enriched in accessible chromatin at promoter-proximal regions. The motif for Stat4, recently implicated in establishing alpha cell identity [85] was concentrated exonic chromatin. Lastly, the motif for Ptf1a, a transcription factor identified in early pancreatic endocrine cell development [86], was preferentially associated with areas of open chromatin at distal intergenic regions.

Between beta and delta cells, no single transcription factor motif overlapped across all five functional categories, nor were there any unique to downstream or exonic regions, as we observed in the prior alpha and delta comparisons (Fig. 8C). Of note, the motif for Smad3, a transcription factor important for islet development [87], as well as the negative regulation of insulin secretion in beta cells via occupancy of the insulin promoter [88], was concentrated in promoter-proximal accessible chromatin. Motifs for Insm1 [77] and Nkx6.1 [89], both key beta cell identity transcription factors, were preferentially associated with accessible chromatin at intronic regions. Lastly, motifs for Fev – recently identified as important for the development and differentiation of the endocrine lineage [90] - and Atf3 – linked to enhancer regions in EndoC-bH1 cells [11] - were enriched in accessible chromatin at distal regions.

Validating motif calls against aggregated islet ChIP datasets

As we observed motif binding site preferences across promoter, intronic, exonic, downstream, or distal chromatin regions, we wished to confirm how accurately predictive DNA motif binding sites conveys true transcription factor binding. To do so, we once again turned to our aggregated pancreatic islet ChIP datasets. We applied the same motif detection method as above on individual ChIP datasets and on all open chromatin – as derived from our ATAC-Seq consensus peak set –and assessed how well predicted motif binding overlapped with true ChIP peaks from our selected list of ChIP-Seq data. We observed strong (57%) true positive and low (8.34%) false negative values for the Rfx DNA-motif's ability to predict all Rfx ChIP binding sites (Supplemental Table 4A). We then limited this same test to Rfx6 ChIP-Seq binding sites (35.65%) that were shared across 1.19% of all open chromatin ATAC peaks (Supplemental Table 3A-B). Notably, when comparing DNA-motif predictions for Rfx6 against these shared Rfx6 ChIP-Seq binding sites, we observed that 65.10% were true positives, while 8.81% were false predictions (Supplemental Table 4B). However, we also observed a broad distribution in true positive values across these DNA-binding motifs (4.71-65.10%), and also noted a relatively low range of false positives (1.68-8.81%). This is

reflective of a limitation in the ability for DNA-motif's to consistently predict true transcription factor binding.

Determining overlap of differentially enriched chromatin with islet ChIP and histone datasets

Given that may of the DNA-motifs associated with transcription factors do not have available ChIP-Seq datasets derived from mouse pancreatic islets, we sought to understand whether differential chromatin between cell types could be associated to transcription factors and histone markers integral to pancreatic islet cell fate that do have available ChIP data, as opposed to relying only on predictive motifs. As we previously observed strong enrichment of transcription factor binding sites across all open islet chromatin, we wanted to confirm if this overlap is augmented in differentially enriched chromatin associated with pancreatic islet transcription factor binding sites via aggregated islet ChIP data - Pdx1, Nkx6-1, Neurod1, Insm1, Foxa2, Nkx2-2, Rfx6, and MafA - and select, key histone marks — H3K27ac [91], H3K4me3 [91], and H3K4me1 [12]. Our intention in integrating these data was in anticipation that they may help delineate whether enhancer regions are poised (defined by: H3k4me1) [92], or active (defined by: H3k27ac and H3k4me1) [93], and whether promoter regions are active (defined by: H3k4me3) [94]. Indeed, we observed that differentially enriched peaks within our comparisons occurred at much higher rates than random chance across a majority of transcription factor ChIP data associated with islet cell identity (Supplemental Fig. 7A-C) as well as all predictive histone marker regions (Supplemental Fig. 7D-F). This further supported our hypothesis that open chromatin, and now specifically differentially enriched chromatin, would be directly associated with the transcription factors responsible for shaping islet cell-specific gene expression patterns and identity. As differentially enriched chromatin is associated with cell-identity regulatory networks, we inquired to selectively evaluate these regions for enhancers that may be relevant to pancreatic islet cell identity.

Identifying and visualizing putative islet cell-type specific enhancers via epiRomics

These transcription factor and histone ChIP datasets were then fed into an R package named *epiRomics* that we developed to identify putative enhancer regions involved in pancreatic islet cell identity. We defined enhancer regions by the co-localization of H3k27ac and H3k4me1 histone modifications within islet cell chromatin. We stringently narrowed our definition further by requiring these regions to also have transcription factor binding sites - Pdx1, Nkx6-1, Neurod1, Insm1, Foxa2, Nkx2-2, Rfx6, MafA, Isl1, Kat2b, Ldb1, and Gata6 – defined by islet ChIP data that are either ubiquitously or selectively expressed across the three islet cell types (Fig. 2E-F; Supplemental Fig. 2E; Supplemental Fig. 8A-I). This first pass resulted in 28,647 putative enhancer regions (Supplemental Dataset 3). We then filtered this list against chromatin accessible regions from our ATAC-Seq data sets of alpha, beta, and delta cells, resulting in 16,651 putative active enhancers (Fig. 9A). To further increase our confidence in these enhancer calls, we crossed our putative enhancer regions against the curated FANTOM5 curated enhancer database [95]. This resulted in a conservative list of 3,535 putative enhancer regions. Of these 2,347 were inaccessible to at least two out of three islet endocrine cell types (Fig. 9B) (Supplemental Dataset 4).

In both putative enhancer lists, we found that 39.8-43.2% of the enhancer regions we identified were common across all cell types, supporting the theory that related cell types of a common origin would have a sizeable commonality of similar regulatory regions involved in development and maintenance (Fig. 9A-B). Interestingly, between 1.53–10.1% of called enhancers were associated with accessible chromatin unique to each cell type. Enhancer regions selective to beta cells were identified at the highest frequency (~10%), while alpha and delta enhancers made up ~2% of the list.

Upon evaluating whether or not our putative enhancer list would recapitulate two previous mouse pancreatic islet studies delineating enhancers, we confirmed that our approach was able to independently identify an established intronic enhancer on the *Slc30a8* gene, demonstrated to be regulated in part by the

Pdx1 transcription factor (Fig. 9C) [96]. Our approach also supported a previously identified promoter-proximal enhancer region targeting *Pdx1*, with co-occurring binding sites for islet transcription factors Insm1, Neurod1, and Foxa2 (Fig. 9D) [77].

Given that our approach corroborated enhancers identified through complementary methods in previous mouse islet studies, we investigated cell-specific or common putative enhancer candidates by evaluating those with the highest number of transcription factor co-binding sites from our list. One of the top predicted beta cell-unique putative enhancer regions is located on the sixth exon of the *Slc35d2* gene and aligns with eight different ChIP co-localization binding sites (Fig. 10A). An alpha cell-unique putative enhancer located at a distal-intergenic region ~30kb upstream of *Dusp10*, overlapped precisely with six sites of co-binding from various transcription factors (Fig. 10B). A delta cell-unique region at a distal-intergenic enhancer region ~21kb upstream of *Gm20745* aligned closely with no fewer than 12 sites of co-binding from multiple transcription factors. (Fig. 10C). And finally, a common enhancer located ~32kb upstream of *Snap25* – a gene expressed in alpha, beta, and delta cells, associated with a total of 17 co-binding sites of aggregated transcription factors. (Fig. 10D).

We noted further examples of enhancer regions that are inaccessible to beta, but present in both alpha and delta (Supplemental Fig. 9A) cells, or others with chromatin accessibility across all cell types with an adjacent, intronic enhancer region uniquely available to beta cells alone (Supplemental Fig. 9B). In particular, the *Slc2a2* gene shares common open chromatin across alpha, beta, and delta cells. However, beta cells have a gained accessible chromatin region on the first intron identified as a putative enhancer and which overlaps with six co-binding sites. Finally, we noted more putative regions that were enriched in both alpha and beta cells, and present in delta cells as well (Supplemental Fig. 9C-D).

Discussion

The high quality ATAC-Seq data derived from this study is the first dataset of its kind from FACS-purified mouse alpha, beta, and delta cells. Moreover, while bulk-ATAC-Seq data from human alpha and beta cells have previously been reported, our data are the first to report on chromatin accessibility of delta cells and combine these cell's chromatin landscapes to companion transcriptome data. We believe that our data will provide a useful resource that complements our companion transcriptome data that we reported previously using the exact same combination of reporter strains [6]. Leveraging these data allowed us to confirm previous findings in a human ATAC-Seq study evaluating alpha and beta cells, which suggested that alpha cells are poised but repressed from becoming beta cells [5], and present evidence that supports that delta cells might be similarly epigenetically poised to adopt a beta cell like gene expression pattern. We also now harmonized our ATAC-Seq and RNA-Seq data with a wealth of -omics levels data from our colleagues, resulting in a comprehensive multi-layered omics overview that includes histone modifications and transcription factor binding sites. Finally, we made these data accessible through an intuitive interface that we developed to be navigated without any bioinformatics experience.

In evaluating the chromatin landscape of alpha, beta, and delta cells, we noted that over half of accessible chromatin in any of the cell types corresponded to promoter-proximal regions (~25%) and intronic regions (~32%), even though a much smaller fraction of the genome is represented by promoter-proximal sites. This underscores that a substantial portion of regulatory activity occurs directly at genic regions themselves. The enrichment of promoter-proximal and intronic open chromatin we observed in mouse islet cells agrees with previous findings in human studies [5, 38]. The strong presence of intronic peaks supported previously established findings of how enhancers on introns can act as suppressors [33] or drivers of gene expression [97] – in one instance, how *Pdx1* regulates the expression *Slc30a8* through an intronic enhancer [96] – and suggested that these intron regions of accessible chromatin may play a role in cell identity (Fig. 8D). Our findings of a large number of peaks residing at distal-intergenic regions (~35%) agree with previous research identifying and emphasizing the role of distal intergenic regions acting as

enhancers (Fig. 10C-D, Supplemental Fig. 9C-D) in pancreatic islet identity and functional beta cell behavior, and through linkage with T2D GWAS studies that link these regions to beta cell dysregulation [19, 20, 23, 24, 98, 99].

Upon evaluating differences in chromatin accessibility between pairwise comparisons, we discovered that overall, differentially enriched ATAC-Seq peaks in alpha or delta cells were more likely to occur at promoter-proximal regions adjacent to the TSS, whereas peaks enriched in beta cells were often found in distal intergenic or intronic regions, suggesting different mechanisms regulating alpha and delta cell fate specification (Fig. 5B, 5F). When comparing differentially enriched TSS-associated chromatin and respective gene expression, we observed a strong association between chromatin accessibility and gene expression. However, both alpha and delta cells showed a preference in putatively poised genes when either was compared to beta cells. Of note, *MafA* is a key transcription factor enriched in beta cells that shows abundant chromatin accessibility in both alpha and delta cells (Fig. 2E) but is only expressed in beta cells. Another notable example is *Pdx1*, which shows poised TSS enrichment in alpha cells, but is only expressed in beta and delta cells (Fig. 2F).

Moreover, a majority of alpha- and delta- repressed genes showed intermediate expression in the transdifferentiated populations (Fig. 7A-B), further supporting that these are indeed putatively poised. These observations are in line with prior data that suggest that alpha cells are epigenetically poised to become beta cells, but are prevented from assuming beta cell transcriptional programs by repressive regulators at key beta-specific transcription factors [5, 15]. Our observations here also fit reports of adult or juvenile transdifferentiation of alpha-to-beta, or delta-to-beta, respectively [5, 67, 100], although the contribution of these processes to beta cell regeneration is uncertain [101].

After evaluated motif binding on differentially enriched chromatin, we found that Irx2 and Insm1 motifs are enriched at promoter-proximal regions of cell-specific alpha peaks when comparing alpha and beta cells, suggesting that they directly drive gene expression or repression in alpha cells by binding to uniquely accessible chromatin [4, 77]. For Irx2, this indicates that it directly drives gene expression or

repression in alpha cells. For Insm1, which is expressed more uniformly in all three endocrine cell types, the role that it plays at promoter-proximal accessible chromatin is more complex and cannot be as readily inferred. Between alpha and delta cells, the DNA binding motif for Pbx3, a transcription factor implicated in driving *Sst* expression in delta cells, was found preferentially enriched in accessible chromatin associated with promoter-proximal peaks [84], while the DNA motif for *Atf2*, identified as an enriched alpha cell motif in a previous human study, was found to be preferential enriched to chromatin associated with intronic peaks [5]. Between beta and delta cells, Insm1 and Nkx6.1 had motifs enriched at intronic chromatin regions [77, 89], while Fev – recently identified in pancreatic islet development - and Atf3 – linked to enhancer regions in EndoC-bH1 cells - were identified as preferential to chromatin associated with distal-intergenic regions via their DNA binding motifs [11, 90].

Utilizing the R package *epiRomics*, we were able to derive a set of 16,651 putative enhancer peaks. Of these regions, 16.7% of enhancers were shared between beta and alpha cells, and 17.8% were shared between beta and delta cells, as opposed to the 8.18% shared between alpha and delta cells. Of note, our approach identified previously identified intronic enhancers, such as the one located on *Slc30a8 that is* regulated in part via the binding of the transcription actor Pdx1 and a promoter-proximal enhancer region upstream of the Pdx1 that is associated with Insm1, Neurod1, and Foxa2 binding, also identified by our approach (Fig. 9D) [77]. One final example of an enhancer is situated on the first intron of *Slc2a2.* having unique chromatin accessibility to beta cells, coupled with multiple transcription factor binding sites, including Pdx1, MafA, and Nkx6.1, could possibly explain the expression of the gene in beta cells while it is near undetectable between alpha and delta cells (Supplemental Fig.9B). *Slc2a2* plays a necessary role glucose-stimulated insulin secretion [102], with a recent study identifying a downstream enhancer regulating *Slc2a2* requiring the co-occupancy of both MafA and Neurod1, but also noting that complex epigenetic interactions occur beyond the scope of this distal region [103].

One limitation of our approach was that we were constrained to using protein data available to the field. The substantial majority are transcription factors associated with beta cells, with the results reflective

of this limitation. For instance, 10.1% of the enhancer regions called were unique to beta cells, whereas we were only able to identify 1.53-2.47% unique to either delta or alpha cells (Fig. 2E-F; Supplemental Fig. 2E; Supplemental Fig. 8A-I). The over-representation of beta cell-specific enhancer regions is probably explained by the fact that ChIP data for alpha and delta cell-specific enhancers obtained from pure populations of primary alpha and delta cells does not exist. While the majority of the transcription factors here are associated with beta cells, these data are still informative as delta and alpha regions with absence of beta-cell transcription factor binding may be areas regulated through other layers of epigenetics, such as methylation, or via alpha- or delta- specific transcription factors for which no ChIP data is currently available [94]. While further validation of these regions lays beyond the scope of this study, such information would be readily integrated in the future in the multi-omics resource we described here.

In conclusion, we provide a comprehensive snapshot of the characterization of chromatin similarities and differences between mouse alpha, beta, and delta cells. Here, we identify certain TSS genic regions that present as putatively poised in either alpha or delta cells and demonstrate intermediate expression of these genes in beta cells of a non beta-lineage (either alpha or delta transdifferentiated). We also provide a novel approach to identify active enhancers in these cell types through the use of these data alongside data integrated from the field using our package, *epiRomics*, first confirming enhancers identified in previous studies, and then showcasing novel regions with potential for further exploration. Taken together, we have demonstrated that the integration of chromatin accessibility data via ATAC-Seq with other epigenomic data can help further delineate regulatory regions and help answer outstanding questions in the field. Studies and resources such as these are relevant in such that they also function as a supportive resource for integrative research. Given this, we have made these data along with those aggregated through our approach as an interactive resource available on our website.

Conclusion

Here we have established a comprehensive picture of chromatin accessibility between major islet endocrine cell types and present the novel chromatin landscape of delta cells. We identified differential chromatin accessibility at promoter-proximal regions in both alpha cells and delta cells, when compared to beta cells. This finding was in line with a previous study in human islets, and further builds on previous literature in the field suggesting that both alpha and delta cells can transdifferentiate into beta cells. We also identified preferentially binding pattern differences across the annotated genome in transcription factor DNA-motifs across differentially enriched chromatin. Our evaluation of whether chromatin enrichment at the gene body is always correlated with gene expression enrichment also demonstrated that transcriptional regulation plays a role in determining cell fate rather than chromatin dynamics alone. Lastly, we devised and provided a simple approach to utilize and integrate a subset of these epigenomic datasets – ChIP and histone - alongside our ATAC-Seq chromatin data integrated with our previously published transcriptome of the FACS-purified alpha, beta, and delta cells through the development of an R package, *epiRomics*. This allowed for the visualization of integrated epigenomic data, and furthermore applies a novel approach to identify putative enhancer regions, enabling a high-resolution overview of key regions that may be responsible for driving cell fate decisions in pancreatic islet cell types. We have made this an interactive resource publicly available at https://www.huisinglab.com/epiRomics_2021/index.html. We believe our data and the tool we developed to visualize it to be a valuable resource to our field in pursuit of a full understanding of the epigenetic control over islet gene expression.

Acknowledgements

Conceptualization, Methodology, Investigation and Review, A.M.M., M.O.H, T.v.d.M.; Experimental Design, A.M.M., T.v.d.M., M.O.H.; Writing and Editing, and Visualization, A.M.M. and M.O.H.; Formal Analysis, A.M.M.; Supervision, Project Administration, Mentorship, and Funding Acquisition, M.O.H. M.O.H. received funding and consulting fees from Crinetics Inc. The other authors

report no conflicts of interest. M.O.H is the guarantor of this work and takes responsibility for the integrity and accuracy of the data analyzed within this manuscript.

This work was supported by grants from the National Institutes of Health (NIDDK110276) and the Juvenile Diabetes Research Foundation (CDA-2-2013-54 and 2-SRA-2021-1054-M-N) to M.O.H. A.M.M. was supported by the Stephen F. and Bettina A. Sims Immunology Fellowship. This work was also supported by the University of California, Davis Flow Cytometry Shared Resource Laboratory with funding from the NCI (P30 CA093373) and NIH S10 (OD018223) awards.

Figures

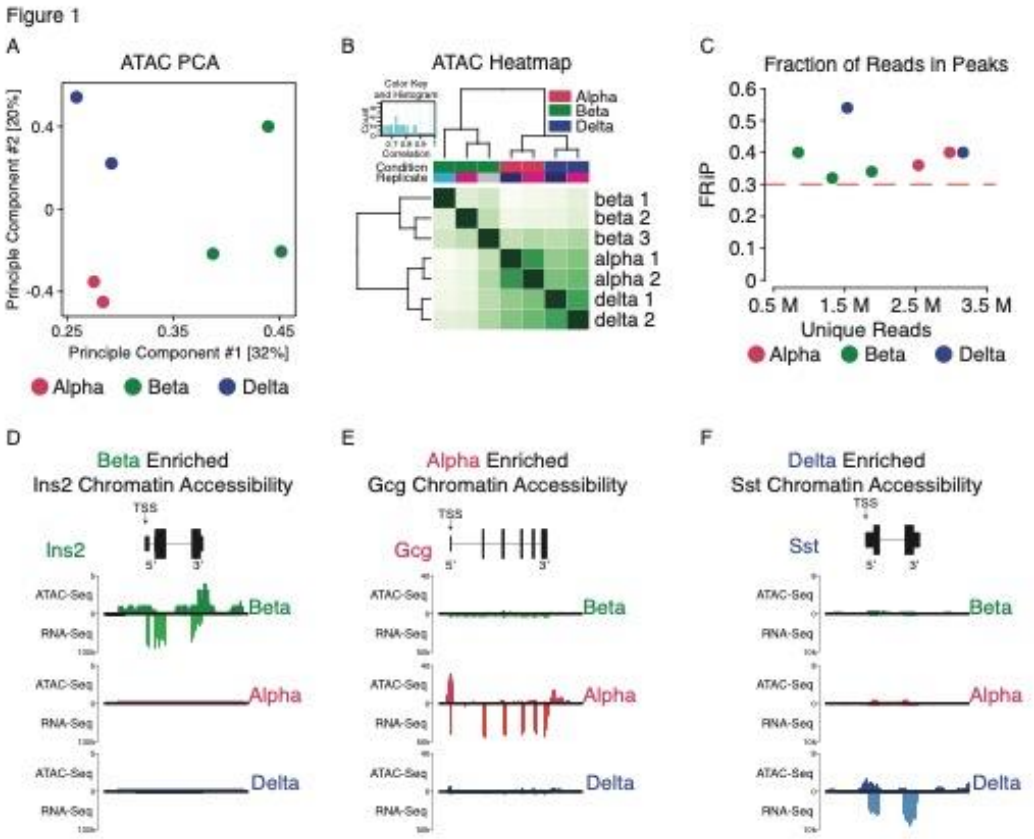


Figure 1
A
ATAC PCA
Principle Component #2 [20%]
0.4
0
-0.4
0.25
0.35
0.45
Principle Component #1 [32%]
Alpha
Beta
Delta
B
ATAC Heatmap
Alpha
Beta
Delta
beta 1
beta 2
beta 3
alpha 1
alpha 2
delta 1
delta 2
C
Fraction of Reads in Peaks
FRiP
0.6
0.5
0.4
0.3
0.2
0.1
0
0.5 M
1.5 M
2.5 M
3.5 M
Unique Reads
Alpha
Beta
Delta
D
Beta Enriched
Ins2 Chromatin Accessibility
TSS
Ins2
ATAC-Seq
RNA-Seq
Beta
Alpha
Delta
E
Alpha Enriched
Gcg Chromatin Accessibility
TSS
Gcg
ATAC-Seq
RNA-Seq
Beta
Alpha
Delta
F
Delta Enriched
Sst Chromatin Accessibility
TSS
Sst
ATAC-Seq
RNA-Seq
Beta
Alpha
Delta

Figure 2

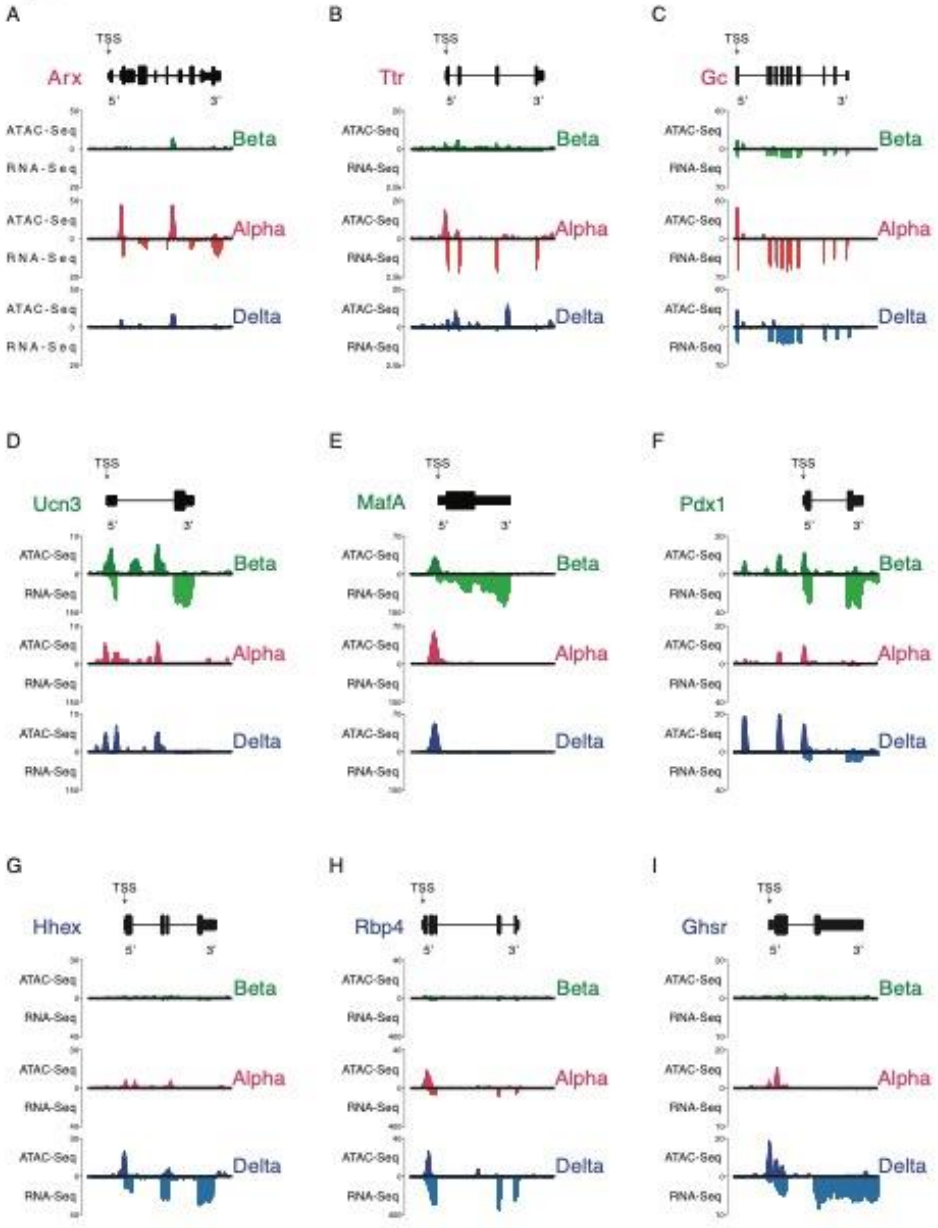

A
Arx
TSS
5'
3'
ATAC-Seq
RNA-Seq
Beta
Alpha
Delta
B
Ttr
C
Gc
D
Ucn3
E
MafA
F
Pdx1
G
Hhex
H
Rbp4
I
Ghsr

Figure 3

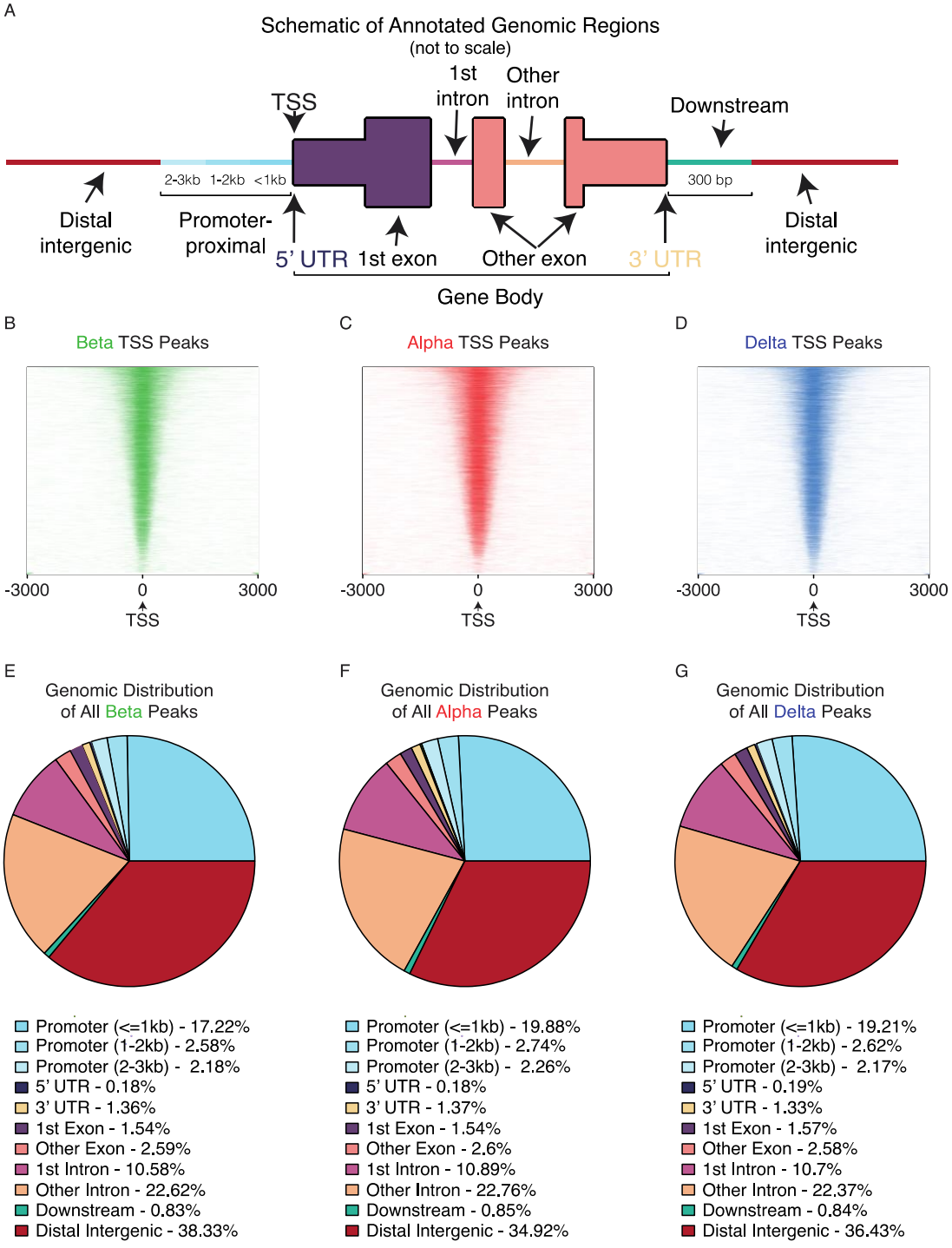

A
Schematic of Annotated Genomic Regions
(not to scale)
TSS
1st intron
Other intron
Downstream
2-3kb
1-2kb
<1kb
300 bp
Distal intergenic
Promoter-proximal
5' UTR
1st exon
Other exon
3' UTR
Distal intergenic
Gene Body
B
Beta TSS Peaks
C
Alpha TSS Peaks
D
Delta TSS Peaks
-3000
0
3000
TSS
E
Genomic Distribution of All Beta Peaks
F
Genomic Distribution of All Alpha Peaks
G
Genomic Distribution of All Delta Peaks
Promoter (<=1kb) - 17.22%
Promoter (1-2kb) - 2.58%
Promoter (2-3kb) - 2.18%
5' UTR - 0.18%
3' UTR - 1.36%
1st Exon - 1.54%
Other Exon - 2.59%
1st Intron - 10.58%
Other Intron - 22.62%
Downstream - 0.83%
Distal Intergenic - 38.33%
Promoter (<=1kb) - 19.88%
Promoter (1-2kb) - 2.74%
Promoter (2-3kb) - 2.26%
5' UTR - 0.18%
3' UTR - 1.37%
1st Exon - 1.54%
Other Exon - 2.6%
1st Intron - 10.89%
Other Intron - 22.76%
Downstream - 0.85%
Distal Intergenic - 34.92%
Promoter (<=1kb) - 19.21%
Promoter (1-2kb) - 2.62%
Promoter (2-3kb) - 2.17%
5' UTR - 0.19%
3' UTR - 1.33%
1st Exon - 1.57%
Other Exon - 2.58%
1st Intron - 10.7%
Other Intron - 22.37%
Downstream - 0.84%
Distal Intergenic - 36.43%

Figure 4

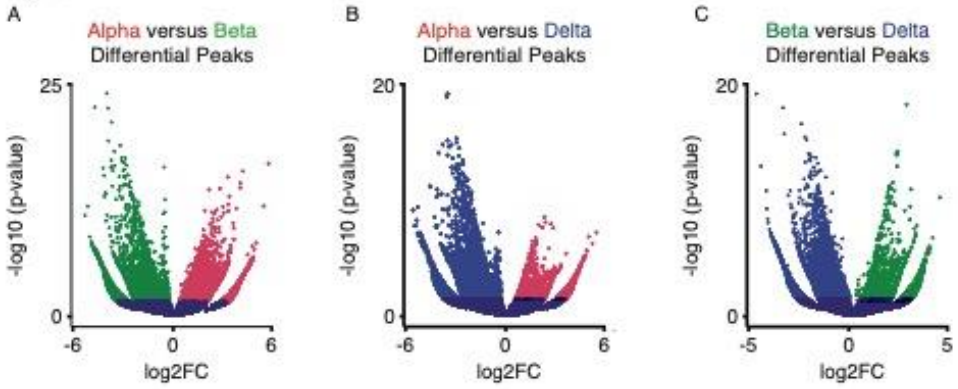

A
Alpha versus Beta
Differential Peaks
25
0
-log10 (p-value)
-6
0
6
log2FC
B
Alpha versus Delta
Differential Peaks
20
0
-log10 (p-value)
-6
0
6
log2FC
C
Beta versus Delta
Differential Peaks
20
0
-log10 (p-value)
-5
0
5
log2FC

Figure 5

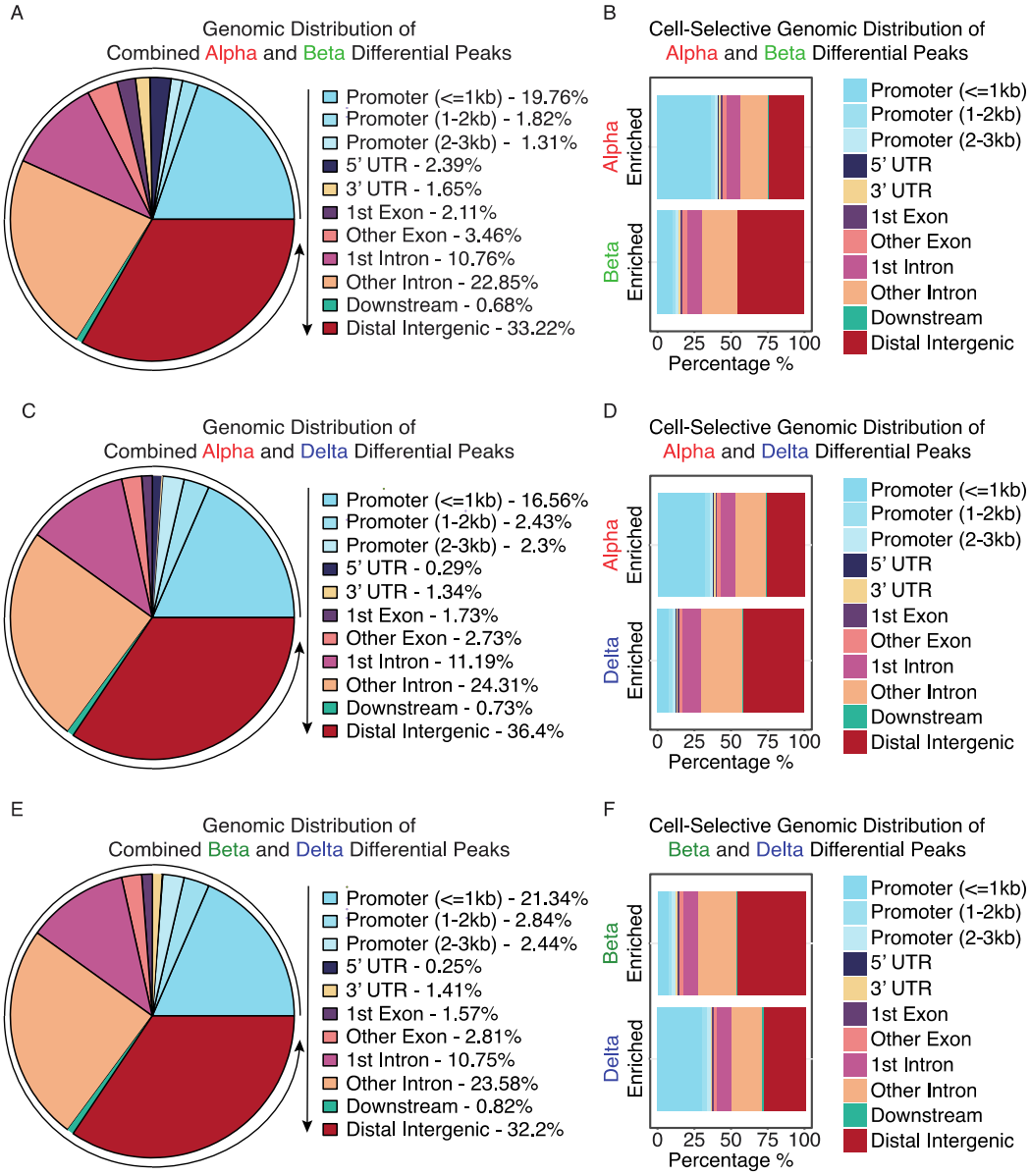

A
Genomic Distribution of
Combined Alpha and Beta Differential Peaks
Promoter (<=1kb) - 19.76%
Promoter (1-2kb) - 1.82%
Promoter (2-3kb) - 1.31%
5' UTR - 2.39%
3' UTR - 1.65%
1st Exon - 2.11%
Other Exon - 3.46%
1st Intron - 10.76%
Other Intron - 22.85%
Downstream - 0.68%
Distal Intergenic - 33.22%
B
Cell-Selective Genomic Distribution of
Alpha and Beta Differential Peaks
Alpha Enriched
Beta Enriched
0 25 50 75 100
Percentage %
Promoter (<=1kb)
Promoter (1-2kb)
Promoter (2-3kb)
5' UTR
3' UTR
1st Exon
Other Exon
1st Intron
Other Intron
Downstream
Distal Intergenic
C
Genomic Distribution of
Combined Alpha and Delta Differential Peaks
Promoter (<=1kb) - 16.56%
Promoter (1-2kb) - 2.43%
Promoter (2-3kb) - 2.3%
5' UTR - 0.29%
3' UTR - 1.34%
1st Exon - 1.73%
Other Exon - 2.73%
1st Intron - 11.19%
Other Intron - 24.31%
Downstream - 0.73%
Distal Intergenic - 36.4%
D
Cell-Selective Genomic Distribution of
Alpha and Delta Differential Peaks
Alpha Enriched
Delta Enriched
0 25 50 75 100
Percentage %
Promoter (<=1kb)
Promoter (1-2kb)
Promoter (2-3kb)
5' UTR
3' UTR
1st Exon
Other Exon
1st Intron
Other Intron
Downstream
Distal Intergenic
E
Genomic Distribution of
Combined Beta and Delta Differential Peaks
Promoter (<=1kb) - 21.34%
Promoter (1-2kb) - 2.84%
Promoter (2-3kb) - 2.44%
5' UTR - 0.25%
3' UTR - 1.41%
1st Exon - 1.57%
Other Exon - 2.81%
1st Intron - 10.75%
Other Intron - 23.58%
Downstream - 0.82%
Distal Intergenic - 32.2%
F
Cell-Selective Genomic Distribution of
Beta and Delta Differential Peaks
Beta Enriched
Delta Enriched
0 25 50 75 100
Percentage %
Promoter (<=1kb)
Promoter (1-2kb)
Promoter (2-3kb)
5' UTR
3' UTR
1st Exon
Other Exon
1st Intron
Other Intron
Downstream
Distal Intergenic

Figure 6

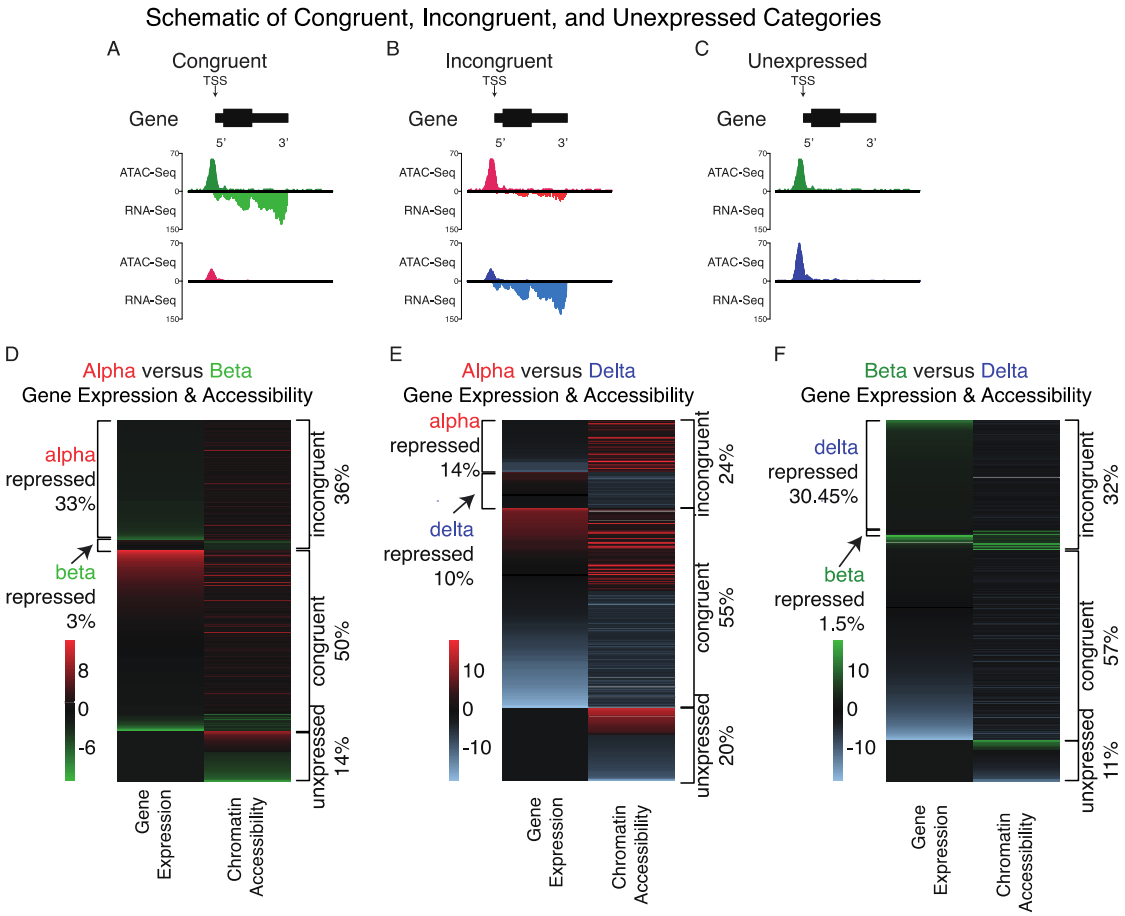

Schematic of Congruent, Incongruent, and Unexpressed Categories
A
Congruent
B
Incongruent
C
Unexpressed
TSS
Gene
5'
3'
ATAC-Seq
RNA-Seq
D
Alpha versus Beta
Gene Expression & Accessibility
alpha repressed 33%
beta repressed 3%
incongruent 36%
congruent 50%
unxpressed 14%
E
Alpha versus Delta
Gene Expression & Accessibility
alpha repressed 14%
delta repressed 10%
incongruent 24%
congruent 55%
unxpressed 20%
F
Beta versus Delta
Gene Expression & Accessibility
delta repressed 30.45%
beta repressed 1.5%
incongruent 32%
congruent 57%
unxpressed 11%
Gene Expression
Chromatin Accessibility

Figure 7

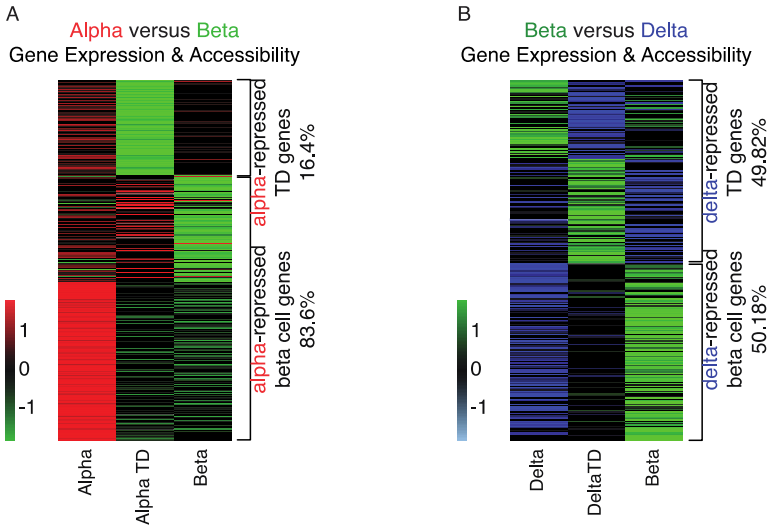

A
Alpha versus Beta
Gene Expression & Accessibility
alpha-repressed TD genes 16.4%
alpha-repressed beta cell genes 83.6%
1
0
-1
Alpha
Alpha TD
Beta
B
Beta versus Delta
Gene Expression & Accessibility
delta-repressed TD genes 49.82%
delta-repressed beta cell genes 50.18%
1
0
-1
Delta
DeltaTD
Beta

Figure 8

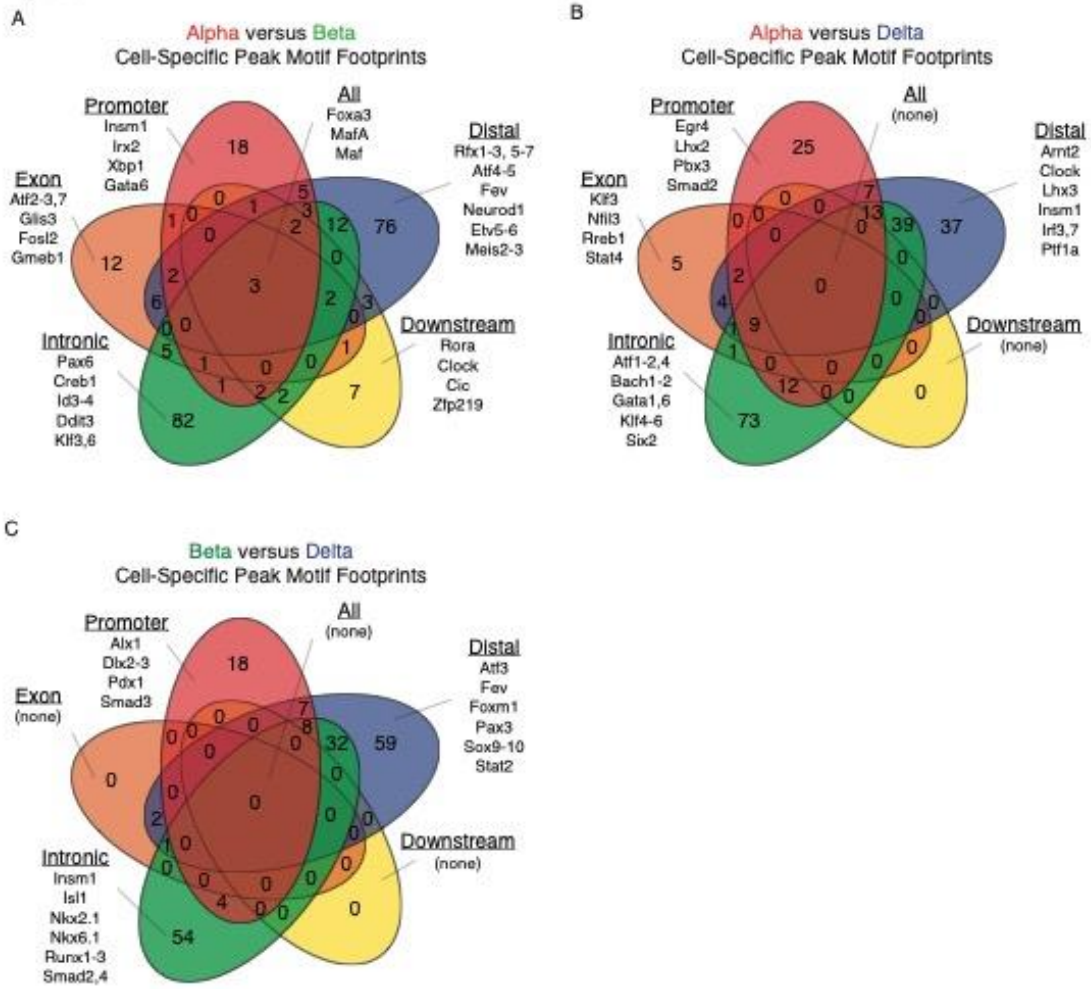

A
Alpha versus Beta
Cell-Specific Peak Motif Footprints
Promoter
Insm1
Irx2
Xbp1
Gata6
All
Foxa3
MafA
Maf
Distal
Rfx1-3, 5-7
Atf4-5
Fev
Neurod1
Etv5-6
Meis2-3
Exon
Atf2-3,7
Glis3
Fosl2
Gmeb1
Intronic
Pax6
Creb1
Id3-4
Ddit3
Klf3,6
Downstream
Rora
Clock
Cic
Zfp219
B
Alpha versus Delta
Cell-Specific Peak Motif Footprints
Promoter
Egr4
Lhx2
Pbx3
Smad2
All
(none)
Distal
Arnt2
Clock
Lhx3
Insm1
Irf3,7
Ptf1a
Exon
Klf3
Nfil3
Rreb1
Stat4
Intronic
Atf1-2,4
Bach1-2
Gata1,6
Klf4-6
Six2
Downstream
(none)
C
Beta versus Delta
Cell-Specific Peak Motif Footprints
Promoter
Alx1
Dlx2-3
Pdx1
Smad3
All
(none)
Distal
Atf3
Fev
Foxm1
Pax3
Sox9-10
Stat2
Exon
(none)
Intronic
Insm1
Isl1
Nkx2.1
Nkx6.1
Runx1-3
Smad2,4
Downstream
(none)

Figure 9

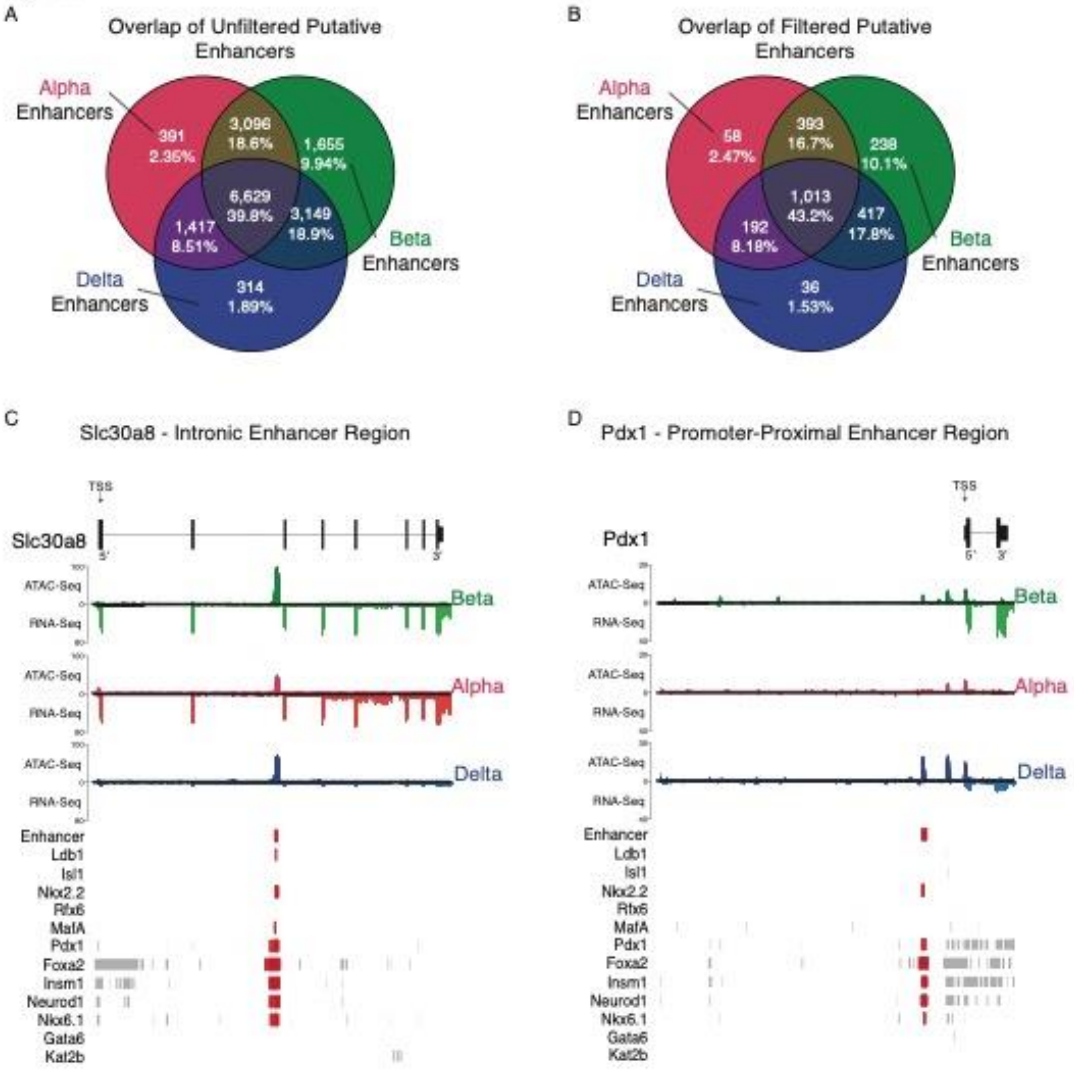
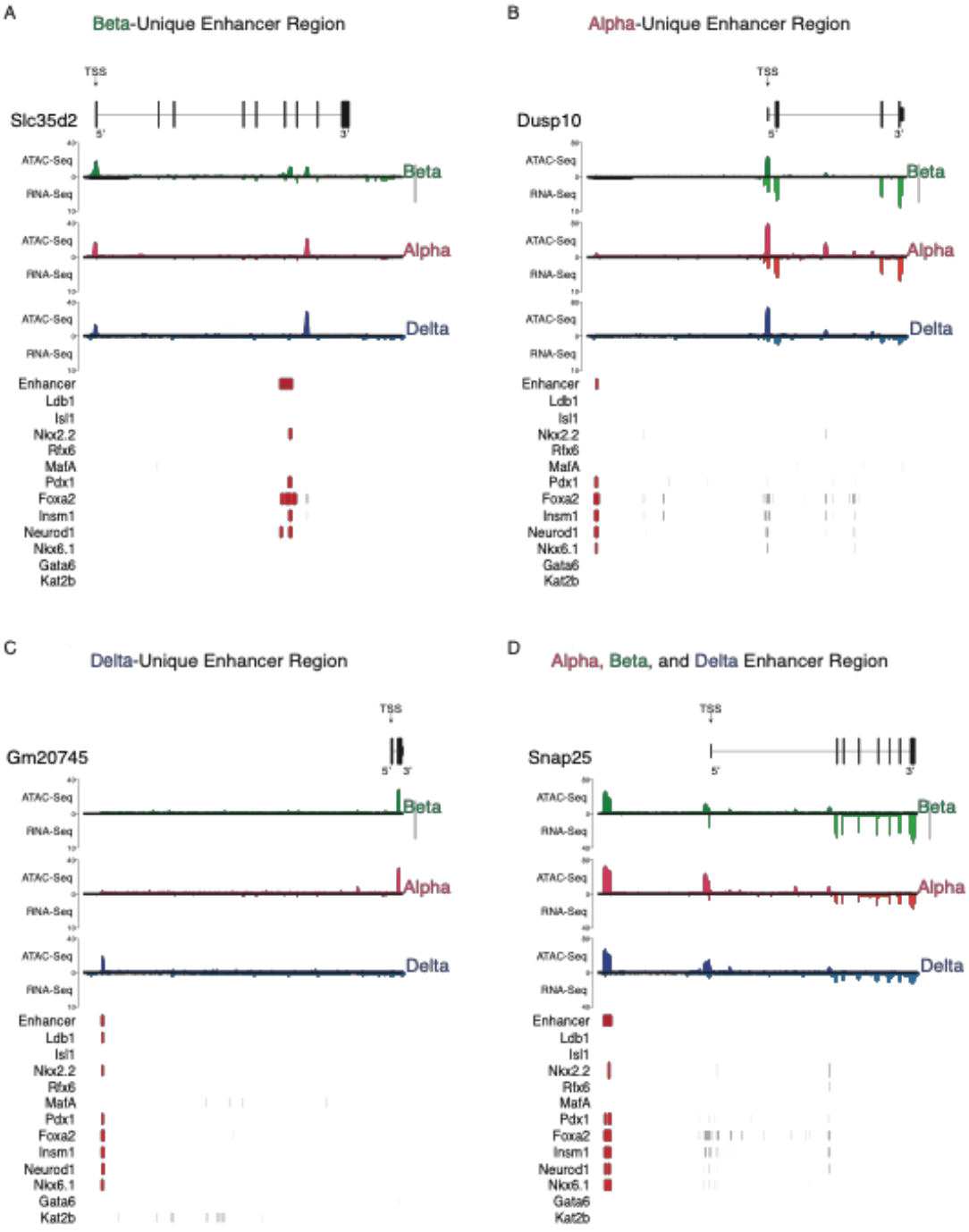

A
Overlap of Unfiltered Putative Enhancers
Alpha Enhancers
Beta Enhancers
Delta Enhancers
391
2.35%
3,096
18.6%
1,655
9.94%
6,629
39.8%
1,417
8.51%
3,149
18.9%
314
1.89%
B
Overlap of Filtered Putative Enhancers
Alpha Enhancers
Beta Enhancers
Delta Enhancers
58
2.47%
393
16.7%
238
10.1%
1,013
43.2%
192
8.18%
417
17.8%
36
1.53%
C
Slc30a8 - Intronic Enhancer Region
TSS
Slc30a8
ATAC-Seq
RNA-Seq
Beta
Alpha
Delta
Enhancer
Ldb1
Isl1
Nkx2.2
Rfx6
MafA
Pdx1
Foxa2
Insm1
Neurod1
Nkx6.1
Gata6
Kat2b
D
Pdx1 - Promoter-Proximal Enhancer Region
TSS
Pdx1
ATAC-Seq
RNA-Seq
Beta
Alpha
Delta
Enhancer
Ldb1
Isl1
Nkx2.2
Rfx6
MafA
Pdx1
Foxa2
Insm1
Neurod1
Nkx6.1
Gata6
Kat2b

Figure 10


A
Beta-Unique Enhancer Region
Slc35d2
TSS
5'
3'
ATAC-Seq
RNA-Seq
Beta
Alpha
Delta
Enhancer
Ldb1
Isl1
Nkx2.2
Rfx6
MafA
Pdx1
Foxa2
Insm1
Neurod1
Nkx6.1
Gata6
Kat2b
B
Alpha-Unique Enhancer Region
Dusp10
C
Delta-Unique Enhancer Region
Gm20745
D
Alpha, Beta, and Delta Enhancer Region
Snap25

Supplemental Table 1

A

Sample	# Cells	Sex	Library Prep	Library Type	Library Size (bp)	Total Reads	Total Reads Mapped	Unique Reads Mapped	FRiP
alpha 1	20000	F	NexteraDNA	single-end	444	28.1 M	18.5 M	2.5 M	0.37
alpha 2	20000	M	NexteraDNA	single-end	475	27.6 M	19.8 M	3.0 M	0.41
beta 1	20000	M	NexteraDNA	single-end	469	31.9 M	21.9 M	0.9 M	0.42
beta 2	20000	F	NexteraDNA	single-end	~450	36.1 M	24.8 M	1.3 M	0.42
beta 3	20000	M	NexteraDNA	single-end	~450	40.2 M	27.6 M	1.9 M	0.35
beta 4	20000	F	NexteraDNA	single-end	457	32.4 M	23.2 M	3.4 M	0.11
delta 1	20000	F	NexteraDNA	single-end	~450	42 M	29.9 M	3.2 M	0.42
delta 2	13500	M	NexteraDNA	single-end	~450	35.4 M	24.1 M	1.5 M	0.55

Supplemental Table 2

A

ChIP (Transcription Factor) Datasets	
Marker	**Accession Info**
MafA	GSE30298
Nkx2.2	GSE79785
Rfx6	GSE62844
Neurod1	GSE54046
Foxa2	GSE54046
Isl1	GSE84759
Kat2b	GSE78860
Ldb1	GSE84759
Nkx6.1	GSE40975
Pdx1	E-MTAB-1143
Gata6	GSE57090
Insm1	GSE54046

B

ChIP (Histone) Datasets	
Marker	**Accession Info**
H3k27ac	GSE110648
H3k4me3	GSE110648
H3k4me1	GSE68618
H2ak119ub	GSE110648
H3k27me1	GSE110648
H3k36me3	GSE110648
H3K9me3	GSE110648
H3k9ac	GSE87530
H3k27me3	GSE110648

Supplemental Table 3

A

Percent of ATAC Peaks Overlapping with Transcription Factor Binding Sites	
Transcription Factor	**Percent Overlap**
Foxa2	29.07%
Insm1	28.40%
Neurod1	20.09%
Pdx1	19.49%
Nkx6.1	15.44%
Nkx2.2	4.56%
MafA	1.56%
Rfx6	1.19%
Gata6	1.02%
Ldb1	0.41%
Kat2b	0.35%
Isl1	0.31%

B

Percent of Transcription Factor Binding Sites Overlapping with ATAC Peaks	
Transcription Factor	**Percent Overlap**
Nkx2.2	63.79%
Neurod1	55.07%
Insm1	51.49%
Pdx1	50.85%
Nkx6.1	43.56%
Gata6	40.20%
Rfx6	35.65%
Foxa2	34.86%
Isl1	9.95%
Ldb1	5.46%
MafA	3.27%
Kat2b	0.35%

Supplemental Table 4

A

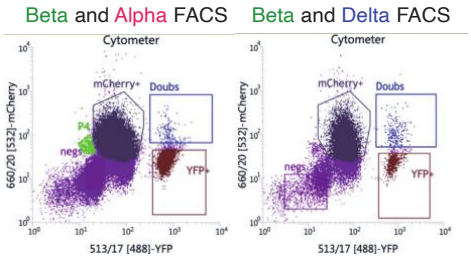
Transcription Factor Motif Calls against ChIP Peaks Validation				
Transcription Factor	False Negative	True Positive	False Positive	True Negative
Rfx6	43.00%	57%	8.34%	91.66%
Gata6	77.58%	22.42%	1.78%	98.22%
Foxa2	78.17%	21.83%	6.38%	93.62%
Nkx6.1	82.05%	17.95%	6.89%	93.11%
Nkx2.2	89.48%	10.52%	2.98%	97.02%
Insm1	93.39%	6.61%	3.63%	96.37%
Pdx1	95.51%	4.49%	1.26%	98.74%
Isl1	98.33%	1.67%	0.51%	99.49%
MafA	99.41%	0.59%	1.19%	98.81%

B

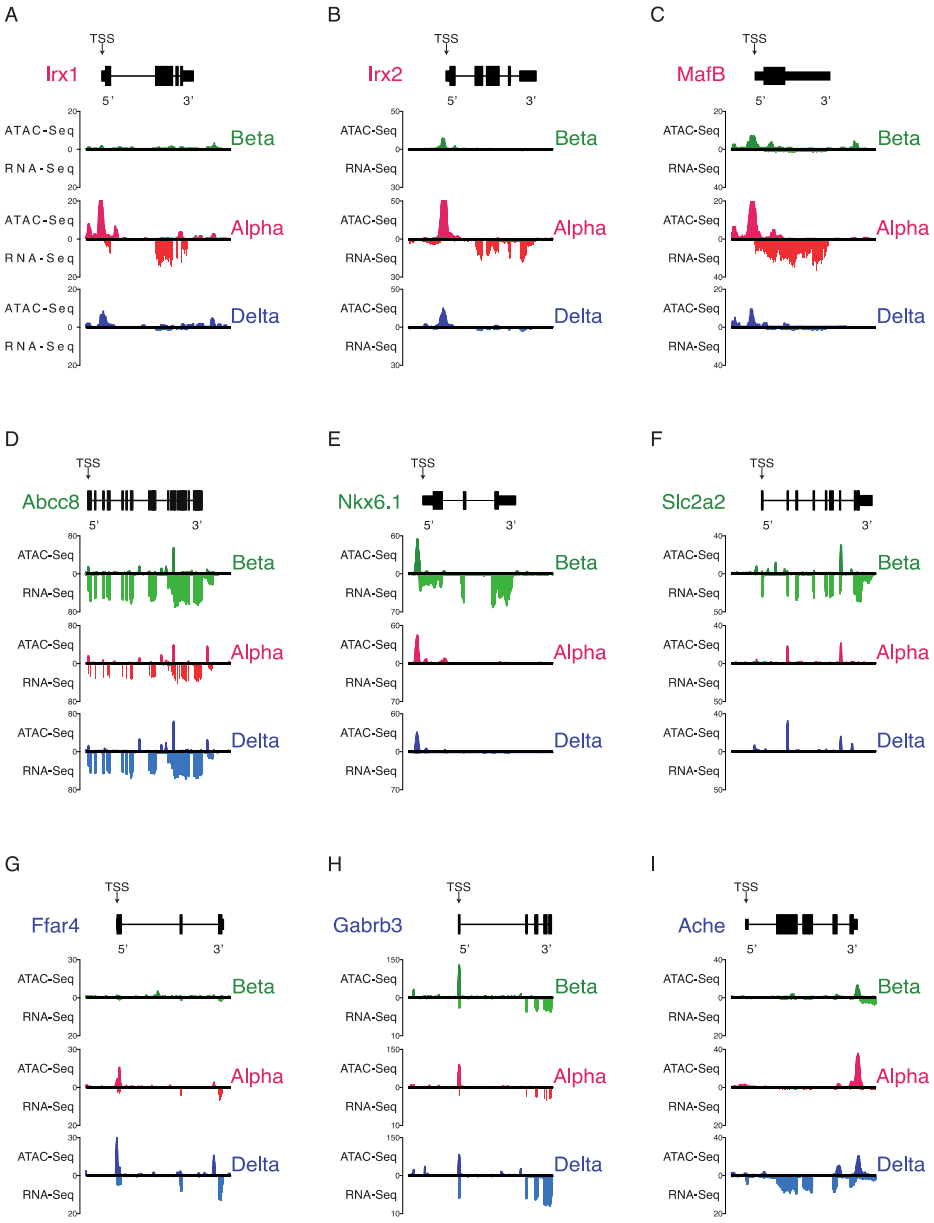
Transcription Factor Motif Calls against ChIP Peaks (Open ATAC only) Validation				
Transcription Factor	False Negative	True Positive	False Positive	True Negative
Rfx6	34.90%	65.10%	8.81%	91.19%
Foxa2	80.11%	19.89%	8.92%	91.08%
Gata6	83.80%	16.20%	1.98%	98.02%
Nkx6.1	88.61%	11.39%	8.74%	91.26%
Nkx2.2	90.28%	9.72%	3.69%	96.31%
MafA	91.29%	8.71%	1.73%	98.27%
Insm1	91.74%	8.26%	3.97%	96.03%
Isl1	94.12%	5.88%	0.74%	99.26%
Pdx1	95.30%	4.71%	1.68%	98.32%

Supplemental Figure 1

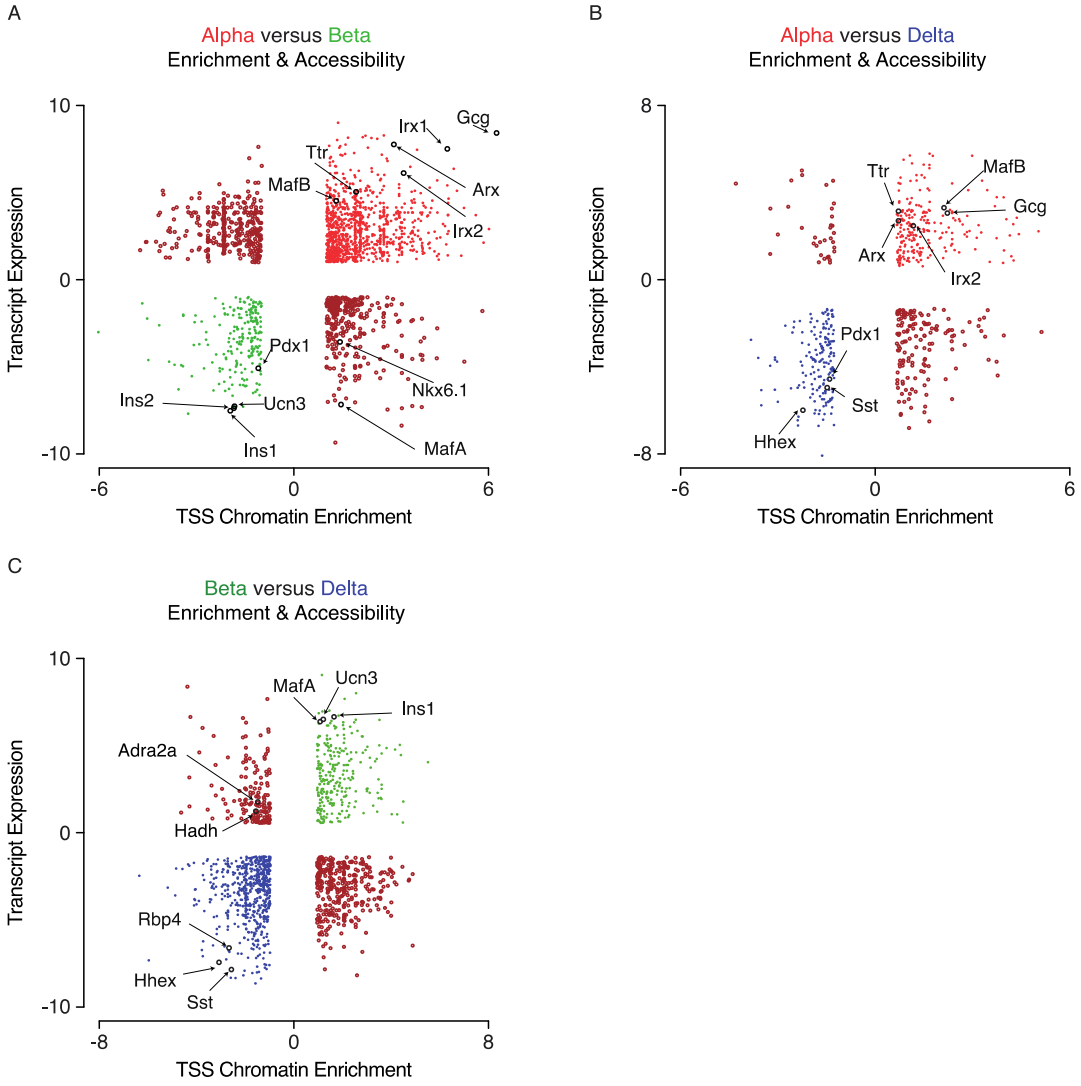
A



Supplemental Figure 2

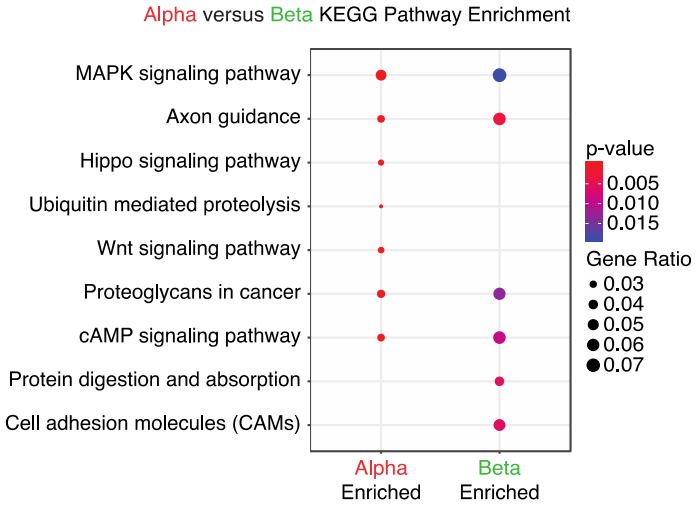


Supplemental Figure 3

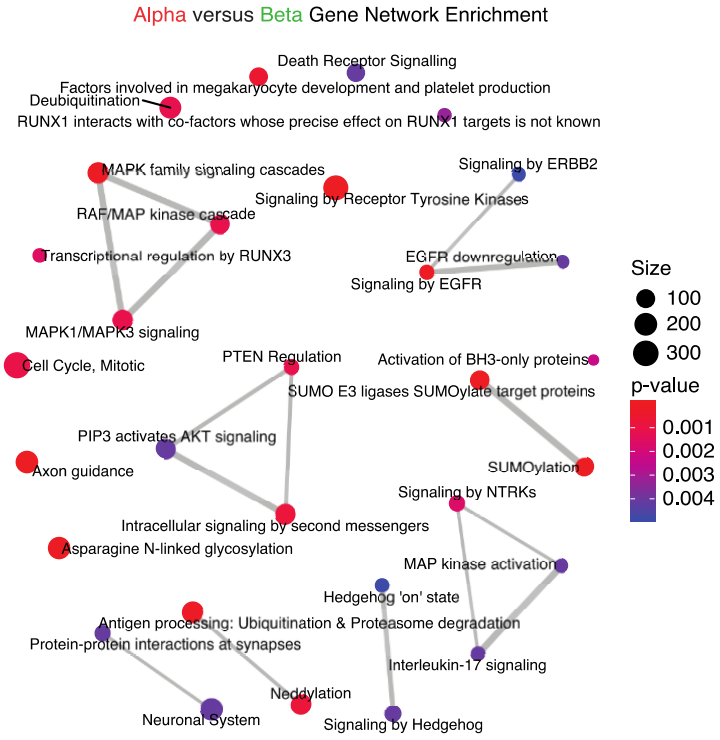


Supplemental Figure 4

A

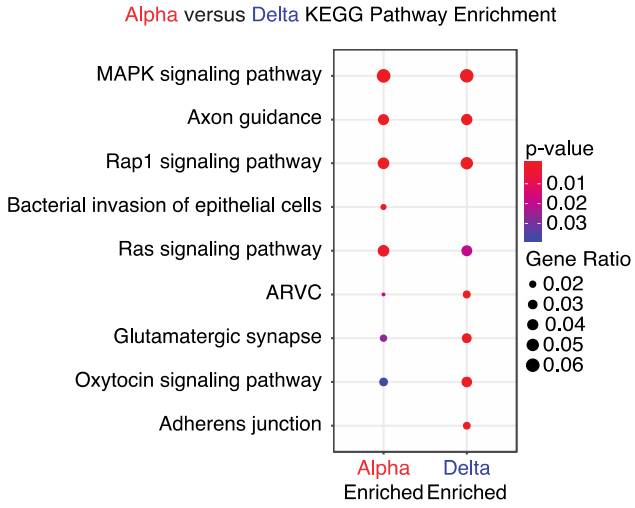


B

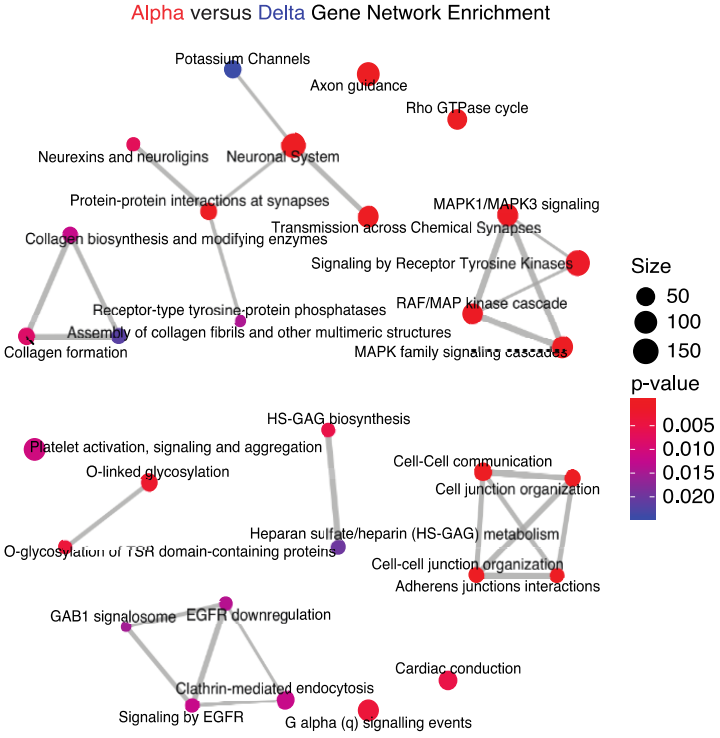


Supplemental Figure 5

A

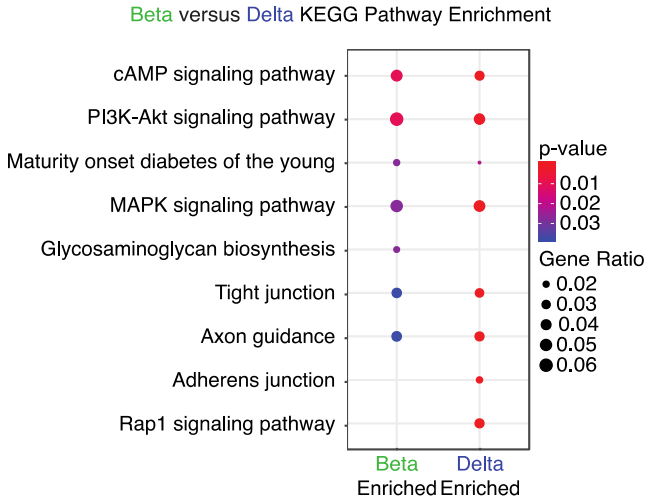


B

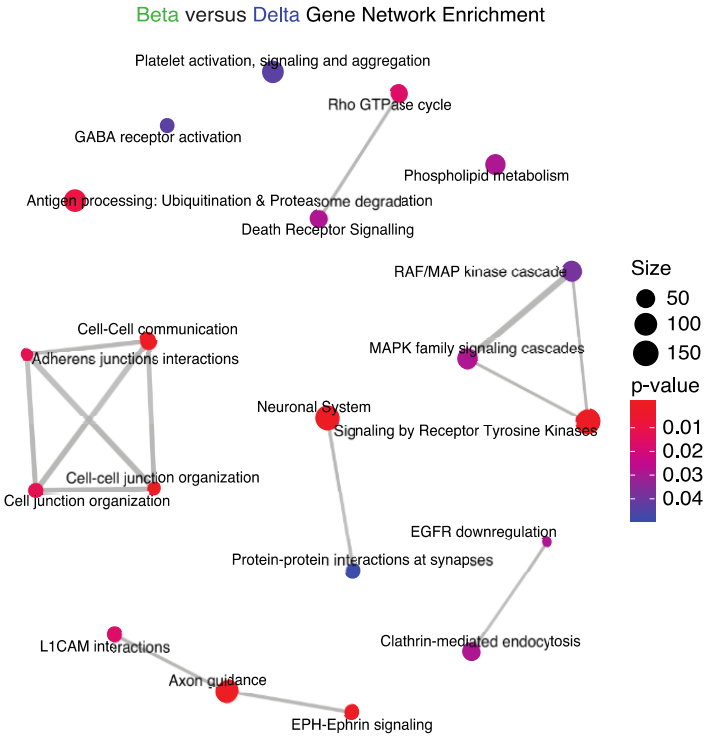


Supplemental Figure 6

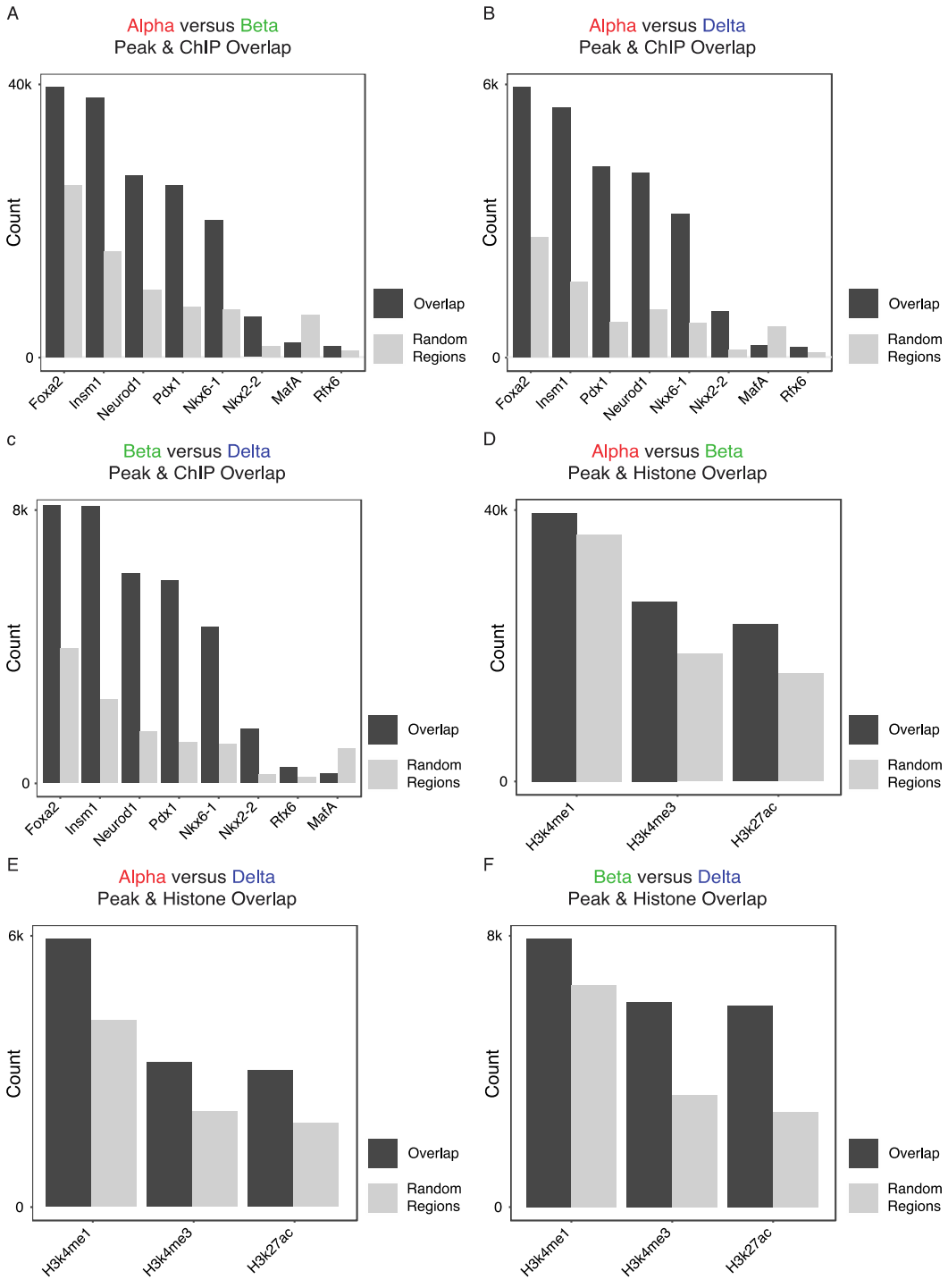
A



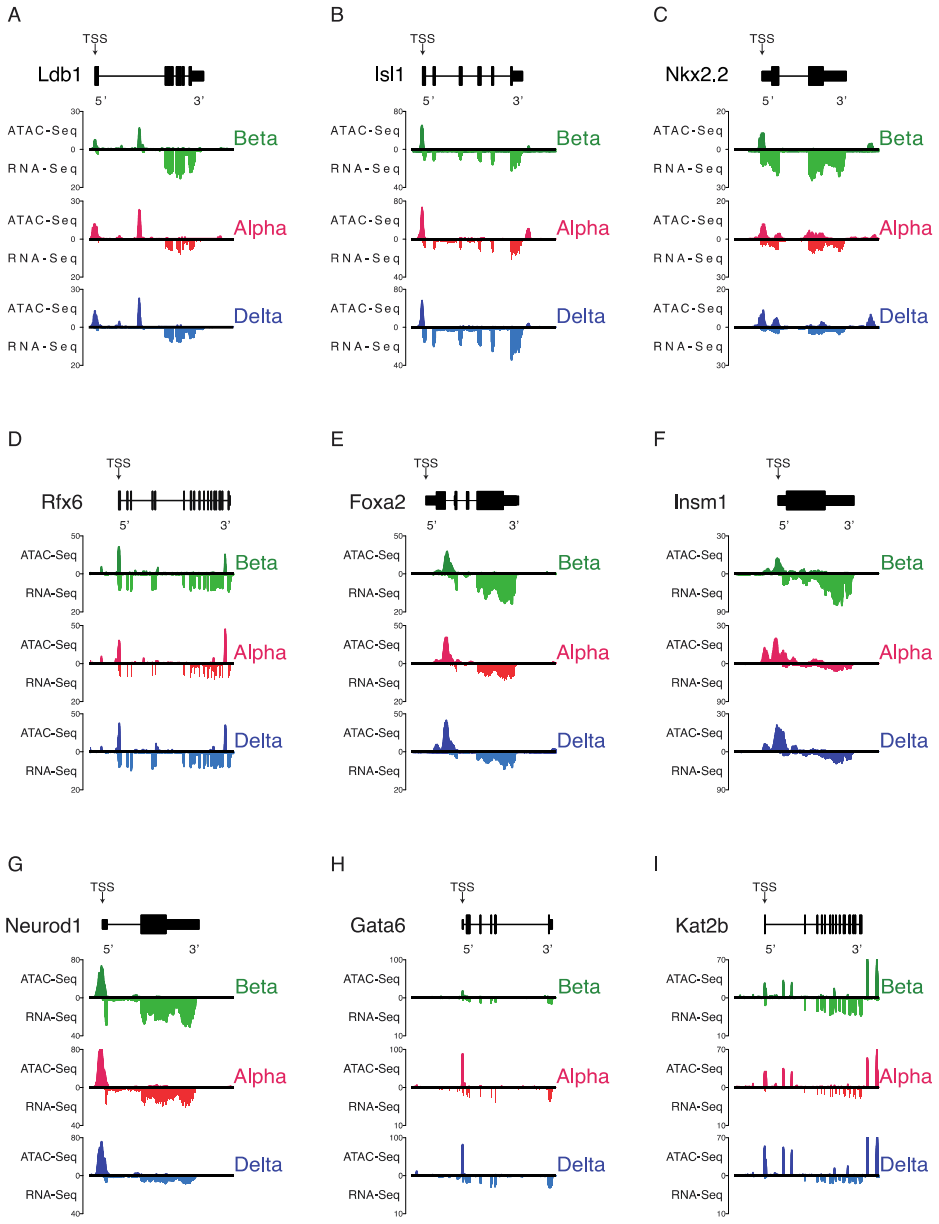
B



Supplemental Figure 7



Supplemental Figure 8



Supplemental Figure 9

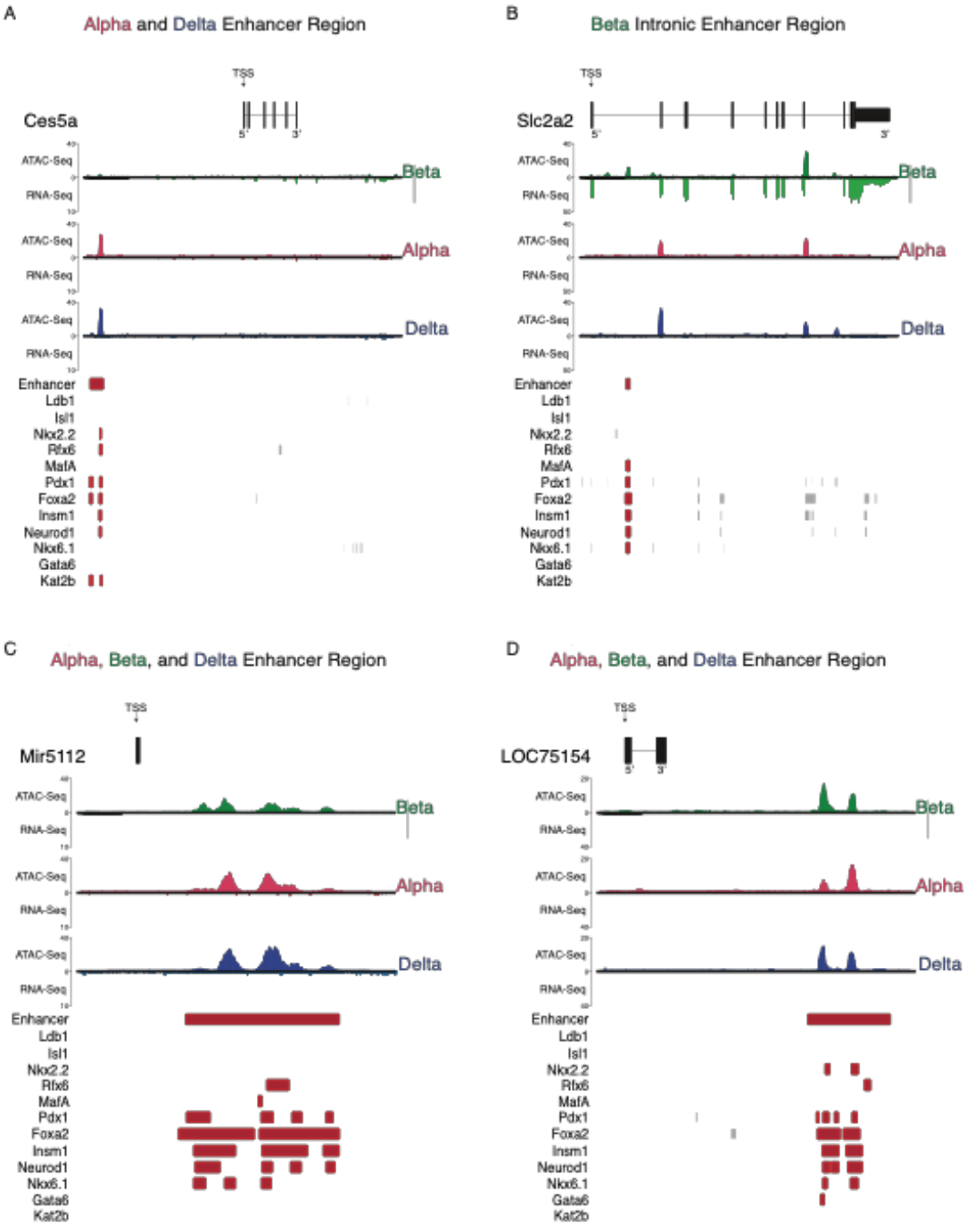


Figure Legends

Main Figures

Figure 1 – Validating alpha, beta, and delta chromatin accessibility ATAC Seq. A: Dimensional reduction through principal component analysis across seven samples from all three cell types (See Supplemental Table 1 for details). All three cell type's replicates clustered closer together and separate from other cell types. B: Heatmap further confirming quality of replicates and similarity between replicates within each cell type. C: Fraction of Reads in Peaks (FRiP) score evaluation across samples, confirming high library complexity irrespective of depth of sequencing. D-F: Confirming chromatin accessibility at the TSS (arrows) against bulk RNA-Seq expression in key islet cell type-specific marker gene regions - *Ins2, Gcg,* and *Sst* - in beta, alpha, and delta cells, respectively.

Figure 2 - Validating chromatin accessibility ATAC Seq alongside companion RNA-Seq expression in alpha, beta, and delta cells against hallmark genes governing its respective cell's identity. All genes are oriented for 5' to 3' end. A-C: Chromatin accessibility and transcript expression across alpha cell hallmark genes *Arx, Ttr,* and *Gc*. D-F: Chromatin accessibility and transcript expression across beta cell hallmark genes *Ucn3, Esr1,* and *Pdx1*. G-I: Chromatin accessibility and transcript expression across delta cell hallmark genes *Hhex, Rbp4,* and *Ghsr*.

Figure 3 – Evaluating chromatin accessibility ATAC Seq similarities and differences across all three cell types. A: Schematic of annotated genomic regions – promoter proximal, intronic, exonic, distal-intergenic, or downstream. B-D: TSS peak (defined as 3kb up or downstream each respective gene) chromatin accessibility density across beta, alpha, and delta cells. E-G: Distribution of chromatin peaks within each cell type across the annotated genome.

Figure 4 – Comparing chromatin accessibility through differential enrichment analysis across alpha, beta, and delta cells. A: Differential chromatin accessibility peaks between alpha and beta ATAC Seq data. A total of 18,409 peaks were considered differentially enriched at p-value <= 0.05 (Supplemental Dataset 1). B: Differential chromatin accessibility peaks between alpha and delta ATAC Seq data. A total of 12,722 peaks were considered differentially enriched at p-value <= 0.05 (Supplemental Dataset 1). C: Differential chromatin accessibility peaks between beta and delta ATAC Seq data. A total of 16,913 were considered differentially enriched at p-value <= 0.05 (Supplemental Dataset 1).

Figure 5 – Regional differences and characteristics of differentially enriched peaks between alpha, beta, and delta cells. A: Distribution of regional preference across the annotated genome of differentially enriched peaks between alpha and beta cells. B: Regional preference breakdown of differentially enriched

peaks between alpha and beta cells, indicating prevalence of enrichment for each cell type and genomic annotation. Differentially enriched chromatin favored promoter-proximal peaks in alpha cells, and distal-intergenic regions in beta cells. C: Distribution of regional preference across the annotated genome of differentially enriched peaks between alpha and delta cells. D: Regional preference breakdown of differentially enriched peaks between alpha and delta cells, indicating prevalence of enrichment for each cell type and genomic annotation. Differentially enriched chromatin favored promoter-proximal peaks in alpha cells, and distal-intergenic regions in delta cells. E: Distribution of regional preference across the annotated genome of differentially enriched peaks between beta and delta cells. F: Regional preference breakdown of differentially enriched peaks between beta and delta cells, indicating prevalence of enrichment for each cell type and genomic annotation. Differentially enriched chromatin favored promoter-proximal peaks in delta cells, and distal-intergenic regions in beta cells.

Figure 6 – Differentially enriched chromatin at TSS genic regions and their respective gene's expression between alpha, beta, and delta cells. A-C: Schematic of 'congruent', 'incongruent', and 'unexpressed' categories used to determine the association of enriched chromatin at TSS genic regions and respective gene expression. D: Differentially enriched chromatin at TSS genic regions and their respective gene's expression between alpha and beta cells. The majority (50%) of genes with enriched chromatin at promoter-proximal regions around their TSS had correlated gene expression (congruent). Another 36% of chromatin enriched TSS regions showed repressed gene expression for each cell type (alpha repressed (33%) or beta repressed (3%)), and finally, 14% were unexpressed. E: Differentially enriched chromatin at TSS genic regions and their respective gene's expression between alpha and delta cells. The majority (55%) of genes with enriched chromatin had correlated gene expression (congruent). Another 24% of chromatin enriched TSS genic regions showed repressed gene expression for each cell type (alpha repressed (14%) or delta repressed (10%)), and finally, 20% showed no expression. F: Differentially enriched chromatin at TSS genic regions and their respective gene's expression between beta and delta cells. The majority (57%) of genes with enriched chromatin had correlated gene expression (congruent). Another 32% of chromatin enriched TSS regions showed repressed gene expression for each cell type (beta repressed (1.5%) or delta repressed (30.45%)), and finally, 11% showed no difference.

Figure 7 – Gene expression of poised genes enriched in beta cells with a non-beta cell lineage. A: Evaluating alpha repressed genes (Fig. 6A) across alpha, alpha transdifferentiated, and beta cell transcriptomes. The great majority (83.6%) of genes repressed in alpha cells showed intermediate expression in alpha transdifferentiated cells, and highest expression in beta cells, further validating that alpha cells are poised to become beta cells, with a subset (16.4%) of those genes required for the transition. B: Evaluating delta repressed genes (Fig. 6C) across delta, delta transdifferentiated, and beta cell

transcriptomes. Around half (50.18%) of genes repressed in delta cells showed intermediate expression in delta transdifferentiated cells, and highest expression in beta cells, validating that delta cells – to a lesser extent than alpha – are also poised to become beta cells, with the remainder of genes (49.82%) required for the transition.

Figure 8 – Evaluating expressed, cell-specific transcription factor footprints on differentially enriched peaks across cell types. A: Evaluating cell-specific transcription factor footprints on differentially enriched peaks for alpha and beta cells, suggesting transcription factor preference for these peaks across the functionally annotated genome. Notably, three known transcription factors were predicted to overlap all defined regions of the genome, whereas others showed preference for binding at either promoter, exon, intron, or distal regions, suggesting different mechanisms of regulation. B: Evaluating cell-specific transcription factor footprints on differentially enriched peaks for alpha and delta cells, suggesting transcription factor preference for these peaks across the functionally annotated genome. No known transcription factor was predicted to bind to all defined regions of the genome, with the majority binding to either intronic, distal, or promoter areas. C: Evaluating cell-specific transcription factor footprints on differentially enriched peaks for beta and delta cells, suggesting transcription factor preference for these peaks across the functionally annotated genome. No known transcription factor was found predicted to bind to all defined regions of the genome, with the great majority showing a preference for distal, intronic, or promoter regions.

Figure 9 – Putative enhancer detection overlap between the three cell types. A: First-pass overlap of unfiltered putative enhancers called with our novel package, epiRomics. Open chromatin regions in at least one cell type were crossed against two informative histone marks - H3k27ac and H3k4me1 – and transcription factor binding data to call putative enhancer regions. A total of 28,647 regions were identified (Supplemental Dataset 3). 39.8% of putative enhancer calls had chromatin accessible to all three cell types, suggestive of pancreatic endocrine cell development and maintenance involvement. The overlap of enhancer calls with open chromatin between any two cells type was 8.51% - 18.9%. Between 1.89% - 9.94% of calls were unique to one cell type alone. B: First-pass enhancer calls were filtered against the curated FANTOM5 database delineating all identified enhancers in the mouse genome. This resulted in a much more conservative list of 3,535 regions identified (Supplemental Dataset 4). The distribution of enhancers unique or common between cell types remained comparable, with 43.2% identified across all three cell types, and 1.53% - 10.1% unique to a cell type. C: Confirming an enhancer on the second intron of *Slc30a8*, identified in a previous study, with 14 sites of co-binding from multiple transcription factors. D: Confirming an a promoter-proximal enhancer (~1kb upstream) of the gene that codes for the transcription actor *Pdx1*, with 9 sites of co-binding from multiple transcription factors.

Figure 10– Visualizing novel, putative enhancer detection between cell types. A: Visualizing a beta-unique enhancer region. An exonic enhancer region selected from our filtered enhancer call list, with 8 sites of co-binding from various transcription factors relevant to pancreatic islet cell identity and maintenance [1]. B: Visualizing an alpha-unique enhancer region; a distal-intergenic enhancer region (~30kb upstream of *Dusp10*) selected from our filtered enhancer call list, with 6 sites of co-binding from various transcription factors. C: Visualizing a delta-unique enhancer region. A distal-intergenic enhancer region (~21kb upstream of *Gm20745*) selected from our filtered enhancer call list, with 12 sites of co-binding from various transcription factors. D: Visualizing a non-unique enhancer region common across all three cell types. A distal-intergenic enhancer region (~32kb upstream of *Snap25*) selected from our filtered enhancer call list, with 17 sites of co-binding from various transcription factors.

Supplemental Tables

Supplemental Table 1 – Quality control metrics across all ATAC-Seq replicates described.

Supplemental Table 2 – Aggregated dataset description and reference. A: Pancreatic islet ChIP Seq transcription factor data aggregated to identify enhancer and enhancer regions. B: Pancreatic islet histone data aggregated to identify enhancer and enhancer regions. The final approach utilized two histone marks deemed most relevant at delineating putative enhancer regions while taking into account a risk of both false positives and false negatives.

Supplemental Table 3 – Validating open chromatin peaks against known pancreatic islet ChIP binding sites. A: Evaluating the extent of open chromatin– as defined by our ATAC-Seq consensus peak set – contained binding sites for known, pancreatic islet transcription factors. Percent of open chromatin with associated binding sites ranged from 0.31-29.07%. The transcription factors Foxa2, Insm1, and Neurod1 had the highest number of binding sites. B: Evaluating the extent of each ChIP-Seq experiment's binding site calls overlapped with open chromatin. Percent of overlap ranged from 0.35-63.79%. Nkx2.2, Neurod1, and Insm1 had the greatest overlap.

Supplemental Table 4 – Validating motif-calling approach against known ChIP binding sites. A: Pancreatic islet ChIP Seq transcription factor peak calls analyzed by the motif-calling method to determine sensitivity and specificity. True positive calls ranged from 0.59-57%, and false positives ranged from 1.19-8.34%. B: Pancreatic islet ChIP Seq transcription factor peak calls limited to open chromatin determined by the consensus peak set analyzed by the motif-calling method to determine sensitivity and specificity. True positive calls ranged from 4.71-65.10%, and false positives ranged from 1.68-8.81%.

Supplemental Figures

Supplemental Figure 1 – FACS sorting gates used to isolate alpha, beta, and delta cells through our mouse reporter lines. FACS sorting gates isolating beta cells (Ins2-mCherry+) from either alpha (Gcg-YFP+) or delta cells (Sst-YFP+). Double negatives are non-beta and non-alpha or non-delta cells. Double positives (mCherry/YFP+) represent cells with both Ins2 expression and Gcg or Sst expression, reflective of transdifferentiated beta cells. These were not included in any of the samples.

Supplemental Figure 2 – Validating more chromatin accessibility ATAC Seq and companion RNA-Seq expression in alpha, beta, and delta cells against hallmark genes governing its respective cell's identity. All genes are oriented for 5' to 3' end. A-C: Chromatin accessibility and transcript expression across alpha cell hallmark genes *Irx1, Irx2,* and *MafB*. D-F: Chromatin accessibility and transcript expression across beta cell hallmark genes *Abcc8, Nkx6.1,* and *Slc2a2.* G-I: Chromatin accessibility and transcript expression across delta cell hallmark genes *Ffar4, Gabrb3,* and *Ache.*

Supplemental Figure 3 – Chromatin enrichment does not always correlate with associated gene expression. Select hallmark genes defining demonstrating congruent and incongruent chromatin and gene enrichment for cell-specific markers. A: Differentially enriched chromatin at TSS regions and respective gene expression between alpha and beta cells. The majority of cell-specific markers show TSS-enrichment within the cell type of expression. Notably, *Nkx6.1* and *MafA* show TSS enrichment in alpha cells, despite being transcription factors associated with beta cells. B: Differentially enriched chromatin at TSS regions and respective gene expression between alpha and delta cells. C: Differentially enriched chromatin at TSS regions and respective gene expression between beta and beta cells. The majority of cell-specific markers show TSS-enrichment within the cell type of expression.

Supplemental Figure 4 – Evaluating KEGG and gene network enrichment across differentially enriched peaks between alpha and beta cells. A: KEGG enrichment of differentially enriched peaks identified pathways common between the two cell types, or unique to one. B: Gene network enrichment indicative of possible functions of differentially enriched chromatin regions between the two cell types.

Supplemental Figure 5 – Evaluating KEGG and gene network enrichment across differentially enriched peaks between alpha and delta cells. A: KEGG enrichment of differentially enriched peaks identified pathways common between the two cell types, or unique to one. B: Gene network enrichment indicative of possible functions of differentially enriched chromatin regions between the two cell types.

Supplemental Figure 6 – Evaluating KEGG and gene network enrichment across differentially enriched peaks between beta and delta cells. A: KEGG enrichment of differentially enriched peaks

identified pathways common between the two cell types, or unique to one. B: Gene network enrichment indicative of possible functions of differentially enriched chromatin regions between the two cell types.

Supplemental Figure 7 – Verifying transcription factor binding sites and histone mark occurrence at chromatin peaks to determine significance (observed versus expected). A-C: Transcription factors on chromatin regions deemed enriched between differentially enriched chromatin across all three pairwise comparisons.. The majority of transcription factors used in our analysis were deemed statistically significant when observed compared to predicted. D-F: Histone mark occurrence on chromatin regions deemed enriched between differentially enriched chromatin across all three pairwise comparisons. All histone marks used in our analysis were deemed statistically significant when observed compared to predicted.

Supplemental Figure 8 – Aggregated transcription factor ATAC Seq and companion RNA-Seq expression in alpha, beta, and delta cells. All genes are oriented for 5' to 3' end. A-I: Chromatin accessibility and gene expression for aggregated ChIP datasets.

Supplemental Figure 9 – Further illustration of enhancer calls. A: Visualizing a common alpha and delta enhancer region, unavailable in beta cells. B: Further illustration of a beta-unique enhancer region, occurring on the first intron of *Slc2a2,* with 6 co-binding sites for multiple transcription actors. C-D: Two examples of called enhancer regions common across all three cell types. Both are in distal-intergenic regions of the genome and exhibit high transcription factor co-binding activity.

Supplemental Datasets

Supplemental Dataset 1 – Annotated consensus chromatin peak set across alpha, beta and delta cells, along with differential enrichment results between the three pairwise comparisons.

Supplemental Dataset 2 – Congruent and incongruent genes of differentially expressed genes between the three pairwise comparisons. Congruent genes showed gene expression in the same direction as chromatin accessibility enrichment, whereas incongruent genes had opposing expression and enrichment.

Supplemental Dataset 3 – Unfiltered putative enhancer calls defined by open chromatin region in at least one of three cell types, overlapping the histone markers H3K27ac and H3K4me1. These regions were crossed against pancreatic islet transcription factors to identify enhancer regions.

Supplemental Dataset 4 – Filtered putative enhancer calls defined by open chromatin region in at least one of three cell types, overlapping the histone markers H3K27ac and H3K4me1. These regions were crossed against pancreatic islet transcription factors to identify enhancer regions. Last, these data were filtered for regions occurring on curated enhancer calls in the mouse genome using the FANTOM5 database.

References

1. Benner C, van der Meulen T, Caceres E, Tigyi K, Donaldson CJ, Huising MO: **The transcriptional landscape of mouse beta cells compared to human beta cells reveals notable species differences in long non-coding RNA and protein-coding gene expression**. *BMC genomics* 2014, **15**:620.
2. Adriaenssens AE, Svendsen B, Lam BY, Yeo GS, Holst JJ, Reimann F, Gribble FM: **Transcriptomic profiling of pancreatic alpha, beta and delta cell populations identifies delta cells as a principal target for ghrelin in mouse islets**. *Diabetologia* 2016, **59**(10):2156-2165.
3. Nica AC, Ongen H, Irminger JC, Bosco D, Berney T, Antonarakis SE, Halban PA, Dermitzakis ET: **Cell-type, allelic, and genetic signatures in the human pancreatic beta cell transcriptome**. *Genome Res* 2013, **23**(9):1554-1562.
4. Dorrell C, Schug J, Lin CF, Canaday PS, Fox AJ, Smirnova O, Bonnah R, Streeter PR, Stoeckert CJ, Jr., Kaestner KH *et al*: **Transcriptomes of the major human pancreatic cell types**. *Diabetologia* 2011, **54**(11):2832-2844.
5. Ackermann AM, Wang Z, Schug J, Naji A, Kaestner KH: **Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes**. *Mol Metab* 2016, **5**(3):233-244.
6. DiGruccio MR, Mawla AM, Donaldson CJ, Noguchi GM, Vaughan J, Cowing-Zitron C, van der Meulen T, Huising MO: **Comprehensive alpha, beta and delta cell transcriptomes reveal that ghrelin selectively activates delta cells and promotes somatostatin release from pancreatic islets**. *Mol Metab* 2016, **5**(7):449-458.
7. Andrey G, Mundlos S: **The three-dimensional genome: regulating gene expression during pluripotency and development**. *Development* 2017, **144**(20):3646-3658.
8. Xie R, Everett LJ, Lim HW, Patel NA, Schug J, Kroon E, Kelly OG, Wang A, D'Amour KA, Robins AJ *et al*: **Dynamic chromatin remodeling mediated by polycomb proteins orchestrates pancreatic differentiation of human embryonic stem cells**. *Cell Stem Cell* 2013, **12**(2):224-237.
9. Thurner M, Shenhav L, Wesolowska-Andersen A, Bennett AJ, Barrett A, Gloyn AL, McCarthy MI, Beer NL, Efrat S: **Genes Associated with Pancreas Development and Function Maintain Open Chromatin in iPSCs Generated from Human Pancreatic Beta Cells**. *Stem Cell Reports* 2017, **9**(5):1395-1405.
10. Duren Z, Chen X, Jiang R, Wang Y, Wong WH: **Modeling gene regulation from paired expression and chromatin accessibility data**. *Proc Natl Acad Sci U S A* 2017, **114**(25):E4914-E4923.
11. Lawlor N, Marquez EJ, Orchard P, Narisu N, Shamim MS, Thibodeau A, Varshney A, Kursawe R, Erdos MR, Kanke M *et al*: **Multiomic Profiling Identifies cis-Regulatory Networks Underlying Human Pancreatic beta Cell Identity and Function**. *Cell Rep* 2019, **26**(3):788-801 e786.
12. Avrahami D, Li C, Zhang J, Schug J, Avrahami R, Rao S, Stadler MB, Burger L, Schubeler D, Glaser B *et al*: **Aging-Dependent Demethylation of Regulatory Elements Correlates with Chromatin State and Improved beta Cell Function**. *Cell Metab* 2015, **22**(4):619-632.
13. Dayeh T, Volkov P, Salo S, Hall E, Nilsson E, Olsson AH, Kirkpatrick CL, Wollheim CB, Eliasson L, Ronn T *et al*: **Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion**. *PLoS Genet* 2014, **10**(3):e1004160.
14. Dhawan S, Tschen SI, Zeng C, Guo T, Hebrok M, Matveyenko A, Bhushan A: **DNA methylation directs functional maturation of pancreatic beta cells**. *J Clin Invest* 2015, **125**(7):2851-2860.
15. Bramswig NC, Everett LJ, Schug J, Dorrell C, Liu C, Luo Y, Streeter PR, Naji A, Grompe M, Kaestner KH: **Epigenomic plasticity enables human pancreatic alpha to beta cell reprogramming**. *J Clin Invest* 2013, **123**(3):1275-1284.

16. Golson ML, Kaestner KH: **Epigenetics in formation, function, and failure of the endocrine pancreas**. *Mol Metab* 2017, **6**(9):1066-1076.
17. Pullen TJ, Rutter GA: **When less is more: the forbidden fruits of gene repression in the adult beta-cell**. *Diabetes Obes Metab* 2013, **15**(6):503-512.
18. Bhandare R, Schug J, Le Lay J, Fox A, Smirnova O, Liu C, Naji A, Kaestner KH: **Genome-wide analysis of histone modifications in human pancreatic islets**. *Genome Res* 2010, **20**(4):428-433.
19. Pasquali L, Gaulton KJ, Rodriguez-Segui SA, Mularoni L, Miguel-Escalada I, Akerman I, Tena JJ, Moran I, Gomez-Marin C, van de Bunt M *et al*: **Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants**. *Nat Genet* 2014, **46**(2):136-143.
20. van Arensbergen J, Dussaud S, Pardanaud-Glavieux C, Garcia-Hurtado J, Sauty C, Guerci A, Ferrer J, Ravassard P: **A distal intergenic region controls pancreatic endocrine differentiation by acting as a transcriptional enhancer and as a polycomb response element**. *PloS one* 2017, **12**(2):e0171508.
21. Cebola I: **Pancreatic Islet Transcriptional Enhancers and Diabetes**. *Curr Diab Rep* 2019, **19**(12):145.
22. Miguel-Escalada I, Bonas-Guarch S, Cebola I, Ponsa-Cobas J, Mendieta-Esteban J, Atla G, Javierre BM, Rolando DMY, Farabella I, Morgan CC *et al*: **Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes**. *Nat Genet* 2019, **51**(7):1137-1148.
23. Greenwald WW, Chiou J, Yan J, Qiu Y, Dai N, Wang A, Nariai N, Aylward A, Han JY, Kadakia N *et al*: **Pancreatic islet chromatin accessibility and conformation reveals distal enhancer networks of type 2 diabetes risk**. *Nat Commun* 2019, **10**(1):2078.
24. Tennant BR, Robertson AG, Kramer M, Li L, Zhang X, Beach M, Thiessen N, Chiu R, Mungall K, Whiting CJ *et al*: **Identification and analysis of murine pancreatic islet enhancers**. *Diabetologia* 2013, **56**(3):542-552.
25. Greenwald WW, Li H, Benaglio P, Jakubosky D, Matsui H, Schmitt A, Selvaraj S, D'Antonio M, D'Antonio-Chronowska A, Smith EN *et al*: **Subtle changes in chromatin loop contact propensity are associated with differential gene regulation and expression**. *Nat Commun* 2019, **10**(1):1054.
26. Campbell SA, Hoffman BG: **Chromatin Regulators in Pancreas Development and Diabetes**. *Trends Endocrinol Metab* 2016, **27**(3):142-152.
27. Muller C, Leutz A: **Chromatin remodeling in development and differentiation**. *Curr Opin Genet Dev* 2001, **11**(2):167-174.
28. Alvarez-Dominguez JR, Donaghey J, Rasouli N, Kenty JHR, Helman A, Charlton J, Straubhaar JR, Meissner A, Melton DA: **Circadian Entrainment Triggers Maturation of Human In Vitro Islets**. *Cell Stem Cell* 2020, **26**(1):108-122 e110.
29. Mellor J: **The dynamics of chromatin remodeling at promoters**. *Mol Cell* 2005, **19**(2):147-157.
30. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ: **Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position**. *Nat Methods* 2013, **10**(12):1213-1218.
31. Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J *et al*: **Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation**. *Cell* 2012, **148**(1-2):84-98.
32. Birnbaum RY, Clowney EJ, Agamy O, Kim MJ, Zhao J, Yamanaka T, Pappalardo Z, Clarke SL, Wenger AM, Nguyen L *et al*: **Coding exons function as tissue-specific enhancers of nearby genes**. *Genome Res* 2012, **22**(6):1059-1068.
33. Rose AB: **Introns as Gene Regulators: A Brick on the Accelerator**. *Front Genet* 2018, **9**:672.
34. Kulaeva OI, Nizovtseva EV, Polikanov YS, Ulianov SV, Studitsky VM: **Distant activation of transcription: mechanisms of enhancer action**. *Mol Cell Biol* 2012, **32**(24):4892-4897.

35. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ: **ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide**. *Current protocols in molecular biology / edited by Frederick M Ausubel [et al]* 2015, **109**:21 29 21-21 29 29.
36. Gaulton KJ, Nammo T, Pasquali L, Simon JM, Giresi PG, Fogarty MP, Panhuis TM, Mieczkowski P, Secchi A, Bosco D *et al*: **A map of open chromatin in human pancreatic islets**. *Nat Genet* 2010, **42**(3):255-259.
37. Arda HE, Tsai J, Rosli YR, Giresi P, Bottino R, Greenleaf WJ, Chang HY, Kim SK: **A Chromatin Basis for Cell Lineage and Disease Risk in the Human Pancreas**. *Cell Syst* 2018, **7**(3):310-322 e314.
38. Bysani M, Agren R, Davegardh C, Volkov P, Ronn T, Unneberg P, Bacos K, Ling C: **ATAC-seq reveals alterations in open chromatin in pancreatic islets from subjects with type 2 diabetes**. *Sci Rep* 2019, **9**(1):7785.
39. Thurner M, van de Bunt M, Torres JM, Mahajan A, Nylander V, Bennett AJ, Gaulton KJ, Barrett A, Burrows C, Bell CG *et al*: **Integration of human pancreatic islet genomic data refines regulatory mechanisms at Type 2 Diabetes susceptibility loci**. *eLife* 2018, **7**:1-30.
40. Rai V, Quang DX, Erdos MR, Cusanovich DA, Daza RM, Narisu N, Zou LS, Didion JP, Guan Y, Shendure J *et al*: **Single-cell ATAC-Seq in human pancreatic islets and deep learning upscaling of rare cells reveals cell-specific type 2 diabetes regulatory signatures**. *Mol Metab* 2020, **32**:109-121.
41. Chiou J, Zeng C, Cheng Z, Han JY, Schlichting M, Miller M, Mendez R, Huang S, Wang J, Sui Y *et al*: **Single-cell chromatin accessibility identifies pancreatic islet cell type- and state-specific regulatory programs of diabetes risk**. *Nat Genet* 2021, **53**(4):455-466.
42. Kleiber T, Davidson G, Mengus G, Martianov I, Davidson I: **Single cell transcriptomics reveal trans-differentiation of pancreatic beta cells following inactivation of the TFIID subunit Taf4**. *Cell death & disease* 2021, **12**(8):790.
43. Mawla AM, Huising MO: **epiRomics: a multi-omics R package to identify and visualize enhancers**. *bioRxiv* 2021:2021.2008.2019.456732.
44. Taniguchi H, He M, Wu P, Kim S, Paik R, Sugino K, Kvitsiani D, Fu Y, Lu J, Lin Y *et al*: **A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex**. *Neuron* 2011, **71**(6):995-1013.
45. Herrera PL: **Adult insulin- and glucagon-producing cells differentiate from two independent cell lineages**. *Development* 2000, **127**(11):2317-2322.
46. Huising MO, van der Meulen T, Vaughan JM, Matsumoto M, Donaldson CJ, Park H, Billestrup N, Vale WW: **CRFR1 is expressed on pancreatic beta cells, promotes beta cell proliferation, and potentiates insulin secretion in a glucose-dependent manner**. *Proc Natl Acad Sci U S A* 2010, **107**(2):912-917.
47. van der Meulen T, Xie R, Kelly OG, Vale WW, Sander M, Huising MO: **Urocortin 3 marks mature human primary and embryonic stem cell-derived pancreatic alpha and beta cells**. *PloS one* 2012, **7**(12):e52181.
48. Srinivas S, Watanabe T, Lin CS, William CM, Tanabe Y, Jessell TM, Costantini F: **Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus**. *BMC Dev Biol* 2001, **1**:4.
49. Andrews SR: **FastQC: A Quality Control Tool for High Throughput Sequence Data**. In. Online; 2015.
50. Joshi NA FJ: **A sliding-window, adaptive, quality-based trimming tool for FastQ files** *Online* 2011.
51. Williams CR, Baccarella A, Parrish JZ, Kim CC: **Trimming of sequence reads alters RNA-Seq gene expression estimates**. *BMC Bioinformatics* 2016, **17**:103.
52. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S *et al*: **GENCODE: the reference human genome annotation for The ENCODE Project**. *Genome Res* 2012, **22**(9):1760-1774.

53. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2**. *Nat Methods* 2012, **9**(4):357-359.
54. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W *et al*: **Model-based analysis of ChIP-Seq (MACS)**. *Genome Biol* 2008, **9**(9):R137.
55. Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, Brown GD, Gojis O, Ellis IO, Green AR *et al*: **Differential oestrogen receptor binding is associated with clinical outcome in breast cancer**. *Nature* 2012, **481**(7381):389-393.
56. Gontarz P, Fu S, Xing X, Liu S, Miao B, Bazylianska V, Sharma A, Madden P, Cates K, Yoo A *et al*: **Comparison of differential accessibility analysis strategies for ATAC-seq data**. *Sci Rep* 2020, **10**(1):10150.
57. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, Hilton JA, Jain K, Baymuradov UK, Narayanan AK *et al*: **The Encyclopedia of DNA elements (ENCODE): data portal update**. *Nucleic acids research* 2018, **46**(D1):D794-D801.
58. Yan F, Powell DR, Curtis DJ, Wong NC: **From reads to insight: a hitchhiker's guide to ATAC-seq data analysis**. *Genome Biol* 2020, **21**(1):22.
59. Schep AN, Wu B, Buenrostro JD, Greenleaf WJ: **chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data**. *Nat Methods* 2017, **14**(10):975-978.
60. Gel B, Diez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R: **regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests**. *Bioinformatics* 2016, **32**(2):289-291.
61. Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ: **Software for computing and annotating genomic ranges**. *PLoS Comput Biol* 2013, **9**(8):e1003118.
62. Schep A: **motifmatchr: Fast Motif Matching in R**. In., 1.12.0 edn. BioConductor: R; 2020.
63. Kulakovskiy IV, Vorontsov IE, Yevshin IS, Sharipov RN, Fedorova AD, Rumynskiy EI, Medvedeva YA, Magana-Mora A, Bajic VB, Papatsenko DA *et al*: **HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis**. *Nucleic acids research* 2018, **46**(D1):D252-D259.
64. Yu G, Wang LG, He QY: **ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization**. *Bioinformatics* 2015, **31**(14):2382-2383.
65. Yu G, He QY: **ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization**. *Mol Biosyst* 2016, **12**(2):477-479.
66. Yu G, Wang LG, Han Y, He QY: **clusterProfiler: an R package for comparing biological themes among gene clusters**. *OMICS* 2012, **16**(5):284-287.
67. van der Meulen T, Mawla AM, DiGruccio MR, Adams MW, Nies V, Dolleman S, Liu S, Ackermann AM, Caceres E, Hunter AE *et al*: **Virgin Beta Cells Persist throughout Life at a Neogenic Niche within Pancreatic Islets**. *Cell Metab* 2017, **25**(4):911-926 e916.
68. Blodgett DM, Nowosielska A, Afik S, Pechhold S, Cura AJ, Kennedy NJ, Kim S, Kucukural A, Davis RJ, Kent SC *et al*: **Novel Observations From Next-Generation RNA Sequencing of Highly Purified Human Adult and Fetal Islet Cell Subsets**. *Diabetes* 2015, **64**(9):3172-3181.
69. Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, Sandstrom R, Ma Z, Davis C, Pope BD *et al*: **A comparative encyclopedia of DNA elements in the mouse genome**. *Nature* 2014, **515**(7527):355-364.
70. Starks RR, Biswas A, Jain A, Tuteja G: **Combined analysis of dissimilar promoter accessibility and gene expression profiles identifies tissue-specific genes and actively repressed networks**. *Epigenetics Chromatin* 2019, **12**(1):16.
71. de la Torre-Ubieta L, Stein JL, Won H, Opland CK, Liang D, Lu D, Geschwind DH: **The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis**. *Cell* 2018, **172**(1-2):289-304 e218.

72. Daugherty AC, Yeo RW, Buenrostro JD, Greenleaf WJ, Kundaje A, Brunet A: **Chromatin accessibility dynamics reveal novel functional enhancers in C. elegans**. *Genome Res* 2017, **27**(12):2096-2107.
73. Neph S, Stergachis AB, Reynolds A, Sandstrom R, Borenstein E, Stamatoyannopoulos JA: **Circuitry and dynamics of human transcription factor regulatory networks**. *Cell* 2012, **150**(6):1274-1286.
74. Gowd V, Gurukar A, Chilkunda ND: **Glycosaminoglycan remodeling during diabetes and the role of dietary factors in their modulation**. *World J Diabetes* 2016, **7**(4):67-73.
75. Khoo C, Yang J, Weinrott SA, Kaestner KH, Naji A, Schug J, Stoffers DA: **Research resource: the pdx1 cistrome of pancreatic islets**. *Mol Endocrinol* 2012, **26**(3):521-533.
76. Taylor BL, Liu FF, Sander M: **Nkx6.1 is essential for maintaining the functional state of pancreatic beta cells**. *Cell Rep* 2013, **4**(6):1262-1275.
77. Jia S, Ivanov A, Blasevic D, Muller T, Purfurst B, Sun W, Chen W, Poy MN, Rajewsky N, Birchmeier C: **Insm1 cooperates with Neurod1 and Foxa2 to maintain mature pancreatic beta-cell function**. *EMBO J* 2015, **34**(10):1417-1433.
78. Gutierrez GD, Bender AS, Cirulli V, Mastracci TL, Kelly SM, Tsirigos A, Kaestner KH, Sussel L: **Pancreatic beta cell identity requires continual repression of non-beta cell programs**. *J Clin Invest* 2017, **127**(1):244-259.
79. Piccand J, Strasser P, Hodson DJ, Meunier A, Ye T, Keime C, Birling MC, Rutter GA, Gradwohl G: **Rfx6 maintains the functional identity of adult pancreatic beta cells**. *Cell Rep* 2014, **9**(6):2219-2232.
80. Ediger BN, Lim HW, Juliana C, Groff DN, Williams LT, Dominguez G, Liu JH, Taylor BL, Walp ER, Kameswaran V *et al*: **LIM domain-binding 1 maintains the terminally differentiated state of pancreatic beta cells**. *J Clin Invest* 2017, **127**(1):215-229.
81. Rabhi N, Denechaud PD, Gromada X, Hannou SA, Zhang H, Rashid T, Salas E, Durand E, Sand O, Bonnefond A *et al*: **KAT2B Is Required for Pancreatic Beta Cell Adaptation to Metabolic Stress by Controlling the Unfolded Protein Response**. *Cell Rep* 2016, **15**(5):1051-1061.
82. Martinelli P, Madriles F, Canamero M, Pau EC, Pozo ND, Guerra C, Real FX: **The acinar regulator Gata6 suppresses KrasG12V-driven pancreatic tumorigenesis in mice**. *Gut* 2016, **65**(3):476-486.
83. Sander M, Neubuser A, Kalamaras J, Ee HC, Martin GR, German MS: **Genetic analysis reveals that PAX6 is required for normal transcription of pancreatic hormone genes and islet development**. *Genes Dev* 1997, **11**(13):1662-1673.
84. Ampofo E, Nalbach L, Menger MD, Laschke MW: **Regulatory Mechanisms of Somatostatin Expression**. *Int J Mol Sci* 2020, **21**(11).
85. Vanheer L, Schiavo AA, Van Haele M, Haesen T, Janiszewski A, Chappell J, Roskams T, Cnop M, Pasque V: **Revealing the Key Regulators of Cell Identity in the Human Adult Pancreas**. 2020.
86. Burlison JS, Long Q, Fujitani Y, Wright CV, Magnuson MA: **Pdx-1 and Ptf1a concurrently determine fate specification of pancreatic multipotent progenitor cells**. *Dev Biol* 2008, **316**(1):74-86.
87. El-Gohary Y, Tulachan S, Guo P, Welsh C, Wiersch J, Prasadan K, Paredes J, Shiota C, Xiao X, Wada Y *et al*: **Smad signaling pathways regulate pancreatic endocrine development**. *Dev Biol* 2013, **378**(2):83-93.
88. Lin HM, Lee JH, Yadav H, Kamaraju AK, Liu E, Zhigang D, Vieira A, Kim SJ, Collins H, Matschinsky F *et al*: **Transforming growth factor-beta/Smad3 signaling regulates insulin gene transcription and pancreatic islet beta-cell function**. *J Biol Chem* 2009, **284**(18):12246-12257.
89. Henseleit KD, Nelson SB, Kuhlbrodt K, Hennings JC, Ericson J, Sander M: **NKX6 transcription factor activity is required for alpha- and beta-cell development in the pancreas**. *Development* 2005, **132**(13):3139-3149.

90. Byrnes LE, Wong DM, Subramaniam M, Meyer NP, Gilchrist CL, Knox SM, Tward AD, Ye CJ, Sneddon JB: **Lineage dynamics of murine pancreatic development at single-cell resolution**. *Nat Commun* 2018, **9**(1):3922.
91. Lu TT, Heyne S, Dror E, Casas E, Leonhardt L, Boenke T, Yang CH, Sagar, Arrigoni L, Dalgaard K *et al*: **The Polycomb-Dependent Epigenome Controls beta Cell Dysfunction, Dedifferentiation, and Diabetes**. *Cell Metab* 2018, **27**(6):1294-1308 e1297.
92. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA *et al*: **Histone H3K27ac separates active from poised enhancers and predicts developmental state**. *Proc Natl Acad Sci U S A* 2010, **107**(50):21931-21936.
93. Spicuglia S, Vanhille L: **Chromatin signatures of active enhancers**. *Nucleus* 2012, **3**(2):126-131.
94. Calo E, Wysocka J: **Modification of enhancer chromatin: what, how, and why?** *Mol Cell* 2013, **49**(5):825-837.
95. Lizio M, Harshbarger J, Abugessaisa I, Noguchi S, Kondo A, Severin J, Mungall C, Arenillas D, Mathelier A, Medvedeva YA *et al*: **Update of the FANTOM web resource: high resolution transcriptome of diverse cell types in mammals**. *Nucleic acids research* 2017, **45**(D1):D737-D743.
96. Pound LD, Hang Y, Sarkar SA, Wang Y, Milam LA, Oeser JK, Printz RL, Lee CE, Stein R, Hutton JC *et al*: **The pancreatic islet beta-cell-enriched transcription factor Pdx-1 regulates Slc30a8 gene transcription through an intronic enhancer**. *Biochem J* 2011, **433**(1):95-105.
97. Shaul O: **How introns enhance gene expression**. *Int J Biochem Cell Biol* 2017, **91**(Pt B):145-155.
98. Perelis M, Marcheva B, Ramsey KM, Schipma MJ, Hutchison AL, Taguchi A, Peek CB, Hong H, Huang W, Omura C *et al*: **Pancreatic beta cell enhancers regulate rhythmic transcription of genes controlling insulin secretion**. *Science* 2015, **350**(6261):aac4250.
99. Mularoni L, Ramos-Rodriguez M, Pasquali L: **The Pancreatic Islet Regulome Browser**. *Front Genet* 2017, **8**:13.
100. Chera S, Baronnier D, Ghila L, Cigliola V, Jensen JN, Gu G, Furuyama K, Thorel F, Gribble FM, Reimann F *et al*: **Diabetes recovery by age-dependent conversion of pancreatic delta-cells into insulin producers**. *Nature* 2014, **514**(7523):503-507.
101. Feng Y, Qiu WL, Yu XX, Zhang Y, He MY, Li LC, Yang L, Zhang W, Franti M, Ye J *et al*: **Characterizing pancreatic beta-cell heterogeneity in the streptozotocin model by single-cell transcriptomic analysis**. *Mol Metab* 2020, **37**:100982.
102. Dai C, Brissova M, Hang Y, Thompson C, Poffenberger G, Shostak A, Chen Z, Stein R, Powers AC: **Islet-enriched gene expression and glucose-induced insulin secretion in human and mouse islets**. *Diabetologia* 2012, **55**(3):707-718.
103. Ono Y, Kataoka K: **MafA, NeuroD1, and HNF1beta synergistically activate the Slc2a2 (Glut2) gene in beta-cells**. *J Mol Endocrinol* 2021, **67**(3):71-82.

VIII. Chapter 8: epiRomics: a multi-omics R package for identifying and visualizing enhancers

Alex M. Mawla and Mark O. Huising

Final draft, in process of submission, with pre-print available on biorXiv

Contributions to Jointly Authored Works: As first author of this manuscript, I conceived and devised the approach, and wrote and currently maintain the R package. I was responsible for pursuing validation using a previously published dataset in human pancreatic islets, alongside aggregated human transcription factor and histone mark data, later cross-referencing findings of active enhancers detected against SNP data associated with diabetes in human subjects. I was also responsible for writing this short manuscript intended to share the application so that other colleagues in the scientific community may utilize the approach if they choose so. I am currently in the process of adding further validation to this approach in order to submit the manuscript for publication.

Significance of Research: This manuscript details a package I designed intended to provide a straightforward, and clear solution to integrate multi-omics data and delineate enhancers during the course of writing Chapter 7.

epiRomics: a multi-omics R package for identifying and visualizing enhancers

Alex M. Mawla[1], Mark O. Huising[1,2*]

[1]Department of Neurobiology, Physiology & Behavior, College of Biological Sciences, University of California, Davis, CA 95616, USA

[2]Department of Physiology and Membrane Biology, School of Medicine, University of California, Davis, CA 95616, USA

* Corresponding author:

Dr. Mark O. Huising, Ph.D.

Professor

University of California, Davis

Department of Neurobiology, Physiology & Behavior, College of Biological Sciences

Department of Physiology and Membrane Biology, School of Medicine

One Shields Avenue,

Davis, CA, 95616

Ph: (530) 752-4670

Email: mhuising@ucdavis.edu

Abstract

Summary: epiRomics is an R package designed to integrate multi-omics data in order to identify and visualize enhancer regions alongside gene expression and other epigenomic modifications. Regulatory network analysis can be done using combinatory approaches to infer regions of significance such as enhancers, when combining ChIP and histone data. Downstream analysis can identify co-occurrence of these regions of interest with other user-supplied data, such as chromatin availability or gene expression. Finally, this package allows for results to be visualized at high resolution in a stand-alone browser.

Availability and Implementation: epiRomics is released under Artistic-2.0 License. The source code and documents are freely available through Github (https://github.com/Huising-Lab/epiRomics).

Contact: ammawla@ucdavis.edu or mhuising@ucdavis.edu

Supplementary information: Supplementary data, and methods are available online on biorxiv or *Github.*

Introduction

The evaluation of the transcriptional landscape between cell types grants the scientific community a deeper understanding of cellular identity, and helps paint the underlying mechanisms that drive phenotype and function [1]. Bulk RNA sequencing has been a gold standard in the field, followed more recently with the advent of single-omics approaches [2-5]. However, gene expression represents only a single aspect of what is a sophisticated and interlaced network of genetic and epigenomic regulators that drive and determine cell identity, with perturbations leading to dysfunction and, sometimes, disease [6].

Chromatin remodeling is a dynamic process that represents one of the epigenetic layers of cell fate maintenance and identity [7, 8]. Approaches such as DNase I hypersensitive site sequencing (DNASE-Seq) [9] and Assay for transposase-accessible chromatin using sequencing (ATAC-Seq) [10, 11], are commonly used to compare chromatin accessibility between cell types and states. Chromatin immunoprecipitation Sequencing (ChIP-Seq) is another approach used to assess the epigenomic regulators driven by specific transcription factors acting as either activators or suppressors on the genic region, or at distal-intergenic regions, associated with enhancer activity [12-16]. Alternatively, ChIP-Seq is used in conjunction with antibodies that pull down specific histone modifications associated with regions of chromatin, whose presence can be used to infer whether the region is active, poised or repressed [17, 18]. The interrogation of transcription factor binding is not enough to infer transcriptional behavior, as whether or not chromatin is accessible between cell types, and whether the region is active, poised, or repressed, demarked by co-occurrences of histone marks, must be considered in order to fully evaluate the biology [17, 19]. Lastly, methylation of DNA, quantifiable through whole-genome bisulfite sequencing analysis (BS-Methyl Seq) [20], can help determine whether accessible chromatin that has recruited the correct histone marks is even available for transcription factor binding and recruitment of co-modulators to drive or suppress gene transcription within cell types [21, 22].

While tools exist to compare these different layers in a pairwise manner, a gap exists to integrate multiple omics layers quickly, and easily to generate high resolution visuals in order to derive more biological meaning behind results. We developed a novel 'epigenomics in R', epiRomics, package to solve this issue. We designed epiRomics to accept either browser extensible data (BED) [23] or bigwig [24] files as input for any of the aforementioned types of data. Inclusion of functional annotations, *i.e.* FANTOM [25], single nucleotide polymorphism (SNP) data from GWAS [26, 27], or Ultra Conserved Non Coding Elements (UCNEs) [28] is also possible – in order to more fully integrate many slices of the genetic and epigenetic pie.

Functions

epiRomics takes in a user-submitted comma-separated values (csv) file containing hard paths to all BED or bigwig formatted files, optional hexadecimal (hex) color code associations for each file and a user-defined label to group each input data set (*e.g.,* ChIP, ATAC, RNA, functional, etc.). The *epiRomics_build_dB* command quickly generates a comprehensive and easily accessible variable of the class "epiRomics-class" containing a GenomicRanges (*GRanges)* object [29] that tracks each of these submitted data, along with all other data related to the species, pulled automatically from the UCSC genome database [23]. The epiRomics-class variable can easily be integrated with other packages, and the user can also save these data in a csv format for further manual exploration in excel, or other comparable third-party tools.

Putative enhancers can efficiently be called, and then categorized separately – active, poised, repressed, etc., through the *epiRomics_putative_enhancers* function, which will consider user provided histone data. For example, the histone marks H3K27ac and H3K4me1 are commonly used to demark active enhancer regions [17, 30-32] outputting an epiRomics-class variable for downstream use within the package, or outside. This variable can be used further to identify key enhanceosome regions with evidence

of co-binding of multiple user-selected transcription factors by implementing the *epiRomics_putative_enhanceosome* command. These data can also be filtered against functional data annotations, such as methylation calls, FANTOM, SNP regions, or UCNEs, through the use of *epiRomics_putative_enhancers_filtered.* A side function is provided within the package making use of decision trees [33] in order to classify which transcription factors were most meaningfully associated with different enhancer types, through the use of *epiRomics_predictors*.

For visualization of these differently classified regions, and integration with bigwig data such as gene expression or chromatin availability between cell types, the tool *epiRomics_track_layer* can be used. This makes use of the package *GViz [34]* to generate resolution, publication-quality encapsulated postscript (eps) files. Specific calls for enhancer regions provided by *epiRomics_putative_enhanceosome* can be visualized. Conversely, if users have specific regions or genes of interest they wish to evaluate, they can do so using *epiRomics_region_of_interest.*

These tools were designed to allow biological relevance to be determined from the integrated multi-omics data that is available for a particular tissue or cell type. For example, a common enhancer region may be present between cell types of a common progenitor, with chromatin accessible across all cell types, methylation may block activity in one cell type, but not the other. Drug treatment, or healthy versus diseased comparisons can quickly be made, and the multitude of SNPs amassed via GWAS can be seamlessly connected to narrow in on deleterious variants that may contribute to disease.

Results

epiRomics is developed as an R package to be made available through Bioconductor [35], and is available under Artistic-2.0 License. epiRomics is designed to integrate a multitude of -omics data – in either BED or bigwig format – in order to identify regions of regulatory interest, such as enhancers, and provide sophisticated, high quality resolution visuals in EPS format for use in publications. Users with little programming experience can use epiRomics to encode colors for individual tracks, cross-reference diverse

types of -omics data – such as ATAC- and RNA- Seq, and produce strong candidate lists for putative enhancers common or unique to cell types. Finally, epiRomics is easy to use, with a full walkthrough with sample data available through its companion vignette.

Funding

This work was supported by grants from the National Institutes of Health (NIDDK110276), the Juvenile Diabetes Research Foundation (2-SRA-2021-1054-M-N) and American Diabetes Association (1-19-IBS-078) to M.O.H. A.M.M. was supported by the Stephen F. and Bettina A. Sims Immunology Fellowship and the Amazon AWS Machine Learning Research Fellowship.

Conflict of interest

The authors have no conflicts of interests to declare.

Getting Started with EpiRomics

Authors: Alex M. Mawla & Mark O. Huising.

2021-10-05

Contents

Abstract

Summary epiRomics is an R package designed to integrate multi-omics data in order to identify and visualize enhancer regions alongside gene expression and other epigenomic modifications. Regulatory network analysis can be done using combinatory approaches to infer regions of significance such as enhancers, when combining ChIP and histone data. Downstream analysis can identify co-occurrence of these regions of interest with

other user-supplied data, such as chromatin availability or gene expression. Finally, this package allows for results to be visualized at high resolution in a stand-alone browser.

Availability and Implementation epiRomics is released under Artistic-2.0 License. The source code and documents are freely available through Github (https://github.com/Huising-Lab/epiRomics).

Contact ammawl@ucdavis.edu or mhuising@ucdavis.edu

Supplementary information Supplementary data, and methods are available online on *biorXiv* or *Github.*

Competing Interest Statement

The authors have declared no competing interest.

Citation

If you use epiRomics in published research, please cite:

Mawla, AM& Huising, MO. **epiRomics: a multi-omics R package to identify and visualize enhancers.**

biorXiv 2021. doi:https://doi.org/10.1101/2021.08.19.456732

Loading the epiRomics package and dependencies for vignette

```
## loading packages

library(epiRomics)

library(TxDb.Hsapiens.UCSC.hg38.knownGene)

library(org.Hs.eg.db)
```

Brief explanation of example data

This package includes some example data to get you started, delineating human pancreatic islet enhancers between alpha and beta cells.

Human pancreatic islet alpha and beta ATAC- and companion RNA- Seq data were retrieved from GEO accession GSE76268 (Ackermann, et al., 2016).

ATAC samples were processed using the ENCODE-DCC ATAC sequencing pipeline, aligning to the hg38 (Harrow, et al., 2012) build of the human genome (Consortium, 2012; Davis, et al., 2018).

Peak calls generated through the pipeline using MACS2 (Zhang, et al., 2008) were analyzed downstream through the BioConductor package DiffBind (Ross-Innes, et al., 2012) in order to identify differentially enriched chromatin regions between the two cell types.

RNA samples were quality controlled using the tool fastp (Chen, et al., 2018), and aligned using STAR (Dobin, et al., 2013) to the hg38 build of the human genome. Wiggle files produced by the STAR aligner were then merged by cell type using UCSC command line tools.

Bigwigs merged by cell type were subsetted to chromosome 1 using UCSC command line tools (Kent, et al., 2010).

ChIP-sequencing peak calls generated using MACS2 for human pancreatic islet transcription factors Foxa2, MafB, Nkx2.2, Nkx6.1, and Pdx1 were retrieved from the EMBL-EBI repository database E-MTAB-1919 (Pasquali, et al., 2014). All peak calls were lifted over to the hg38 genome build using the UCSC genome browser liftOver tool (Kent, et al., 2002).

Histone-sequencing peak calls generated using MACS2 for histones H3k27ac and H3k4me1 were retrieved from GEO accession GSE16256 (Bernstein, et al., 2010), and for histone H2A.Z from the EMBL-EBI repository database E-MTAB-1919 (Pasquali, et al., 2014). All peak calls were lifted over to the hg38 genome build using the UCSC genome browser liftOver tool.

The FANTOM5 human enhancer database (Lizio, et al., 2015) was retrieved, and all regions were lifted over to the hg38 genome build using the UCSC genome browser liftOver tool.

Human ultra-conserved non-coding elements (UCNEs) were retrieved form the UCNE database (Dimitrieva and Bucher, 2012), and all regions were lifted over to the hg38 genome build using the UCSC genome browser liftOver tool.

The human islet regulome database was retrieved (Miguel-Escalada, et al., 2019) and all regions were lifted over to the hg38 genome build using the UCSC genome browser liftOver tool.

How to load and build the database

Lets load and take a look at how to properly format the datasets epiRomics uses to build the initial database.

```
example_epiRomics_Db_sheet <- read.csv(
  file = system.file(
    "extdata",
    "example_epiRomics_Db_sheet_user_paths.csv",
    package = "epiRomics"
  )
)


## Required columns are: name, path, genome, format, and type

## The genome must also be in proper format, e.g. mm10 or hg38

## Type of data can be histone, methyl, SNP, or ChIP.
## ChIP is required for some downstream functions to work appropriately.

## Not run
#head(example_epiRomics_Db_sheet)
```

epiRomics_build_dB constructs a database of class epiRomics with this data sheet

```
epiRomics_dB <- epiRomics_build_dB(
  epiRomics_db_file =
    system.file(
      "extdata",
      "example_epiRomics_Db_sheet_user_paths.csv",
      package = "epiRomics"
    ),
  txdb_organism =
```

```
    "TxDb.Hsapiens.UCSC.hg38.knownGene::TxDb.Hsapiens.UCSC.hg38.knownGene",
  epiRomics_genome = "hg38",
  epiRomics_organism = "org.Hs.eg.db"
)
#> Building enhancers...
#> snapshotDate(): 2021-05-18
#> loading from cache
#> 'select()' returned 1:1 mapping between keys and columns
#> Building promoters...
#> Building 1to5kb upstream of TSS...
#> Building intergenic...
#> Building cds...
#> Building 5UTRs...
#> Building 3UTRs...
#> Building exons...
#> Building first exons...
#> Building introns...
#> Building intron exon boundaries...
#> Building exon intron boundaries...
#> Building CpG islands...
#> Building CpG shores...
#> Building CpG shelves...
#> Building inter-CpG-islands...
#> snapshotDate(): 2021-05-18
#> Building lncRNA transcripts...
#> loading from cache
```

Delineating active enhancers using H3k4me1 and H3k27ac marks as a proxy

There is a lot of flexibility for data exploration here. In this example, we search for putative enhancers using two histone marks known to co-occur at enhancer regions - h3k4me1 & h3k27ac

```
epiRomics_putative_enhancers <-
  epiRomics_enhancers(
    epiRomics_dB,
    epiRomics_histone_mark_1 =
      "h3k4me1",
    epiRomics_histone_mark_2 = "h3k27ac"
  )

## Taking a look, we see a  list of 19,692 putative enhancers

epiRomics_putative_enhancers@annotations
#> GRanges object with 19692 ranges and 0 metadata columns:
#>           seqnames            ranges strand
#>              <Rle>         <IRanges>  <Rle>
#>       [1]     chr1    999886-1000011      *
#>       [2]     chr1   1000228-1000811      *
#>       [3]     chr1   1000850-1001468      *
#>       [4]     chr1   1005007-1006023      *
```

```
#>       [5]     chr1   1013701-1013893      *
#>       ...      ...                ...    ...
#>   [19688]     chrY 12392544-12392994      *
#>   [19689]     chrY 13282680-13282760      *
#>   [19690]     chrY 15455449-15455788      *
#>   [19691]     chrY 19066496-19066508      *
#>   [19692]     chrY 19075542-19075899      *
#>   -------
#>   seqinfo: 595 sequences (1 circular) from hg38 genome
```

Cross-referencing enhancer calls to other databases

FANTOM Enhancer Database

Now we have a list of regions as possible candidates for enhancers, but where do we go from here? One way to increase confidence of these calls is to cross this list against an enhancer database, for instance, FANTOM.

```
## NOTE: This option may not be available for all organisms.

epiRomics_putative_enhancers_filtered_fantom <-
  epiRomics_enhancers_filter(epiRomics_putative_enhancers, epiRomics_dB,
                             epiRomics_type =
                               "hg38_custom_fantom")

## Taking a look, we see a reduced number of 2,749 candidate regions

epiRomics_putative_enhancers_filtered_fantom@annotations
#> GRanges object with 2749 ranges and 0 metadata columns:
#>          seqnames              ranges strand
#>             <Rle>           <IRanges>  <Rle>
#>      [1]     chr1     1021242-1021277      *
#>      [2]     chr1     1021318-1021698      *
#>      [3]     chr1     1079632-1080061      *
#>      [4]     chr1     1080101-1080628      *
#>      [5]     chr1     1128200-1128445      *
#>      ...      ...                 ...    ...
#>   [2745]     chrX 154369950-154370183      *
#>   [2746]     chrX 154371971-154372237      *
#>   [2747]     chrX 154372350-154372695      *
#>   [2748]     chrX 154517139-154517596      *
#>   [2749]     chrX 154734550-154734738      *
#>   -------
#>   seqinfo: 595 sequences (1 circular) from hg38 genome
```

Human Pancreatic Islet Regulome Enhancer Database

We can also filter putative enhancer calls against active enhancers from the human islet regulome database

```
epiRomics_putative_enhancers_filtered_regulome_active <-
  epiRomics_enhancers_filter(epiRomics_putative_enhancers,
```

```
                            epiRomics_dB,
                            epiRomics_type = "hg38_custom_regulome_active")


epiRomics_putative_enhancers_filtered_regulome_active@annotations
#> GRanges object with 6025 ranges and 0 metadata columns:
#>          seqnames              ranges strand
#>             <Rle>           <IRanges>  <Rle>
#>      [1]     chr1     1068896-1068951      *
#>      [2]     chr1     1069171-1069333      *
#>      [3]     chr1     1079632-1080061      *
#>      [4]     chr1     1080101-1080628      *
#>      [5]     chr1     1158358-1158930      *
#>      ...      ...                 ...    ...
#>   [6021]     chrX 153381411-153381523      *
#>   [6022]     chrX 153381677-153381956      *
#>   [6023]     chrX 153382322-153382448      *
#>   [6024]     chrX 153985442-153985689      *
#>   [6025]     chrX 154091801-154091996      *
#>   -------
#>   seqinfo: 595 sequences (1 circular) from hg38 genome
```

Human Pancreatic Islet Regulome Super-Enhancer Database

We can also filter putative enhancer calls against super enhancers from human islet regulome database

```
epiRomics_putative_enhancers_filtered_regulome_super <-
  epiRomics_enhancers_filter(epiRomics_putative_enhancers,
                            epiRomics_dB, epiRomics_type = "hg38_custom_regulome_super")

epiRomics_putative_enhancers_filtered_regulome_super@annotations
#> GRanges object with 2401 ranges and 0 metadata columns:
#>          seqnames            ranges strand
#>             <Rle>         <IRanges>  <Rle>
#>      [1]     chr1   7574092-7574479      *
#>      [2]     chr1   7574640-7575094      *
#>      [3]     chr1   8169274-8169689      *
#>      [4]     chr1   8170112-8170857      *
#>      [5]     chr1   8174089-8174358      *
#>      ...      ...               ...    ...
#>   [2397]    chr22 46109916-46110442      *
#>   [2398]    chr22 46115774-46116154      *
#>   [2399]    chr22 46116326-46116501      *
#>   [2400]     chrX 39813348-39813627      *
#>   [2401]     chrX 39814304-39814607      *
#>   -------
#>   seqinfo: 595 sequences (1 circular) from hg38 genome
```

Human Ultra-Conserved Non-Coding Elements Database

We can also filter putative enhancer calls against Ultra-Conserved Non-Coding Elements

```
epiRomics_putative_enhancers_filtered_ucnes <-
  epiRomics_enhancers_filter(epiRomics_putative_enhancers,
                             epiRomics_dB,
                             epiRomics_type = "hg38_custom_ucnes")

epiRomics_putative_enhancers_filtered_ucnes@annotations
#> GRanges object with 11 ranges and 0 metadata columns:
#>        seqnames              ranges strand
#>           <Rle>           <IRanges>  <Rle>
#>    [1]     chr1 164635220-164635921      *
#>    [2]     chr1 164711914-164712296      *
#>    [3]     chr1 164712350-164713071      *
#>    [4]     chr1 200079185-200079426      *
#>    [5]     chr1 213585694-213586385      *
#>    [6]     chr3   71131859-71132164      *
#>    [7]     chr9 106921420-106921764      *
#>    [8]    chr11 114163425-114164860      *
#>    [9]    chr15   36903894-36904085      *
#>   [10]    chr15   53447393-53447809      *
#>   [11]    chr21   16534340-16534665      *
#>   -------
#>   seqinfo: 595 sequences (1 circular) from hg38 genome
```

Screening for high transcription factor co-binding sites

Biology has established that enhancers can be quite redundant, and not all play an active role in regulating a cell's activity. How can we utilize other epigenomic data in order to identify true enhanceosome regions? One way is to cross this list against all ChIP data of the cell type. A true enhanceosome region should have made it through our filtering thus far, and contain several binding sites for known TFs. Co-binding is expected, and the list is sorted by the highest number of ChIP hits within the region.

```
epiRomics_putative_enhanceosome_fantom <-
  epiRomics_enhanceosome(epiRomics_putative_enhancers_filtered_fantom,
                         epiRomics_dB)
#> >> preparing features information...      2021-10-05 17:43:33
#> >> identifying nearest features...        2021-10-05 17:43:34
#> >> calculating distance from peak to TSS...   2021-10-05 17:43:35
#> >> assigning genomic annotation...         2021-10-05 17:43:35
#> >> adding gene annotation...          2021-10-05 17:44:15
#> 'select()' returned 1:many mapping between keys and columns
#> >> assigning chromosome lengths           2021-10-05 17:44:15
#> >> done...                    2021-10-05 17:44:15

## Taking a look, we see the top candidates meet the criteria we list as expected

epiRomics_putative_enhanceosome_fantom@annotations
#> GRanges object with 2749 ranges and 19 metadata columns:
#>        seqnames              ranges strand |     foxa2      mafb    nkx2_2
#>           <Rle>           <IRanges>  <Rle> | <integer> <integer> <integer>
#>    183     chr1 154418514-154419684      * |         2         2         1
#>   1096     chr9     2242369-2242873      * |         2         1         1
```

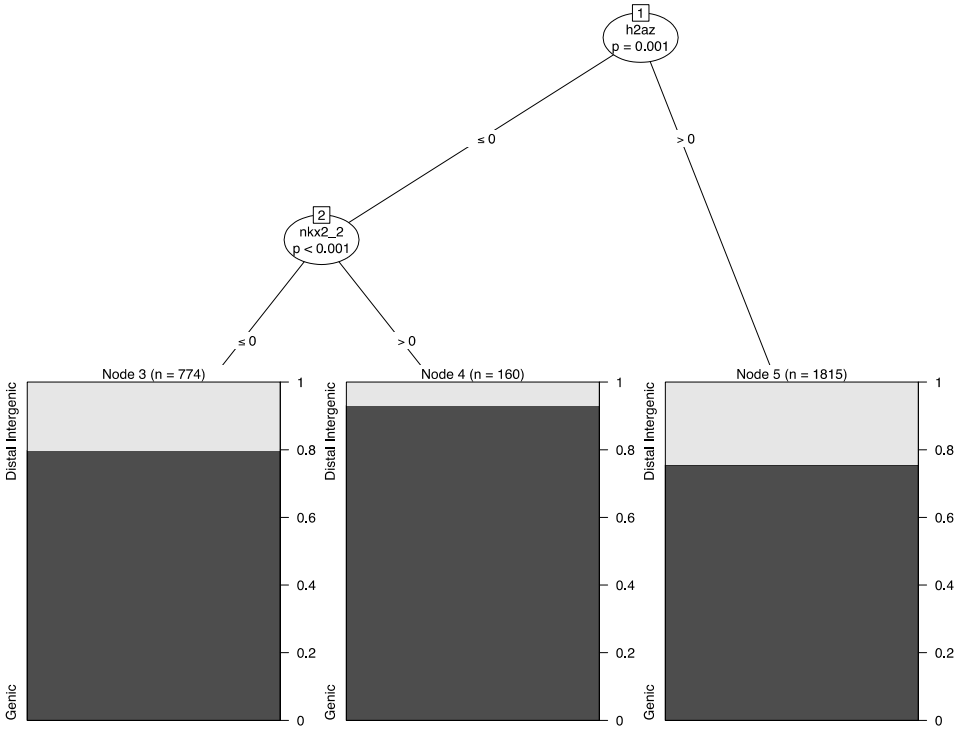
```
#>   2615     chr22   30310745-30311570      * |         2         1         2
#>     34      chr1   10685395-10688670      * |         1         0         1
#>    792      chr6   30748438-30749427      * |         2         1         1
#>    ...       ...                 ...    ... .       ...       ...       ...
#>   2743      chrX 153927339-153927701      * |         0         0         0
#>   2745      chrX 154369950-154370183      * |         0         0         0
#>   2746      chrX 154371971-154372237      * |         0         0         0
#>   2747      chrX 154372350-154372695      * |         0         0         0
#>   2748      chrX 154517139-154517596      * |         0         0         0
#>          nxk6_1      pdx1      h2az ChIP_Hits             annotation   geneChr
#>       <integer> <integer> <integer> <numeric>            <character> <integer>
#>    183         2         1         2        10 Intron (ENST00000622..         1
#>   1096         1         2         1         8      Distal Intergenic         9
#>   2615         1         1         1         8       Promoter (2-3kb)        22
#>     34         1         2         2         7 Intron (ENST00000377..         1
#>    792         1         1         1         7 Intron (ENST00000656..         6
#>    ...       ...       ...       ...       ...                    ...       ...
#>   2743         0         0         0         0       Promoter (<=1kb)        23
#>   2745         0         0         0         0       Promoter (1-2kb)        23
#>   2746         0         0         0         0       Promoter (<=1kb)        23
#>   2747         0         0         0         0       Promoter (1-2kb)        23
#>   2748         0         0         0         0       Promoter (<=1kb)        23
#>        geneStart   geneEnd geneLength geneStrand      geneId      transcriptId
#>        <integer> <integer>  <integer>  <integer> <character>       <character>
#>    183 154429343 154449979      20637          1        3570 ENST00000476006.5
#>   1096   2181571   2186183       4613          1        6595 ENST00000635392.1
#>   2615  30292008  30307890      15883          2       83874 ENST00000403362.5
#>     34  10660737  10693912      33176          2       54897 ENST00000478728.2
#>    792  30743199  30744547       1349          2        8870 ENST00000259874.6
#>    ...       ...       ...        ...        ...         ...                ...
#>   2743 153920715 153926860       6146          2         393 ENST00000422091.1
#>   2745 154348524 154371203      22680          2        2316 ENST00000420627.5
#>   2746 154348529 154371283      22755          2        2316 ENST00000422373.6
#>   2747 154348529 154371283      22755          2        2316 ENST00000422373.6
#>   2748 154506204 154516242      10039          2       60343 ENST00000434658.6
#>        distanceToTSS         ENSEMBL      SYMBOL             GENENAME
#>            <numeric>     <character> <character>            <character>
#>    183         -9659 ENSG00000160712        IL6R interleukin 6 receptor
#>   1096         60798 ENSG00000080503     SMARCA2 SWI/SNF related, mat..
#>   2615         -2855 ENSG00000099992    TBC1D10A TBC1 domain family m..
#>     34          5242 ENSG00000130940       CASZ1   castor zinc finger 1
#>    792         -3891 ENSG00000137331        IER3 immediate early resp..
#>    ...           ...                ...         ...                 ...
#>   2743          -479 ENSG00000089820     ARHGAP4 Rho GTPase activatin..
#>   2745          1020 ENSG00000196924        FLNA              filamin A
#>   2746          -688 ENSG00000196924        FLNA              filamin A
#>   2747         -1067 ENSG00000196924        FLNA              filamin A
#>   2748          -897 ENSG00000071889       FAM3A FAM3 metabolism regu..
#>   -------
#>   seqinfo: 595 sequences (1 circular) from hg38 genome

## Evaluate calls on chromosome 1
```

```
head(as.data.frame(
  epiRomics_putative_enhanceosome_fantom@annotations
  )[as.data.frame(epiRomics_putative_enhanceosome_fantom@annotations)$seqnames
    == "chr1",])
#>     seqnames     start       end width strand foxa2 mafb nkx2_2 nxk6_1 pdx1
#> 183     chr1 154418514 154419684  1171      *     2    2      1      2    1
#> 34      chr1  10685395  10688670  3276      *     1    0      1      1    2
#> 24      chr1   8170112   8170857   746      *     1    1      1      1    1
#> 67      chr1  21638834  21639978  1145      *     1    1      1      1    1
#> 71      chr1  22414515  22414806   292      *     1    1      1      1    1
#> 256     chr1 205318810 205319660   851      *     1    1      2      1    0
#>     h2az ChIP_Hits                                          annotation geneChr
#> 183    2        10   Intron (ENST00000622330.4/3570, intron 1 of 6)       1
#> 34     2         7 Intron (ENST00000377022.8/54897, intron 4 of 20)       1
#> 24     1         6                                Distal Intergenic       1
#> 67     1         6  Intron (ENST00000290101.8/5909, intron 2 of 26)       1
#> 71     1         6                                Distal Intergenic       1
#> 256    1         6                                 Promoter (2-3kb)       1
#>     geneStart   geneEnd geneLength geneStrand    geneId      transcriptId
#> 183 154429343 154449979      20637          1      3570 ENST00000476006.5
#> 34   10660737  10693912      33176          2     54897 ENST00000478728.2
#> 24    8201518   8215207      13690          1 102724539 ENST00000670361.1
#> 67   21596221  21651820      55600          2      5909 ENST00000471600.6
#> 71   22428838  22511763      82926          1      9923 ENST00000650433.1
#> 256 205302063 205321745      19683          2     81788 ENST00000367157.6
#>     distanceToTSS         ENSEMBL    SYMBOL
#> 183         -9659 ENSG00000160712      IL6R
#> 34           5242 ENSG00000130940     CASZ1
#> 24         -30661 ENSG00000227634 LINC01714
#> 67          11842 ENSG00000076864   RAP1GAP
#> 71         -14032 ENSG00000184677    ZBTB40
#> 256          2085 ENSG00000163545     NUAK2
#>                                     GENENAME
#> 183                     interleukin 6 receptor
#> 34                        castor zinc finger 1
#> 24  long intergenic non-protein coding RNA 1714
#> 67               RAP1 GTPase activating protein
#> 71     zinc finger and BTB domain containing 40
#> 256                       NUAK family kinase 2

## Find Index

which(names(epiRomics_putative_enhanceosome_fantom@annotations) == 183)
#> [1] 1
```

Transcription factor decision trees

ChIP dataset repositories are quite sizeable for many organisms and cell types, with the expectation to only grow larger. Many different TFs binding to a putative enhancer region may not be that meaningful in the context of your biological question. A further step would be to ask whether there are co-TFs that pop up together, and whether this pattern varies across the functional annotation of the genome, i.e. does the combination of two TFs on enhanceosomes change on the gene body compared to distal intergenic regions?

```
plot(epiRomics_predictors(epiRomics_putative_enhanceosome_fantom))
```

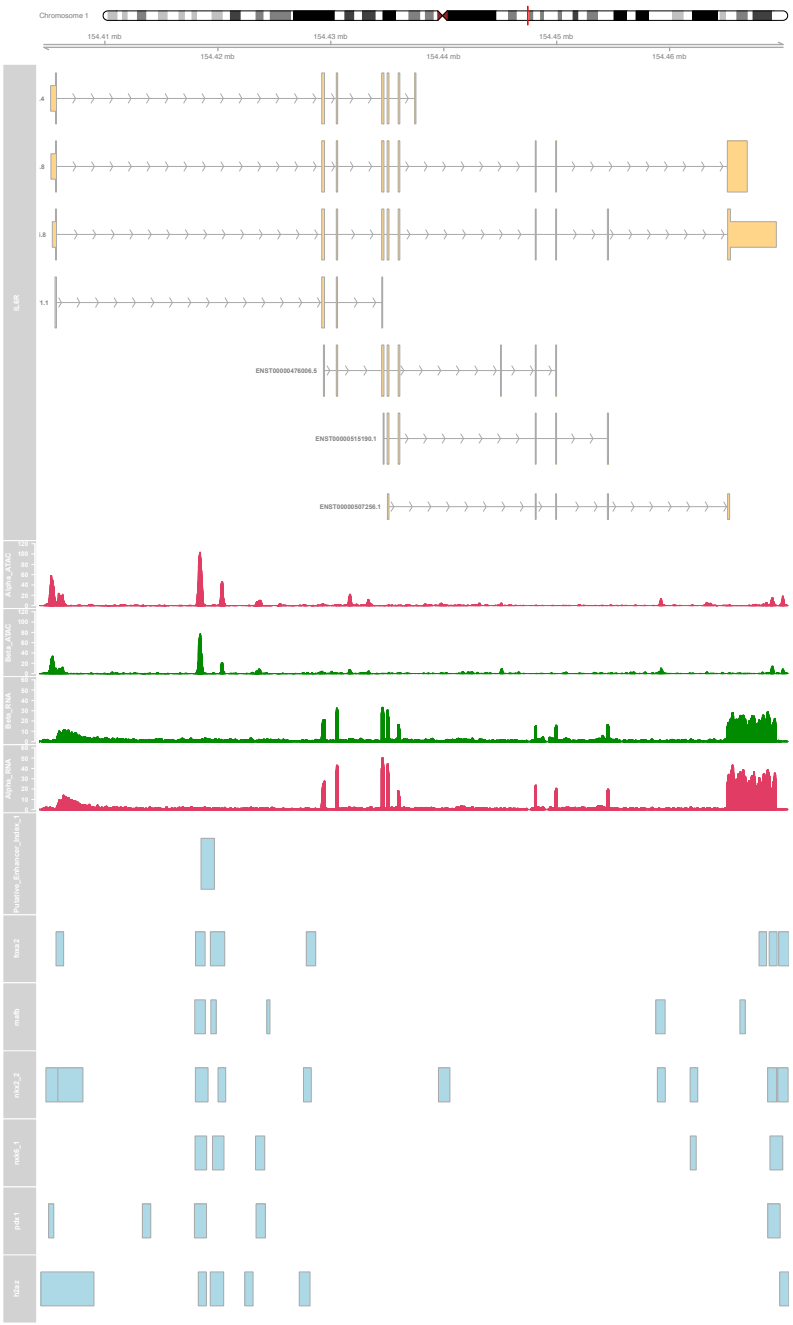


Intersecting and visualizing ATAC- and RNA-Seq data

What if you wanted to visualize co-binding on your FANTOM filtered putative enhancer region? And do you have additional data you want to include for visualization, such as ATAC and RNA Seq? Lets take a look at one of the top hits

```
## Read in ATAC Seq and RNA Seq track bigwigs

## NOTE: These bigwigs are subsetted to chromosome 1.
## Indices not falling on chromosome 1 will return an error.

epiRomics_track_connection <- read.csv(
  system.file(
    "extdata",
    "example_epiRomics_BW_sheet_user_paths.csv",
    package = "epiRomics"
  )
)
```

```
epiRomics_track_layer_human(
  epiRomics_putative_enhanceosome_fantom,
  epiRomics_index =
    which(
      names(epiRomics_putative_enhanceosome_fantom@annotations)
      == 183
    ),
  epiRomics_dB = epiRomics_dB,
  epiRomics_track_connection =
    epiRomics_track_connection
)
#> [1] "not empty"
#> [1] 103.678
#> [1] "not empty"
#> [1] 77.5726
#> [1] "not empty"
#> [1] 33.2945
#> [1] "not empty"
#> [1] 50.4959
```


Chromosome 1
154.41 mb
154.42 mb
154.43 mb
154.44 mb
154.45 mb
154.46 mb
IL6R
ENST00000476006.5
ENST00000515190.1
ENST00000507266.1
Alpha_ATAC
Beta_ATAC
Beta_RNA
Alpha_RNA
Primitive_Enhancer_Index_1
foxa2
mdb
nkx2_2
nr4a_1
pdx1
hoxz

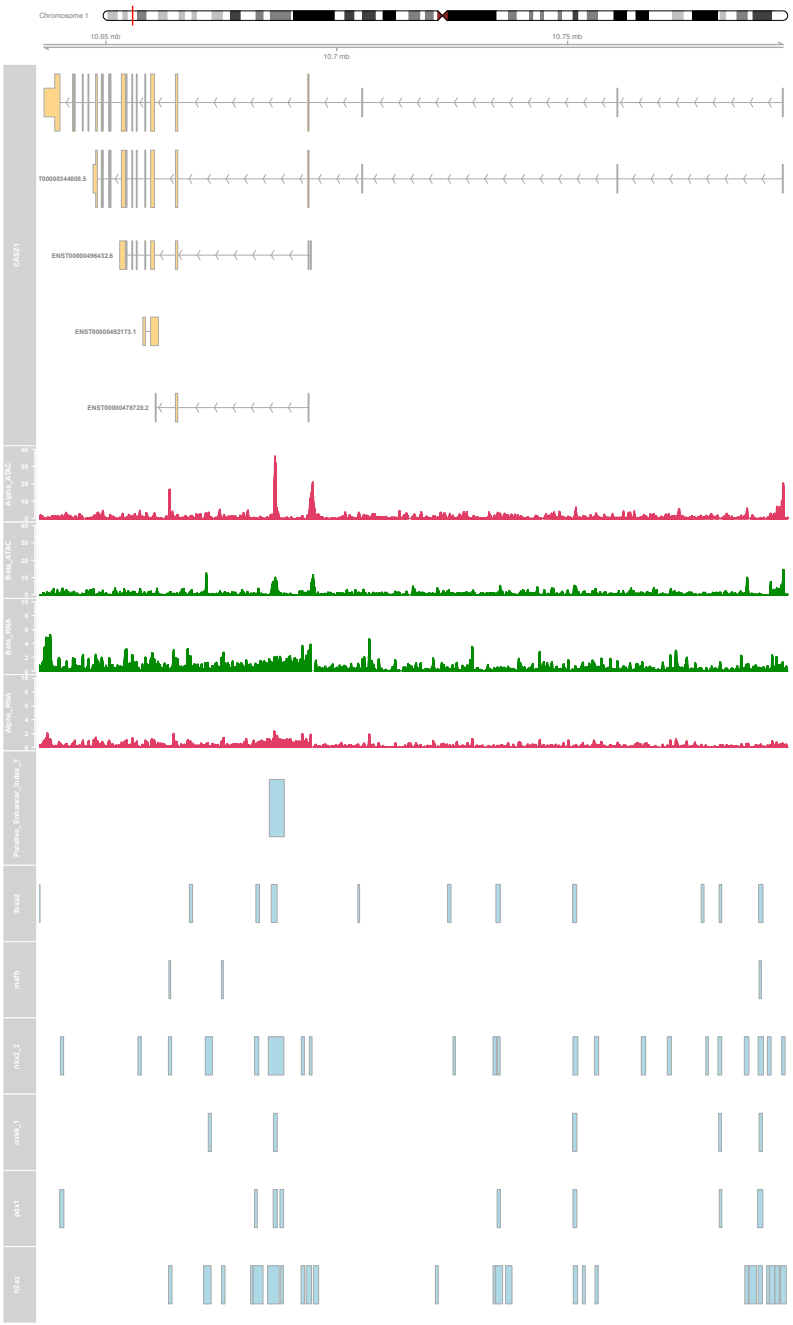
What about a region that overlapped with active enhancers from the human islet regulome database?

```
epiRomics_putative_enhanceosome_regulome_active <-
  epiRomics_enhanceosome(epiRomics_putative_enhancers_filtered_regulome_active,
                         epiRomics_dB)
#> >> preparing features information...      2021-10-05 17:46:34
#> >> identifying nearest features...        2021-10-05 17:46:34
#> >> calculating distance from peak to TSS...   2021-10-05 17:46:36
#> >> assigning genomic annotation...         2021-10-05 17:46:36
#> >> adding gene annotation...           2021-10-05 17:46:45
#> 'select()' returned 1:many mapping between keys and columns
#> >> assigning chromosome lengths           2021-10-05 17:46:45
#> >> done...                     2021-10-05 17:46:45

epiRomics_putative_enhanceosome_regulome_active@annotations
#> GRanges object with 6025 ranges and 19 metadata columns:
#>        seqnames              ranges strand |     foxa2      mafb    nkx2_2
#>           <Rle>           <IRanges>  <Rle> | <integer> <integer> <integer>
#>    456     chr1 154418514-154419684      * |         2         2         1
#>   2082     chr7     1555599-1556082      * |         1         1         1
#>   2572     chr9     2242369-2242873      * |         2         1         1
#>   3421    chr11   65416576-65419753      * |         1         1         2
#>   4709    chr17     7887867-7889135      * |         1         0         2
#>    ...      ...                 ...    ... .       ...       ...       ...
#>   5999     chrX            49184194      * |         0         0         0
#>   6001     chrX   70478965-70479351      * |         0         0         0
#>   6006     chrX 107710676-107711066      * |         0         0         0
#>   6007     chrX 107711430-107711673      * |         0         0         0
#>   6020     chrX 150874055-150874330      * |         0         0         0
#>          nxk6_1      pdx1      h2az ChIP_Hits             annotation   geneChr
#>        <integer> <integer> <integer> <numeric>            <character> <integer>
#>    456         2         1         2        10 Intron (ENST00000622..         1
#>   2082         2         2         1         8       Promoter (<=1kb)         7
#>   2572         1         2         1         8      Distal Intergenic         9
#>   3421         1         1         2         8      Distal Intergenic        11
#>   4709         2         2         1         8       Promoter (<=1kb)        17
#>    ...       ...       ...       ...       ...                    ...       ...
#>   5999         0         0         0         0       Promoter (<=1kb)        23
#>   6001         0         0         0         0       Promoter (<=1kb)        23
#>   6006         0         0         0         0      Distal Intergenic        23
#>   6007         0         0         0         0      Distal Intergenic        23
#>   6020         0         0         0         0 Intron (ENST00000370..        23
#>        geneStart   geneEnd geneLength geneStrand      geneId       transcriptId
#>        <integer> <integer>  <integer>  <integer> <character>        <character>
#>    456 154429343 154449979      20637          1        3570  ENST00000476006.5
#>   2082   1550305   1556120       5816          2      202915  ENST00000441933.5
#>   2572   2181571   2186183       4613          1        6595  ENST00000635392.1
#>   3421  65422798  65445540      22743          1      283131  ENST00000501122.2
#>   4709   7888789   7912755      23967          1        1107 ENST00000330494.12
#>    ...       ...       ...        ...        ...         ...                ...
#>   5999  49175621  49184789       9169          2        4007  ENST00000453382.5
#>   6001  70479118  70499903      20786          1        1741  ENST00000466140.5
#>   6006 107714677 107716401       1725          2        1831  ENST00000486554.1
```

```
#>   6007 107714677 107716401       1725          2        1831  ENST00000486554.1
#>   6020 150814900 150898609      83710          2       83692  ENST00000491877.1
#>        distanceToTSS         ENSEMBL      SYMBOL              GENENAME
#>            <numeric>     <character> <character>           <character>
#>    456         -9659 ENSG00000160712        IL6R interleukin 6 receptor
#>   2082            38 ENSG00000164855    TMEM184A transmembrane protei..
#>   2572         60798 ENSG00000080503     SMARCA2 SWI/SNF related, mat..
#>   3421         -3045 ENSG00000245532       NEAT1 nuclear paraspeckle ..
#>   4709             0 ENSG00000170004        CHD3 chromodomain helicas..
#>    ...           ...             ...         ...                    ...
#>   5999           595 ENSG00000012211    PRICKLE3 prickle planar cell ..
#>   6001             0 ENSG00000082458        DLG3 discs large MAGUK sc..
#>   6006          5335 ENSG00000157514     TSC22D3 TSC22 domain family ..
#>   6007          4728 ENSG00000157514     TSC22D3 TSC22 domain family ..
#>   6020         24279 ENSG00000102181      CD99L2   CD99 molecule like 2
#>   -------
#>   seqinfo: 595 sequences (1 circular) from hg38 genome

## Evaluate calls on chromosome 1

head(as.data.frame(
  epiRomics_putative_enhanceosome_regulome_active@annotations
)[as.data.frame(epiRomics_putative_enhanceosome_regulome_active@annotations)$seqnames
  == "chr1",])
#>     seqnames     start       end width strand foxa2 mafb nkx2_2 nxk6_1 pdx1
#> 456     chr1 154418514 154419684  1171      *     2    2      1      2    1
#> 82      chr1  10685395  10688670  3276      *     1    0      1      1    2
#> 46      chr1   7574092   7574479   388      *     1    1      1      1    1
#> 47      chr1   7574640   7575094   455      *     1    1      1      1    1
#> 49      chr1   8169274   8169689   416      *     1    1      1      1    1
#> 50      chr1   8170112   8170857   746      *     1    1      1      1    1
#>     h2az ChIP_Hits                                          annotation geneChr
#> 456    2        10    Intron (ENST00000622330.4/3570, intron 1 of 6)       1
#> 82     2         7  Intron (ENST00000377022.8/54897, intron 4 of 20)       1
#> 46     1         6 Intron (ENST00000303635.12/23261, intron 6 of 22)       1
#> 47     1         6 Intron (ENST00000303635.12/23261, intron 6 of 22)       1
#> 49     1         6                                Distal Intergenic       1
#> 50     1         6                                Distal Intergenic       1
#>     geneStart   geneEnd geneLength geneStrand    geneId      transcriptId
#> 456 154429343 154449979      20637          1      3570 ENST00000476006.5
#> 82   10660737  10693912      33176          2     54897 ENST00000478728.2
#> 46    7736408   7767856      31449          1     23261 ENST00000495233.5
#> 47    7736408   7767856      31449          1     23261 ENST00000495233.5
#> 49    8201518   8215207      13690          1 102724539 ENST00000670361.1
#> 50    8201518   8215207      13690          1 102724539 ENST00000670361.1
#>     distanceToTSS         ENSEMBL    SYMBOL
#> 456         -9659 ENSG00000160712      IL6R
#> 82           5242 ENSG00000130940     CASZ1
#> 46        -161929 ENSG00000171735    CAMTA1
#> 47        -161314 ENSG00000171735    CAMTA1
#> 49         -31829 ENSG00000227634 LINC01714
#> 50         -30661 ENSG00000227634 LINC01714
#>                                    GENENAME
```

```
#> 456                       interleukin 6 receptor
#> 82                          castor zinc finger 1
#> 46  calmodulin binding transcription activator 1
#> 47  calmodulin binding transcription activator 1
#> 49   long intergenic non-protein coding RNA 1714
#> 50   long intergenic non-protein coding RNA 1714

## Find Index

which(names(epiRomics_putative_enhanceosome_regulome_active@annotations) == 82)
#> [1] 7
```

```
epiRomics_track_layer_human(
  epiRomics_putative_enhanceosome_regulome_active,
  epiRomics_index = which(
    names(
      epiRomics_putative_enhanceosome_regulome_active@annotations
    ) == 82
  ),
  epiRomics_dB = epiRomics_dB,
  epiRomics_track_connection = epiRomics_track_connection
)
#> [1] "not empty"
#> [1] 36.1883
#> [1] "not empty"
#> [1] 14.7321
#> [1] "not empty"
#> [1] 5.30505
#> [1] "not empty"
#> [1] 2.40628
```


Chromosome 1
10.65 mb
10.75 mb
10.7 mb
CASZ1
T00000344008.5
ENST00000496432.6
ENST00000492173.1
ENST00000478728.2
Alpha_ATAC
Beta_ATAC
Beta_RNA
Alpha_RNA
Putative_Enhancer_Index_7
fos2
mafb
nrx2_2
nr4l_1
pdx1
hlx.a

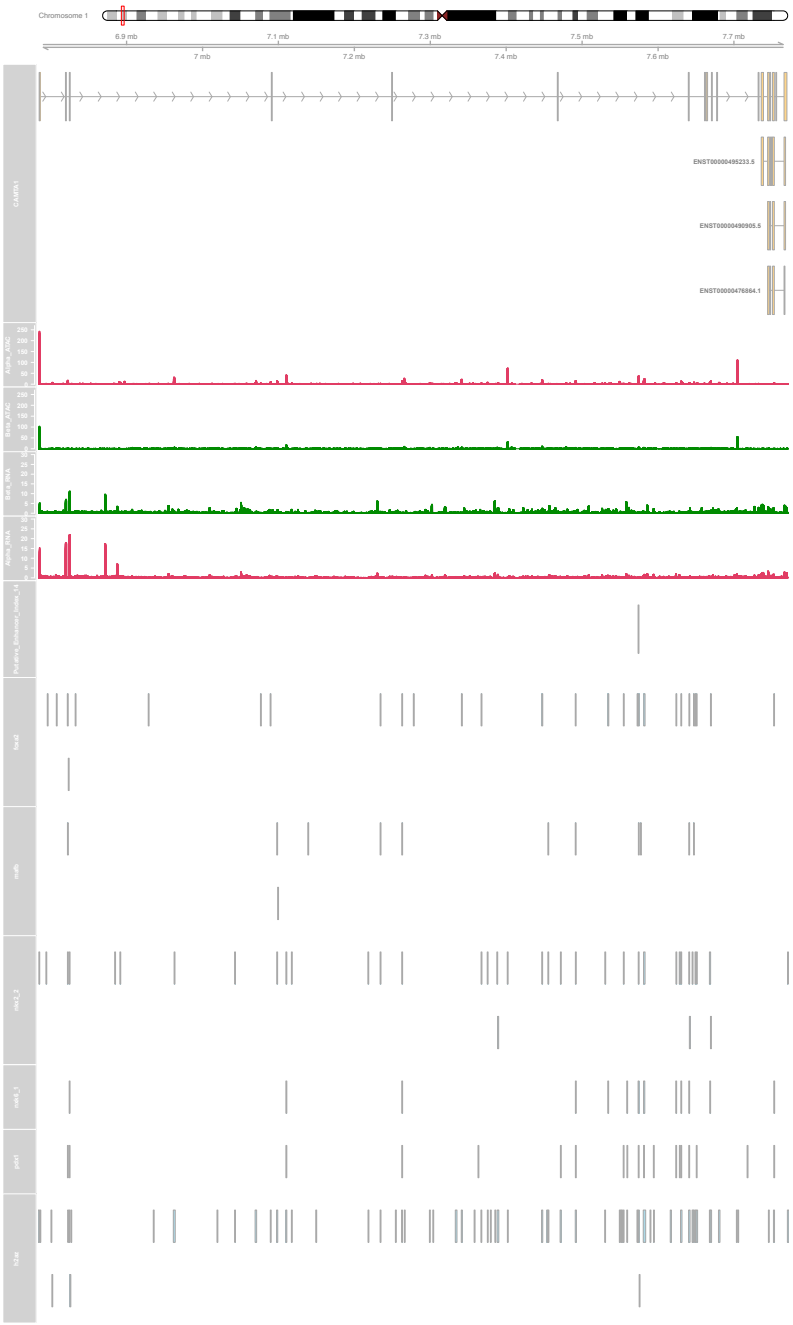
What about a region that overlapped with super enhancers from the human islet regulome database?

```
epiRomics_putative_enhanceosome_regulome_super <-
  epiRomics_enhanceosome(epiRomics_putative_enhancers_filtered_regulome_super,
                         epiRomics_dB)
#> >> preparing features information...      2021-10-05 17:49:52
#> >> identifying nearest features...        2021-10-05 17:49:52
#> >> calculating distance from peak to TSS...   2021-10-05 17:49:53
#> >> assigning genomic annotation...         2021-10-05 17:49:53
#> >> adding gene annotation...           2021-10-05 17:50:00
#> 'select()' returned 1:many mapping between keys and columns
#> >> assigning chromosome lengths           2021-10-05 17:50:01
#> >> done...                     2021-10-05 17:50:01

epiRomics_putative_enhanceosome_regulome_super@annotations
#> GRanges object with 2401 ranges and 19 metadata columns:
#>        seqnames              ranges strand |     foxa2      mafb    nkx2_2
#>           <Rle>           <IRanges>  <Rle> | <integer> <integer> <integer>
#>    166     chr1 154418514-154419684      * |         2         2         1
#>    966     chr9     2242369-2242873      * |         2         1         1
#>   1422    chr11   65416576-65419753      * |         1         1         2
#>     16     chr1   10685395-10688670      * |         1         0         1
#>    764     chr6   30748438-30749427      * |         2         1         1
#>    ...      ...                 ...    ... .       ...       ...       ...
#>   2393    chr22   46103687-46105878      * |         0         0         0
#>   2396    chr22   46108912-46109315      * |         0         0         0
#>   2397    chr22   46109916-46110442      * |         0         0         0
#>   2398    chr22   46115774-46116154      * |         0         0         0
#>   2399    chr22   46116326-46116501      * |         0         0         0
#>          nxk6_1      pdx1      h2az ChIP_Hits             annotation   geneChr
#>        <integer> <integer> <integer> <numeric>            <character> <integer>
#>    166         2         1         2        10 Intron (ENST00000622..         1
#>    966         1         2         1         8      Distal Intergenic         9
#>   1422         1         1         2         8      Distal Intergenic        11
#>     16         1         2         2         7 Intron (ENST00000377..         1
#>    764         1         1         1         7 Intron (ENST00000656..         6
#>    ...       ...       ...       ...       ...                    ...       ...
#>   2393         0         0         0         0 Exon (ENST0000038105..        22
#>   2396         0         0         0         0 Exon (ENST0000043543..        22
#>   2397         0         0         0         0       Promoter (2-3kb)        22
#>   2398         0         0         0         0       Promoter (2-3kb)        22
#>   2399         0         0         0         0       Promoter (2-3kb)        22
#>        geneStart   geneEnd geneLength geneStrand      geneId      transcriptId
#>        <integer> <integer>  <integer>  <integer> <character>       <character>
#>    166 154429343 154449979      20637          1        3570 ENST00000476006.5
#>    966   2181571   2186183       4613          1        6595 ENST00000635392.1
#>   1422  65422798  65445540      22743          1      283131 ENST00000501122.2
#>     16  10660737  10693912      33176          2       54897 ENST00000478728.2
#>    764  30743199  30744547       1349          2        8870 ENST00000259874.6
#>    ...       ...       ...        ...        ...         ...                ...
#>   2393  46112749  46112822         74          1      406883 ENST00000362116.3
#>   2396  46112749  46112822         74          1      406883 ENST00000362116.3
#>   2397  46112749  46112822         74          1      406883 ENST00000362116.3
```

```
#>   2398  46113686  46113768         83          1      406884 ENST00000385140.1
#>   2399  46113686  46113768         83          1      406884 ENST00000385140.1
#>        distanceToTSS         ENSEMBL      SYMBOL              GENENAME
#>            <numeric>     <character> <character>           <character>
#>    166         -9659 ENSG00000160712        IL6R interleukin 6 receptor
#>    966         60798 ENSG00000080503     SMARCA2 SWI/SNF related, mat..
#>   1422         -3045 ENSG00000245532       NEAT1 nuclear paraspeckle ..
#>     16          5242 ENSG00000130940       CASZ1   castor zinc finger 1
#>    764         -3891 ENSG00000137331        IER3 immediate early resp..
#>    ...           ...             ...         ...                    ...
#>   2393         -6871 ENSG00000283990   MIRLET7A3      microRNA let-7a-3
#>   2396         -3434 ENSG00000283990   MIRLET7A3      microRNA let-7a-3
#>   2397         -2307 ENSG00000283990   MIRLET7A3      microRNA let-7a-3
#>   2398          2088 ENSG00000284520    MIRLET7B        microRNA let-7b
#>   2399          2640 ENSG00000284520    MIRLET7B        microRNA let-7b
#>   -------
#>   seqinfo: 595 sequences (1 circular) from hg38 genome

## Evaluate calls on chromosome 1

head(as.data.frame(
  epiRomics_putative_enhanceosome_regulome_super@annotations)[as.data.frame(
    epiRomics_putative_enhanceosome_regulome_super@annotations)$seqnames == "chr1",])
#>     seqnames     start       end width strand foxa2 mafb nkx2_2 nxk6_1 pdx1
#> 166     chr1 154418514 154419684  1171      *     2    2      1      2    1
#> 16      chr1  10685395  10688670  3276      *     1    0      1      1    2
#> 1       chr1   7574092   7574479   388      *     1    1      1      1    1
#> 2       chr1   7574640   7575094   455      *     1    1      1      1    1
#> 3       chr1   8169274   8169689   416      *     1    1      1      1    1
#> 4       chr1   8170112   8170857   746      *     1    1      1      1    1
#>     h2az ChIP_Hits                                          annotation geneChr
#> 166    2        10    Intron (ENST00000622330.4/3570, intron 1 of 6)       1
#> 16     2         7  Intron (ENST00000377022.8/54897, intron 4 of 20)       1
#> 1      1         6 Intron (ENST00000303635.12/23261, intron 6 of 22)       1
#> 2      1         6 Intron (ENST00000303635.12/23261, intron 6 of 22)       1
#> 3      1         6                                  Distal Intergenic       1
#> 4      1         6                                  Distal Intergenic       1
#>     geneStart   geneEnd geneLength geneStrand    geneId      transcriptId
#> 166 154429343 154449979      20637          1      3570 ENST00000476006.5
#> 16   10660737  10693912      33176          2     54897 ENST00000478728.2
#> 1     7736408   7767856      31449          1     23261 ENST00000495233.5
#> 2     7736408   7767856      31449          1     23261 ENST00000495233.5
#> 3     8201518   8215207      13690          1 102724539 ENST00000670361.1
#> 4     8201518   8215207      13690          1 102724539 ENST00000670361.1
#>     distanceToTSS         ENSEMBL    SYMBOL
#> 166         -9659 ENSG00000160712      IL6R
#> 16           5242 ENSG00000130940     CASZ1
#> 1         -161929 ENSG00000171735    CAMTA1
#> 2         -161314 ENSG00000171735    CAMTA1
#> 3          -31829 ENSG00000227634 LINC01714
#> 4          -30661 ENSG00000227634 LINC01714
#>                                     GENENAME
#> 166                     interleukin 6 receptor
```

```
#> 16                          castor zinc finger 1
#> 1   calmodulin binding transcription activator 1
#> 2   calmodulin binding transcription activator 1
#> 3    long intergenic non-protein coding RNA 1714
#> 4    long intergenic non-protein coding RNA 1714

## Find Index

which(names(epiRomics_putative_enhanceosome_regulome_super@annotations) == 1)
#> [1] 14
```

```
epiRomics_track_layer_human(
  epiRomics_putative_enhanceosome_regulome_super,
  epiRomics_index = which(
    names(epiRomics_putative_enhanceosome_regulome_super@annotations) == 1
  ),
  epiRomics_dB = epiRomics_dB,
  epiRomics_track_connection = epiRomics_track_connection
)
#> [1] "not empty"
#> [1] 243.743
#> [1] "not empty"
#> [1] 103.706
#> [1] "not empty"
#> [1] 11.3213
#> [1] "not empty"
#> [1] 22.1498
```

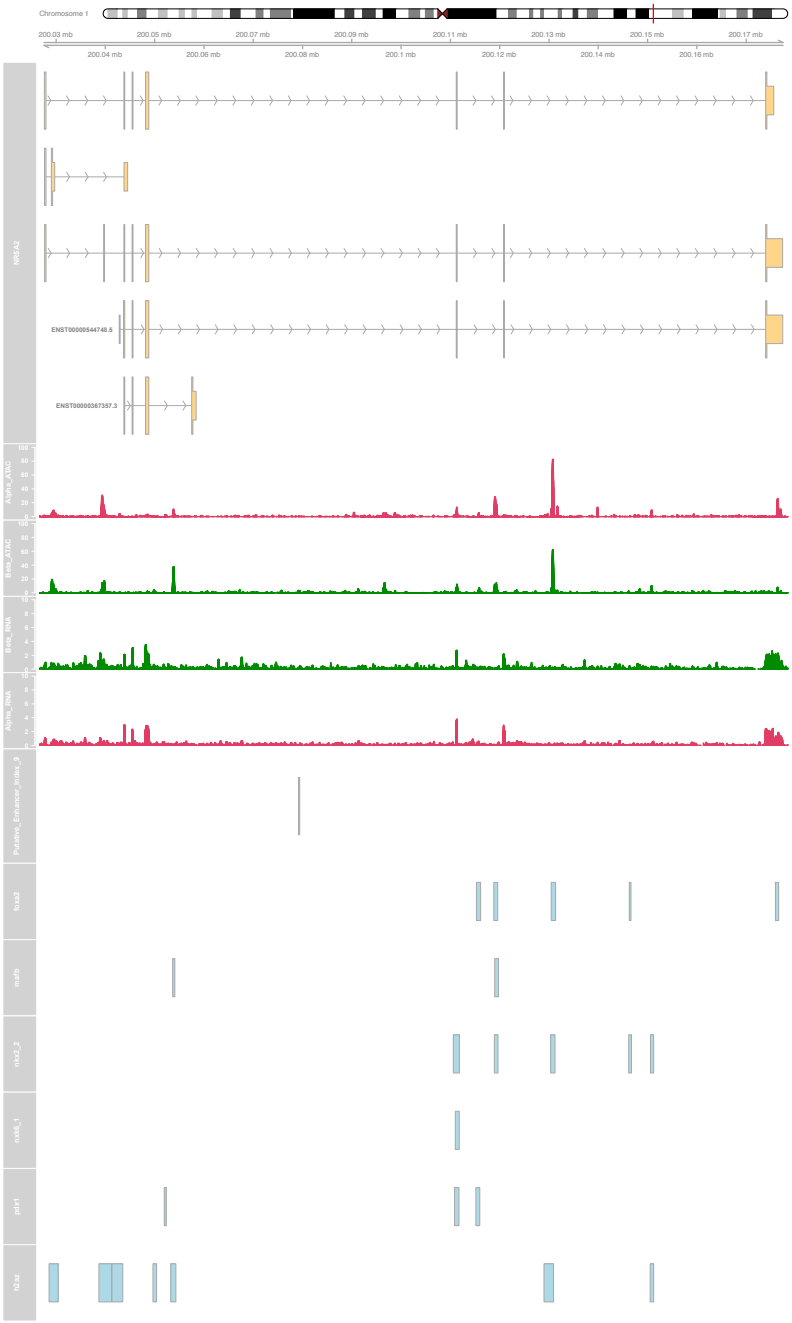

Chromosome 1
6.9 mb
7.1 mb
7.3 mb
7.5 mb
7.7 mb
7 mb
7.2 mb
7.4 mb
7.6 mb
CAMTA1
ENST00000495233.5
ENST00000490905.5
ENST00000476864.1
Alpha_ATAC
Beta_ATAC
Beta_RNA
Alpha_RNA
Promoters_Enhancer_Tracks_v4
head
midb
mid2_3
mid1_1
pon1
hind

Or, about a region that overlapped with ultra-conserved non coding elements?

```
epiRomics_putative_enhanceosome_ucnes <-
  epiRomics_enhanceosome(epiRomics_putative_enhancers_filtered_ucnes, epiRomics_dB)
#> >> preparing features information...      2021-10-05 17:52:00
#> >> identifying nearest features...        2021-10-05 17:52:00
#> >> calculating distance from peak to TSS...   2021-10-05 17:52:01
#> >> assigning genomic annotation...         2021-10-05 17:52:01
#> >> adding gene annotation...           2021-10-05 17:52:08
#> 'select()' returned 1:1 mapping between keys and columns
#> >> assigning chromosome lengths           2021-10-05 17:52:08
#> >> done...                     2021-10-05 17:52:08

epiRomics_putative_enhanceosome_ucnes@annotations
#> GRanges object with 11 ranges and 19 metadata columns:
#>      seqnames              ranges strand |     foxa2      mafb    nkx2_2
#>         <Rle>           <IRanges>  <Rle> | <integer> <integer> <integer>
#>    7     chr9 106921420-106921764      * |         1         1         1
#>    1     chr1 164635220-164635921      * |         0         0         2
#>    6     chr3   71131859-71132164      * |         1         0         0
#>    2     chr1 164711914-164712296      * |         0         0         0
#>    3     chr1 164712350-164713071      * |         0         0         0
#>    8    chr11 114163425-114164860      * |         0         0         0
#>    9    chr15   36903894-36904085      * |         0         0         1
#>   11    chr21   16534340-16534665      * |         0         0         0
#>    4     chr1 200079185-200079426      * |         0         0         0
#>    5     chr1 213585694-213586385      * |         0         0         0
#>   10    chr15   53447393-53447809      * |         0         0         0
#>         nxk6_1      pdx1      h2az ChIP_Hits             annotation   geneChr
#>      <integer> <integer> <integer> <numeric>            <character> <integer>
#>    7         1         0         1         5 Intron (ENST00000472..         9
#>    1         0         1         1         4 Intron (ENST00000420..         1
#>    6         0         0         1         2       Promoter (<=1kb)         3
#>    2         0         0         1         1 Intron (ENST00000420..         1
#>    3         0         0         1         1 Intron (ENST00000420..         1
#>    8         0         0         1         1 Intron (ENST00000335..        11
#>    9         0         0         0         1       Promoter (<=1kb)        15
#>   11         0         0         1         1       Promoter (<=1kb)        21
#>    4         0         0         0         0 Intron (ENST00000236..         1
#>    5         0         0         0         0      Distal Intergenic         1
#>   10         0         0         0         0 Intron (ENST00000662..        15
#>      geneStart   geneEnd geneLength geneStrand      geneId      transcriptId
#>      <integer> <integer>  <integer>  <integer> <character>       <character>
#>    7 106926925 106932462       5538          1       58499 ENST00000480607.5
#>    1 164630981 164799889     168909          1        5087 ENST00000482110.5
#>    6  70959237  71132099     172863          2       27086 ENST00000650188.1
#>    2 164772912 164807571      34660          1        5087 ENST00000558837.5
#>    3 164772912 164807571      34660          1        5087 ENST00000558837.5
#>    8 114180766 114247296      66531          1        7704 ENST00000545851.5
#>    9  36894784  36904067       9284          2        4212 ENST00000559408.1
#>   11  16534952  16607137      72186          1      388815 ENST00000654245.1
#>    4 200043810 200058424      14615          1        2494 ENST00000367357.3
#>    5 213832591 213841041       8451          2   100505832 ENST00000609394.5
```

```
#>   10  53513742  53541080      27339          2      256764 ENST00000614174.4
#>      distanceToTSS         ENSEMBL      SYMBOL              GENENAME
#>          <numeric>     <character> <character>           <character>
#>    7         -5161 ENSG00000148143      ZNF462 zinc finger protein ..
#>    1          4239 ENSG00000185630        PBX1         PBX homeobox 1
#>    6             0 ENSG00000114861       FOXP1        forkhead box P1
#>    2        -60616 ENSG00000185630        PBX1         PBX homeobox 1
#>    3        -59841 ENSG00000185630        PBX1         PBX homeobox 1
#>    8        -15906 ENSG00000109906      ZBTB16 zinc finger and BTB ..
#>    9             0 ENSG00000134138       MEIS2        Meis homeobox 2
#>   11          -287 ENSG00000215386    MIR99AHG mir-99a-let-7c clust..
#>    4         35375 ENSG00000116833       NR5A2 nuclear receptor sub..
#>    5        254656 ENSG00000230461   PROX1-AS1  PROX1 antisense RNA 1
#>   10         93271 ENSG00000166415       WDR72    WD repeat domain 72
#>   -------
#>   seqinfo: 595 sequences (1 circular) from hg38 genome
```

```
epiRomics_track_layer_human(
  epiRomics_putative_enhanceosome_ucnes,
  epiRomics_index = 9,
  epiRomics_dB = epiRomics_dB,
  epiRomics_track_connection = epiRomics_track_connection
)
#> [1] "not empty"
#> [1] 82.795
#> [1] "not empty"
#> [1] 62.847
#> [1] "not empty"
#> [1] 3.55117
#> [1] "not empty"
#> [1] 3.80884
```

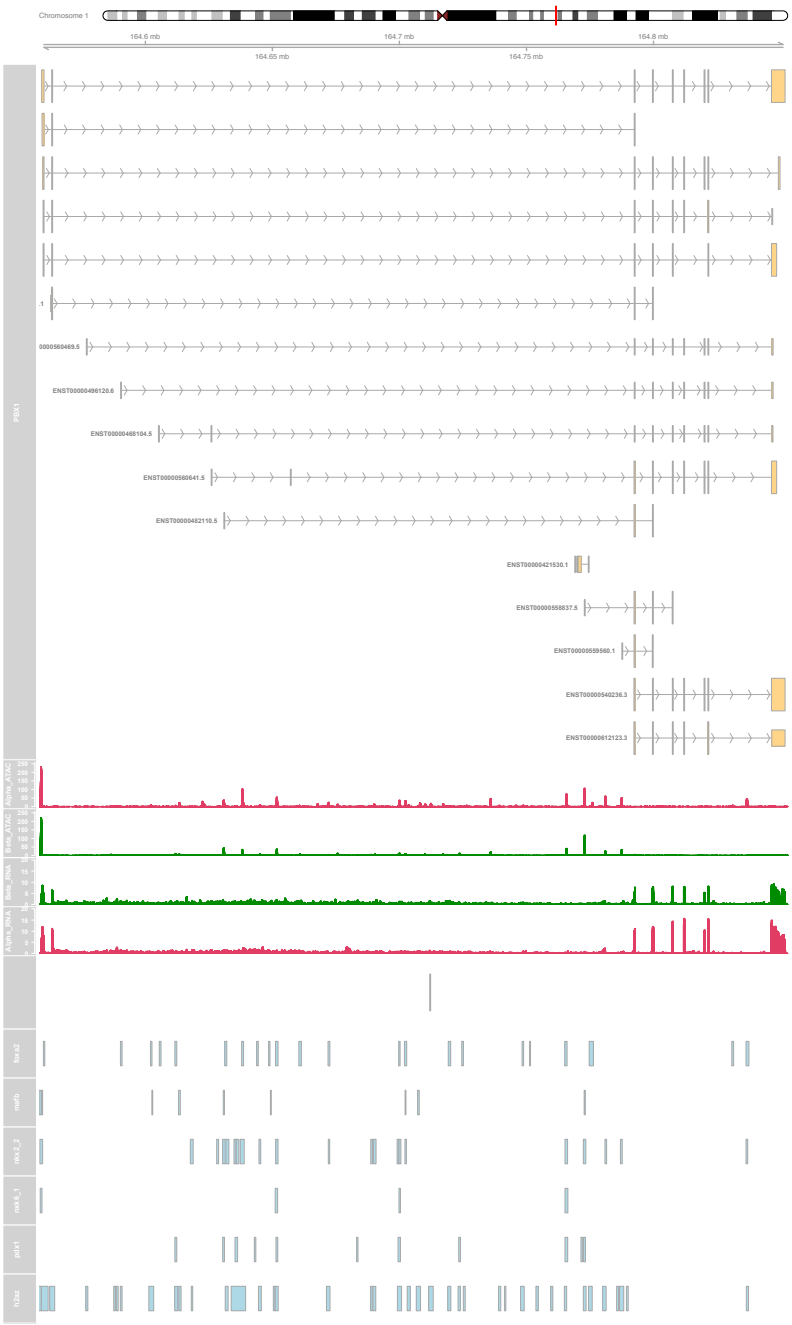

Chromosome 1
200.03 mb
200.05 mb
200.07 mb
200.09 mb
200.11 mb
200.13 mb
200.15 mb
200.17 mb
200.04 mb
200.06 mb
200.08 mb
200.1 mb
200.12 mb
200.14 mb
200.16 mb
NR3A2
ENST00000544748.5
ENST00000367357.3
Alpha_ATAC
Beta_ATAC
Beta_RNA
Alpha_RNA
Putative_Enhancer_Index_9
foxa2
mafb
nkx2_2
nkx6_1
pdx1
hlxb
23

How about applying multiple filters to further increase the confidence of calls?

```
epiRomics_putative_enhancers_filtered_stringent <-
  epiRomics_enhancers_filter(
    epiRomics_enhancers_filter(
      epiRomics_enhancers_filter(
        epiRomics_enhancers_filter(epiRomics_putative_enhancers, epiRomics_dB,
                                   epiRomics_type =
                                     "hg38_custom_fantom"),
        epiRomics_dB,
        epiRomics_type = "hg38_custom_regulome_active"
      ),
      epiRomics_dB,
      epiRomics_type = "hg38_custom_regulome_super"
    ),
    epiRomics_dB,
    epiRomics_type = "hg38_custom_ucnes"
  )


## Here, we see a highly conservative list of putative enhancer calls that overlap with four
## different functional annotations, suggesting the lowest hanging fruit for downstream
## bench-lab validation.
## NOTE: The UCNE database filter caused the greatest reduction in enhancer calls.

epiRomics_putative_enhancers_filtered_stringent@annotations
#> GRanges object with 2 ranges and 0 metadata columns:
#>       seqnames              ranges strand
#>          <Rle>           <IRanges>  <Rle>
#>   [1]     chr1 164711914-164712296      *
#>   [2]     chr1 164712350-164713071      *
#>   -------
#>   seqinfo: 595 sequences (1 circular) from hg38 genome

epiRomics_putative_enhanceosome_stringent <-
  epiRomics_enhanceosome(epiRomics_putative_enhancers_filtered_stringent,
                         epiRomics_dB)
#> >> preparing features information...      2021-10-05 17:54:01
#> >> identifying nearest features...        2021-10-05 17:54:01
#> >> calculating distance from peak to TSS...   2021-10-05 17:54:02
#> >> assigning genomic annotation...        2021-10-05 17:54:02
#> >> adding gene annotation...          2021-10-05 17:54:09
#> 'select()' returned 1:1 mapping between keys and columns
#> >> assigning chromosome lengths           2021-10-05 17:54:09
#> >> done...                    2021-10-05 17:54:09
```

```
epiRomics_track_layer_human(
  epiRomics_putative_enhanceosome_stringent,
  epiRomics_index = 1,
  epiRomics_dB = epiRomics_dB,
  epiRomics_track_connection = epiRomics_track_connection
)
#> [1] "not empty"
```

```
#> [1] 233.705
#> [1] "not empty"
#> [1] 221.083
#> [1] "not empty"
#> [1] 9.31751
#> [1] "not empty"
#> [1] 15.8259
```

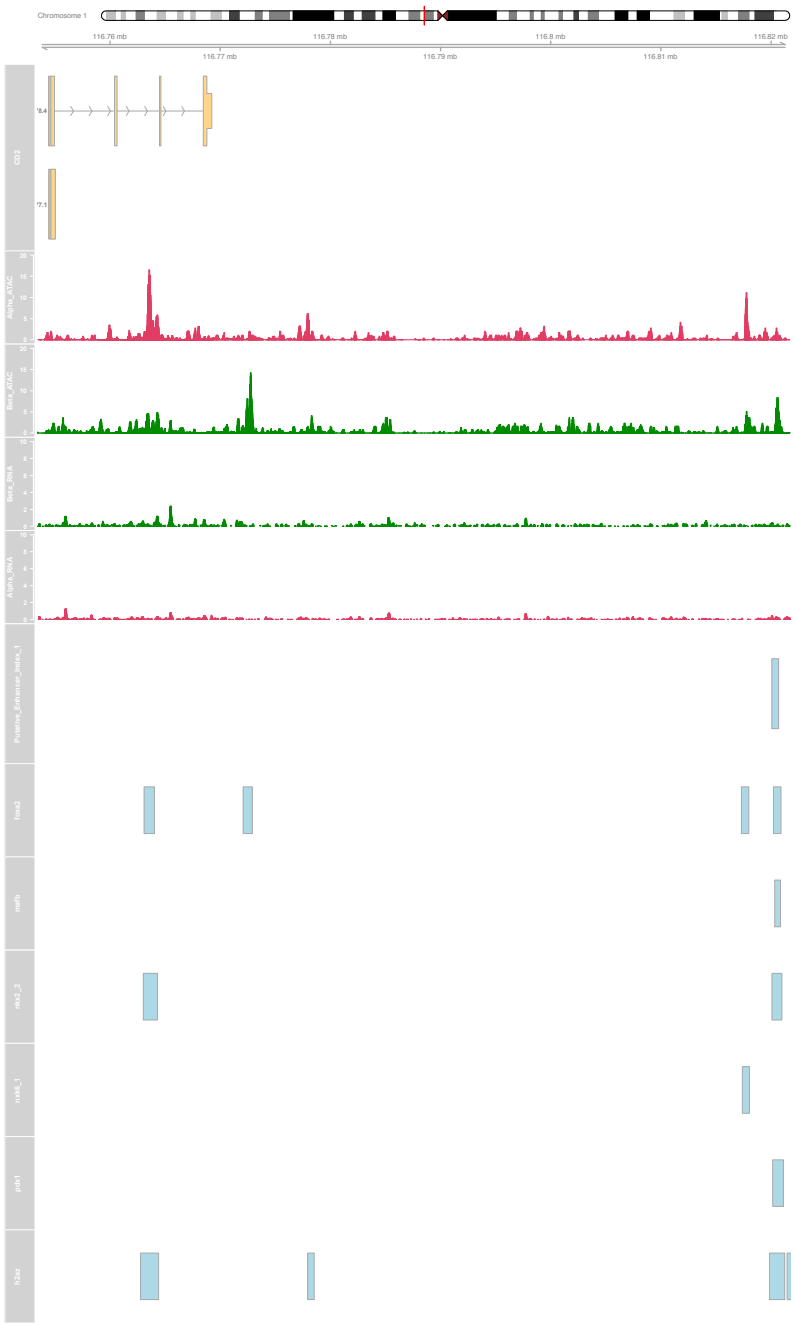

Chromosome 1
164.6 mb
164.65 mb
164.7 mb
164.75 mb
164.8 mb
PIDX1
ENST00000560469.5
ENST00000496120.6
ENST00000468104.5
ENST00000560641.5
ENST00000482110.5
ENST00000421630.1
ENST00000558837.5
ENST00000559560.1
ENST00000540236.3
ENST00000612123.3
Alpha_ATAC
Beta_ATAC
Beta_RNA
Alpha_RNA
fox-a2
meis
mix-2_2
mix6_1
pdx1
h1az

How can we use these putative enhanceosome regions to infer biology between cell states? In this example, we will integrate ATAC-Seq data differential testing showing differences in chromatin accessibility between alpha and beta cells

```
## Read differentially binding data generated with DiffBind.
## DE comparing human alpha and beta cell chromatin.

b.v.a <-
  read.csv(system.file("extdata", "DBA_Beta_Versus_Alpha.csv", package = "epiRomics"))
b.v.a <- GRanges(b.v.a)

# Filter for beta enriched chromatin regions

beta.enriched <- b.v.a[b.v.a$Fold >= 1, ]

# Connect to our putative enhanceosomes

beta_enhancer_regions <-
  epiRomics_regions_of_interest(epiRomics_putative_enhanceosome_fantom, beta.enriched)
```

Now, lets visualize the top candidate region we found after connecting our differential chromatin analysis with the putative enhanceosomes

```
epiRomics_track_layer_human(
  beta_enhancer_regions,
  epiRomics_index = 1,
  epiRomics_dB = epiRomics_dB,
  epiRomics_track_connection = epiRomics_track_connection
)
#> [1] "not empty"
#> [1] 16.5139
#> [1] "not empty"
#> [1] 14.2229
#> [1] "not empty"
#> [1] 2.38379
#> [1] "not empty"
#> [1] 1.29041
```


Chromosome 1
116.76 mb
116.77 mb
116.78 mb
116.79 mb
116.8 mb
116.81 mb
116.82 mb
CD2
Alpha_ATAC
Beta_ATAC
Beta_RNA
Alpha_RNA
Putative_Enhancer_Index_1
foxa2
mafb
nkx2_2
nr4a1_1
pdx1
h2az

Session Information

Here is the output of `sessionInfo()` on the system on which this document was compiled:

```
sessionInfo()
#> R version 4.1.1 (2021-08-10)
#> Platform: x86_64-apple-darwin17.0 (64-bit)
#> Running under: macOS Catalina 10.15.7
#>
#> Matrix products: default
#> BLAS:   /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRblas.0.dylib
#> LAPACK: /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRlapack.dylib
#>
#> locale:
#> [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
#>
#> attached base packages:
#> [1] stats4    parallel  stats     graphics  grDevices utils     datasets
#> [8] methods   base
#>
#> other attached packages:
#>  [1] BSgenome.Hsapiens.UCSC.hg38_1.4.3
#>  [2] BSgenome_1.60.0
#>  [3] rtracklayer_1.52.1
#>  [4] Biostrings_2.60.2
#>  [5] XVector_0.32.0
#>  [6] org.Hs.eg.db_3.13.0
#>  [7] TxDb.Hsapiens.UCSC.hg38.knownGene_3.13.0
#>  [8] GenomicFeatures_1.44.2
#>  [9] AnnotationDbi_1.54.1
#> [10] Biobase_2.52.0
#> [11] GenomicRanges_1.44.0
#> [12] GenomeInfoDb_1.28.4
#> [13] IRanges_2.26.0
#> [14] S4Vectors_0.30.2
#> [15] BiocGenerics_0.38.0
#> [16] epiRomics_0.1.3
#>
#> loaded via a namespace (and not attached):
#>   [1] utf8_1.2.2
#>   [2] tidyselect_1.1.1
#>   [3] RSQLite_2.2.8
#>   [4] htmlwidgets_1.5.4
#>   [5] grid_4.1.1
#>   [6] BiocParallel_1.26.2
#>   [7] scatterpie_0.1.7
#>   [8] munsell_0.5.0
#>   [9] codetools_0.2-18
#>  [10] withr_2.4.2
#>  [11] colorspace_2.0-2
#>  [12] GOSemSim_2.18.1
#>  [13] filelock_1.0.2
#>  [14] knitr_1.36
```

```
#>  [15] rstudioapi_0.13
#>  [16] DOSE_3.18.3
#>  [17] MatrixGenerics_1.4.3
#>  [18] GenomeInfoDbData_1.2.6
#>  [19] polyclip_1.10-0
#>  [20] bit64_4.0.5
#>  [21] farver_2.1.0
#>  [22] treeio_1.16.2
#>  [23] vctrs_0.3.8
#>  [24] generics_0.1.0
#>  [25] TH.data_1.1-0
#>  [26] xfun_0.26
#>  [27] biovizBase_1.40.0
#>  [28] BiocFileCache_2.0.0
#>  [29] party_1.3-9
#>  [30] regioneR_1.24.0
#>  [31] R6_2.5.1
#>  [32] graphlayouts_0.7.1
#>  [33] AnnotationFilter_1.16.0
#>  [34] bitops_1.0-7
#>  [35] cachem_1.0.6
#>  [36] fgsea_1.18.0
#>  [37] gridGraphics_0.5-1
#>  [38] DelayedArray_0.18.0
#>  [39] assertthat_0.2.1
#>  [40] vroom_1.5.5
#>  [41] promises_1.2.0.1
#>  [42] BiocIO_1.2.0
#>  [43] scales_1.1.1
#>  [44] multcomp_1.4-17
#>  [45] ggraph_2.0.5
#>  [46] nnet_7.3-16
#>  [47] enrichplot_1.13.1.992
#>  [48] gtable_0.3.0
#>  [49] tidygraph_1.2.0
#>  [50] sandwich_3.0-1
#>  [51] ensembldb_2.16.4
#>  [52] rlang_0.4.11
#>  [53] splines_4.1.1
#>  [54] lazyeval_0.2.2
#>  [55] dichromat_2.0-0
#>  [56] checkmate_2.0.0
#>  [57] BiocManager_1.30.16
#>  [58] yaml_2.2.1
#>  [59] reshape2_1.4.4
#>  [60] backports_1.2.1
#>  [61] httpuv_1.6.3
#>  [62] qvalue_2.24.0
#>  [63] Hmisc_4.5-0
#>  [64] tools_4.1.1
#>  [65] ggplotify_0.1.0
#>  [66] ggplot2_3.3.5
#>  [67] gplots_3.1.1
```

```
#>  [68] ellipsis_0.3.2
#>  [69] RColorBrewer_1.1-2
#>  [70] Rcpp_1.0.7
#>  [71] plyr_1.8.6
#>  [72] base64enc_0.1-3
#>  [73] progress_1.2.2
#>  [74] zlibbioc_1.38.0
#>  [75] purrr_0.3.4
#>  [76] RCurl_1.98-1.5
#>  [77] prettyunits_1.1.1
#>  [78] rpart_4.1-15
#>  [79] viridis_0.6.1
#>  [80] zoo_1.8-9
#>  [81] SummarizedExperiment_1.22.0
#>  [82] ggrepel_0.9.1
#>  [83] cluster_2.1.2
#>  [84] magrittr_2.0.1
#>  [85] data.table_1.14.2
#>  [86] DO.db_2.9
#>  [87] mvtnorm_1.1-2
#>  [88] ProtGenerics_1.24.0
#>  [89] matrixStats_0.61.0
#>  [90] hms_1.1.1
#>  [91] patchwork_1.1.1
#>  [92] mime_0.12
#>  [93] evaluate_0.14
#>  [94] xtable_1.8-4
#>  [95] XML_3.99-0.8
#>  [96] jpeg_0.1-9
#>  [97] gridExtra_2.3
#>  [98] compiler_4.1.1
#>  [99] biomaRt_2.48.3
#> [100] tibble_3.1.5
#> [101] KernSmooth_2.23-20
#> [102] shadowtext_0.0.9
#> [103] crayon_1.4.1
#> [104] htmltools_0.5.2
#> [105] ggfun_0.0.4
#> [106] later_1.3.0
#> [107] tzdb_0.1.2
#> [108] Formula_1.2-4
#> [109] tidyr_1.1.4
#> [110] aplot_0.1.1
#> [111] libcoin_1.0-9
#> [112] DBI_1.1.1
#> [113] formatR_1.11
#> [114] ChIPseeker_1.28.3
#> [115] tweenr_1.0.2
#> [116] dbplyr_2.1.1
#> [117] MASS_7.3-54
#> [118] rappdirs_0.3.3
#> [119] boot_1.3-28
#> [120] Matrix_1.3-4
```

```
#> [121] readr_2.0.2
#> [122] Gviz_1.36.2
#> [123] igraph_1.2.6
#> [124] TxDb.Hsapiens.UCSC.hg19.knownGene_3.2.2
#> [125] pkgconfig_2.0.3
#> [126] GenomicAlignments_1.28.0
#> [127] coin_1.4-1
#> [128] foreign_0.8-81
#> [129] xml2_1.3.2
#> [130] ggtree_3.0.4
#> [131] yulab.utils_0.0.2
#> [132] stringr_1.4.0
#> [133] VariantAnnotation_1.38.0
#> [134] digest_0.6.28
#> [135] strucchange_1.5-2
#> [136] rmarkdown_2.11
#> [137] fastmatch_1.1-3
#> [138] tidytree_0.3.5
#> [139] htmlTable_2.2.1
#> [140] annotatr_1.18.1
#> [141] restfulr_0.0.13
#> [142] curl_4.3.2
#> [143] gtools_3.9.2
#> [144] shiny_1.7.1
#> [145] Rsamtools_2.8.0
#> [146] modeltools_0.2-23
#> [147] rjson_0.2.20
#> [148] jsonlite_1.7.2
#> [149] lifecycle_1.0.1
#> [150] nlme_3.1-153
#> [151] viridisLite_0.4.0
#> [152] fansi_0.5.0
#> [153] pillar_1.6.3
#> [154] lattice_0.20-45
#> [155] plotrix_3.8-2
#> [156] KEGGREST_1.32.0
#> [157] fastmap_1.1.0
#> [158] httr_1.4.2
#> [159] survival_3.2-13
#> [160] GO.db_3.13.0
#> [161] interactiveDisplayBase_1.30.0
#> [162] glue_1.4.2
#> [163] png_0.1-7
#> [164] BiocVersion_3.13.1
#> [165] bit_4.0.4
#> [166] ggforce_0.3.3
#> [167] stringi_1.7.5
#> [168] blob_1.2.2
#> [169] AnnotationHub_3.0.1
#> [170] caTools_1.18.2
#> [171] latticeExtra_0.6-29
#> [172] memoise_2.0.0
#> [173] dplyr_1.0.7
```

```
#> [174] ape_5.5
```

References

1. Capobianco E: **RNA-Seq Data: A Complexity Journey**. *Comput Struct Biotechnol J* 2014, **11**(19):123-130.
2. Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA: **The technology and biology of single-cell RNA sequencing**. *Mol Cell* 2015, **58**(4):610-620.
3. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szczesniak MW, Gaffney DJ, Elo LL, Zhang X *et al*: **A survey of best practices for RNA-seq data analysis**. *Genome Biol* 2016, **17**:13.
4. Kulkarni A, Anderson AG, Merullo DP, Konopka G: **Beyond bulk: a review of single cell transcriptomics methodologies and applications**. *Curr Opin Biotechnol* 2019, **58**:129-136.
5. Chen G, Ning B, Shi T: **Single-Cell RNA-Seq Technologies and Related Computational Data Analysis**. *Front Genet* 2019, **10**:317.
6. Karczewski KJ, Snyder MP: **Integrative omics for health and disease**. *Nature reviews Genetics* 2018, **19**(5):299-310.
7. Muller C, Leutz A: **Chromatin remodeling in development and differentiation**. *Curr Opin Genet Dev* 2001, **11**(2):167-174.
8. Andrey G, Mundlos S: **The three-dimensional genome: regulating gene expression during pluripotency and development**. *Development* 2017, **144**(20):3646-3658.
9. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE: **High-resolution mapping and characterization of open chromatin across the genome**. *Cell* 2008, **132**(2):311-322.
10. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ: **Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position**. *Nat Methods* 2013, **10**(12):1213-1218.
11. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ: **ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide**. *Current protocols in molecular biology / edited by Frederick M Ausubel [et al]* 2015, **109**:21 29 21-21 29 29.
12. Starks RR, Biswas A, Jain A, Tuteja G: **Combined analysis of dissimilar promoter accessibility and gene expression profiles identifies tissue-specific genes and actively repressed networks**. *Epigenetics Chromatin* 2019, **12**(1):16.
13. de la Torre-Ubieta L, Stein JL, Won H, Opland CK, Liang D, Lu D, Geschwind DH: **The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis**. *Cell* 2018, **172**(1-2):289-304 e218.
14. Daugherty AC, Yeo RW, Buenrostro JD, Greenleaf WJ, Kundaje A, Brunet A: **Chromatin accessibility dynamics reveal novel functional enhancers in C. elegans**. *Genome Res* 2017, **27**(12):2096-2107.
15. Neph S, Stergachis AB, Reynolds A, Sandstrom R, Borenstein E, Stamatoyannopoulos JA: **Circuitry and dynamics of human transcription factor regulatory networks**. *Cell* 2012, **150**(6):1274-1286.
16. Pastor WA, Stroud H, Nee K, Liu W, Pezic D, Manakov S, Lee SA, Moissiard G, Zamudio N, Bourc'his D *et al*: **MORC1 represses transposable elements in the mouse male germline**. *Nat Commun* 2014, **5**:5795.
17. Calo E, Wysocka J: **Modification of enhancer chromatin: what, how, and why?** *Mol Cell* 2013, **49**(5):825-837.
18. Mellor J: **The dynamics of chromatin remodeling at promoters**. *Mol Cell* 2005, **19**(2):147-157.
19. Bemer M: **Unraveling the Complex Epigenetic Mechanisms that Regulate Gene Activity**. 2018, **1675**.

20. Adusumalli S, Mohd Omar MF, Soong R, Benoukraf T: **Methodological aspects of whole-genome bisulfite sequencing analysis**. *Brief Bioinform* 2015, **16**(3):369-379.
21. Arand J, Spieler D, Karius T, Branco MR, Meilinger D, Meissner A, Jenuwein T, Xu G, Leonhardt H, Wolf V *et al*: **In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases**. *PLoS Genet* 2012, **8**(6):e1002750.
22. He XJ, Chen T, Zhu JK: **Regulation and function of DNA methylation in plants and animals**. *Cell Res* 2011, **21**(3):442-465.
23. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D: **The human genome browser at UCSC**. *Genome Res* 2002, **12**(6):996-1006.
24. Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D: **BigWig and BigBed: enabling browsing of large distributed datasets**. *Bioinformatics* 2010, **26**(17):2204-2207.
25. Lizio M, Harshbarger J, Abugessaisa I, Noguchi S, Kondo A, Severin J, Mungall C, Arenillas D, Mathelier A, Medvedeva YA *et al*: **Update of the FANTOM web resource: high resolution transcriptome of diverse cell types in mammals**. *Nucleic acids research* 2017, **45**(D1):D737-D743.
26. Zheng HF, Rong JJ, Liu M, Han F, Zhang XW, Richards JB, Wang L: **Performance of genotype imputation for low frequency and rare variants from the 1000 genomes**. *PloS one* 2015, **10**(1):e0116487.
27. Wang K, Li M, Hakonarson H: **ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data**. *Nucleic acids research* 2010, **38**(16):e164.
28. Dimitrieva S, Bucher P: **Genomic context analysis reveals dense interaction network between vertebrate ultraconserved non-coding elements**. *Bioinformatics* 2012, **28**(18):i395-i401.
29. Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ: **Software for computing and annotating genomic ranges**. *PLoS Comput Biol* 2013, **9**(8):e1003118.
30. Spitz F, Furlong EE: **Transcription factors: from enhancer binding to developmental control**. *Nature reviews Genetics* 2012, **13**(9):613-626.
31. Spicuglia S, Vanhille L: **Chromatin signatures of active enhancers**. *Nucleus* 2012, **3**(2):126-131.
32. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA *et al*: **Histone H3K27ac separates active from poised enhancers and predicts developmental state**. *Proc Natl Acad Sci U S A* 2010, **107**(50):21931-21936.
33. Kingsford C, Salzberg SL: **What are decision trees?** *Nat Biotechnol* 2008, **26**(9):1011-1013.
34. Hahne F, Ivanek R: **Visualizing Genomic Data Using Gviz and Bioconductor**. In: *Statistical Genomics: Methods and Protocols.* Edited by Mathé E, Davis S. New York, NY: Springer New York; 2016: 335-351.
35. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J *et al*: **Bioconductor: open software development for computational biology and bioinformatics**. *Genome Biol* 2004, **5**(10):R80.

IX. Chapter 9: Curriculum Vitae

Publications

1. Huang J*, **Mawla AM***, Huising MO. (2021). Transcriptomic and epigenomic characterizations of alpha and beta cells in lean and ob/ob mice. *Early draft.*

2. Noguchi GM, Huang J, **Mawla AM**, Huising MO. (2021). Functional heterogeneity among pancreatic alpha cells. *Advanced draft.*

3. **Mawla AM**, van der Meulen T, Huising MO. (2021). Chromatin accessibility differences between alpha, beta, and delta cells identifies common endocrine- and cell-specific enhancers. BMC Genomics *Advanced draft.*

4. **Mawla AM**, Huising MO. (2021). There and Back Again (tabaSeq): An unexpected journey between single and bulk sequencing within pancreatic islets. *Early draft.*

5. **Mawla AM**, Huising MO. (2021). epiRomics – a multi-omics R package to identify and visualize enhancers. Pre-print on BiorXiv*, to be submitted*.

6. **Mawla AM**, Huising MO. (2021). RoiViz: an R Package for visualizing ROI intensity from GCAMP6 calcium imaging Pre-print on BiorXiv*, to be submitted*.

7. Lee S, Xu Haixia, **Mawla AM**, Huising MO, Annes JP. (2021). β-cell Dysfunction and Insulinopenic Diabetes are precipitated by Mitochondrial Succinate Dehydrogenase Deficiency. Nature Communications, *In review.*

8. Furterer, A, Gurlo, T, Wang, Q, Gao, F, Rosenberger, M, Pei, L, Nomoto, H, **Mawla, AM**, Huising, MO, Coppola, G, Butler, PC. (2020). IAPP induced beta cell stress recapitulates the islet transcriptome in type 2 diabetes. Diabetologia. *Accepted.*

9. **Mawla AM**, Huising MO. (2019). Navigating the Depths and Avoiding the Shallows of Pancreatic Islet Transcriptomes. Diabetes, 68(7), 1380-1393.

10. van der Meulen T, Lee S, Noordeloos E, Donaldson CJ, Adams MW, Noguchi GM, **Mawla AM**, Huising MO. (2018). Artemether Does Not Turn α Cells into β Cells. Cell Metabolism. 27, 1-8.

11. van der Meulen T, **Mawla AM**, DiGruccio MR, Adams MW, Nies V, Dólleman S, Liu S, Ackermann AM, Cáceres E, Hunter AE, Kaestner KH, Donaldson CJ, Huising MO. (2017). Virgin Beta Cells Persist throughout Life at a Neogenic Niche within Pancreatic Islets. Cell Metabolism. 25, 1-16.

12. Yan D, Tekin D, Bademu G, Foster J, Cengiz F.B, Kannan-Sundhari A, Guo S, Mittal R, Zou B, Grati M, Kabahuma R, Kameswaran M, Lasisi TJ, Adedeji WA, Lasisi AO, Menendez I, Herrera M, Carranza C, Maroofian R, Crosby A, Bensaid M, Mahmoudi S, Behram M,

Mojarrad M, Feng Y, Duman D, **Mawla AM**, Nord AS, Blanton SH, Liu XZ, Tekin M. (2016). Spectrum of DNA variants for non-syndromic deafness in a large cohort from multiple continents. Human Genetics. 135(8), 953–961.

13. DiGruccio MR, **Mawla AM**, Donaldson CJ, Noguchi GM, Vaughan J, Cowing-Zitron C, van der Meulen T, Huising MO. (2016). Comprehensive alpha, beta and delta cell transcriptomes reveal that ghrelin selectively activates delta cells and promotes somatostatin release from pancreatic islets. Molecular Metabolism, 5(7), 449-458.

14. Acknowledgment: Lesh TA, Tanase C, Geib BR, Niendam TA, Yoon JH, Minzenberg MJ, Ragland JD, Solomon M, Carter CS. (2015). A Multimodal Analysis of Antipsychotic Effects on Brain Structure and Function in First-Episode Schizophrenia. JAMA Psychiatry, 72(3), 226.

15. **Mawla AM**. Antibiotic-resistant Bacteria. The Aggie Newspaper February 2012.

Presentations

1. Talk: **Mawla AM**. ATACing the Pancreas: Evaluating Chromatin Accessibility and Transcription in Major Islet Cell Types. Nara Institute International Student Workshop. Ikoma, Japan. November 2020.

2. Poster: **Mawla AM**, Huising MO. Navigating the Depths and Avoiding the Shallows of Pancreatic Islet Cell Transcriptomes. University of California, Davis Human Genomics Symposium. November 2019.

3. Poster: Huang, J, **Mawla AM**, Huising MO. Transcriptomic comparison of alpha and beta cells in lean and ob/ob mice. University of California, Davis Human Genomics Symposium. November 2019.

4. Talk: **Mawla AM**. There and Back Again: An Unexpected Journey between Single and Bulk Sequencing within Pancreatic Islets. UC San Francisco Obesity & Diabetes 2019 Research Retreat. November 2019.

5. Poster: **Mawla, AM**, Huising MO. Navigating the Depths and Avoiding the Shallows of Pancreatic Islet Cell Transcriptomes. UC San Francisco Obesity & Diabetes 2019 Research Retreat. November 2019.

6. Poster: Huang, J, **Mawla AM**, Huising MO. Transcriptomic comparison of alpha and beta cells in lean and ob/ob mice. UC San Francisco Obesity & Diabetes 2019 Research Retreat. November 2019.

7. Talk: **Mawla AM**. Navigating the Depths and Avoiding the Shallows of Pancreatic Islet Transcriptomes. Western Region Islet Study Group. October 2019.

8. Poster: **Mawla, AM**, Huising MO. Navigating the Depths and Avoiding the Shallows of Pancreatic Islet Cell Transcriptomes. AAAS Pacific Division. June 2019.

9. Talk: **Mawla, AM**. A Glance at Newfound Promises and Challenges Within Big Data - Omics Exploring Pancreatic Islets. AAAS Pacific Division, Opportunities and Challenges for Large Data Sets and their Analysis in Contemporary and Future Science. June 2019.

10. Talk **Mawla AM**. Navigating the Depths and Avoiding the Shallows of Pancreatic Islet Transcriptomes. Stanford Diabetes Center Bay Area Islet Biologists Meeting. February 2019.

11. Poster: **Mawla AM**, Huising MO. Navigating the Depths and Avoiding the Shallows of Pancreatic Islet Transcriptomes. UC San Francisco Obesity & Diabetes 2018 Research Retreat. September 2018.

12. Poster: **Mawla AM**, DeLong AN, van der Meulen T, Huising MO. ATAC-ing the Pancreas. UC San Francisco Obesity & Diabetes 2017 Research Retreat. September 2017.

13. Poster: **Mawla AM**, van der Meulen T, Huising MO. Integrative -Omics with Machine Learning to Elucidate Cell Fate. UC San Francisco Obesity & Diabetes 2016 Research Retreat. September 2016.

14. Poster: van der Meulen T, DiGruccio MR, **Mawla AM**, Adams M, Donaldson CJ, Hunter A, Nies V, Dolleman S, Caceres E, Huising MO. Life-long Formation of New Beta Cells at a Neurogenic Niche within Pancreatic Islets through Spontaneous Transdifferentiation. American Diabetes Association 2016 Annual Meeting. June 2016.

15. Presentation: **Mawla AM**. Copy Number Variation Detection. UC Davis Genome Center Bioinformatics Workshop. June 2015.

16. Poster: Lesh TA, **Mawla AM**, Minzenberg M, Yoon J, Niendam TA, Ragland JD, Solomon M, Carter CS. Effects of Past Mild Cannabis Use on Cognition in First Episode Schizophrenia: A Multimodal Analysis of Brain Structure and Function. International Congress on Schizophrenia Research. April 2015.

17. Poster: **Mawla AM**, Nord AS. Expanding the spectrum of mutations identifiable via targeted clinical sequencing using depth of coverage CNV detection. UC Davis Human Genomics Forum. November 2014.

18. Presentation: **Mawla AM**. Clearing the Smoke: The Cognitive, Neuroanatomical, and Functional Impacts of Cannabis Use in Schizophrenia. UC Davis Undergraduate Research Conference. April 2014.

Peer Review

- Reviewed 15+ papers from journals such as Nature Scientific Reports, Journal of Endocrinology, Proceedings of the National Academy of Sciences of the United States of America, Frontiers in Genetics, Diabetologia, and Physiological Genomics under the guidance of my PI.

Professional Affiliations

Chapter Officer – Sigma Xi, The Scientific Research Society, 2019 – Present

- After being nominated as an associate member, spearheaded the renewal of the UC Davis Chapter of Sigma Xi. Currently an active member and chapter officer, in charge of guiding nominations and elections for official chapter positions, as well as organizing seminars and symposiums to garner further membership and promote scientific outreach.

Undergraduate Advisor – Disabilities Issues Administrative Advisory Committee (DIAAC), September 2011 – March 2015

- Assisted in identification, evaluation, and recommendation of actions for problems related to employment, educational opportunities, and removal of architectural barriers to accommodate the needs of the disabled population at UC Davis.

Honors and Awards

Student Research Award – A Glance at Newfound Promises and Challenges Within Big Data-Omics Exploring Pancreatic Islets. AAAS Pacific Division, Opportunities and Challenges for Large Data Sets and their Analysis in Contemporary and Future Science, June 2019

Amazon AWS Machine Learning Research Award, 2019

Stephen F. and Bettina A. Sims Immunology Fellowship, 2017

Top Individual Fundraiser Award – Bonnie J. Addario Lung Cancer Foundation Walk for the Cure, Sacramento, 2017

Professional Development

- GitHub Universe - Virtual December 2020
- Nara Institute International Student Workshop, Ikoma, Japan November 2020
- UC San Diego IGM Symposium on Spatial Omics, San Diego, CA October 2020
- European Association for the Study of Diabetes, Vienna, Austria September 2020
- Stanford Diabetes Center Bay Area Islet Biologists Meeting, Stanford, CA July 2020
- UC San Francisco Obesity & Diabetes Research Retreat, Santa Cruz, CA November 2019
- UC Davis Human Genomics Forum, University of California, Davis November 2019
- Western Region Islet Study Group, Victoria, BC October 2019
- Course-Based Undergraduate Research Experiences (CURE), Institute Certification, Davis, CA July 2019
- AAAS Pacific Division Conference, Ashland, OR June 2019
- Stanford Diabetes Center Bay Area Islet Biologists Meeting, Stanford, CA February 2019
- UC Davis Inaugural Artificial Intelligence & Machine Learning Symposium October 2018
- UC San Francisco Obesity & Diabetes Research Retreat, Santa Cruz, CA September 2018
- UC Davis Human Genomics Forum, University of California, Davis November 2017
- UC San Francisco Obesity & Diabetes Research Retreat, Santa Cruz, CA September 2017
- UC Davis Human Genomics Forum, University of California, Davis November 2016
- UC San Francisco Obesity & Diabetes Research Retreat, Santa Cruz, CA September 2016
- The R User Conference, Stanford University, Stanford, CA June 2016

- BioConductor: Where Software & Biology Connect, Stanford, University, Stanford, CA June 2016
- UC Davis Human Genomics Forum, University of California, Davis November 2014